Nonparametric Transformation Models

Dissertation

zur Erlangung des Doktorgrades
an der Fakultät für Mathematik, Informatik und Naturwissenschaften
Fachbereich Mathematik
der Universität Hamburg

vorgelegt von

Nick Kloodt
Hamburg, 2019

Folgende Gutachter empfehlen die Annahme der Dissertation:
Prof. Dr. Natalie Neumeyer
Prof. Dr. Ingrid Van Keilegom
Prof. Dr. María Dolores Jiménez Gamero

Tag der Disputation: 24. September 2019

Danksagung

Die letzten drei Jahre stellten für mich eine zwar herausfordernde, aber sehr schöne Zeit dar, die natürlich auch durch einige Entbehrungen und Frustration, aber insbesondere durch diverse schöne Erlebnisse und Chancen geprägt war. Ich habe viele Eindrücke und Erfahrungen sammeln dürfen, die ich nicht missen möchte und durch die ich mich nicht nur fachlich, sondern auch persönlich weiterentwickeln konnte. Nachdem sich nun nach der Verteidigung meiner Arbeit die Aufregung etwas gelegt hat und die letzten Stellen dieser Dissertation überarbeitet sind, möchte ich noch einigen Menschen danken, ohne die diese Zeit mit Sicherheit nicht so verlaufen wäre.

Da ist zunächst meine Betreuerin und Erstgutachterin Natalie Neumeyer, die mich eigentlich meine gesamte Studienzeit lang begleitet hat. Natalie, du hast meine akademische Laufbahn mit Abstand am stärksten geprägt. Ich habe mal nachgezählt und seit dem Grundkurs Stochastik im Wintersemester 2012/ 2013 habe ich bis zum Masterabschluss Prüfungsleistungen im Wert von 111 CP (für die Zahlenfreaks 42 im Bachelor und 69 im Master) bei dir absolviert. Besonders deine ruhige und gelassene Art empfinde ich als sehr angenehm und ich habe schon so manches Mal erlebt, dass große Probleme nach einem Gespräch mit dir auf einmal viel kleiner und bewältigbar erschienen. Dafür und für die vielen Möglichkeiten, die du mir geschaffen hast, seien es die bisherigen Konferenzen oder der Forschungsbesuch in Leuven, möchte ich dir sehr herzlich danken.
Ebenfalls möchte ich meiner Zweitgutachterin Ingrid Van Keilegom nicht für nur die Mammutaufgabe, sich durch die 293 Seiten des Prüfungsexemplars meiner Arbeit zu kämpfen, danken, sondern auch für die Möglichkeit, am Anfang des letzten Jahres für sieben Wochen in die Welt der KUL und der kleinen, aber sehr niedlichen (im positivsten Sinne) Studentenstadt Leuven eintauchen zu dürfen. Ich hatte dort eine sehr interessante und schöne Zeit.
Dann möchte ich noch dem gesamten Fachbereich Mathematische Statistik und Stochastische Prozesse der Universität Hamburg danken. Ich habe die Atmosphäre hier immer als offen und freundlich wahrgenommen und bin froh, die letzten drei Jahre ein Teil dieses Bereichs gewesen zu sein und noch ein weiteres Jahr hier bleiben zu dürfen.
Selbstverständlich machen einen solchen Bereich nicht etwa die Mensa oder das Büro im Geomatikum (dem ich an dieser Stelle für die heißen Sommer und die kalten Winter danken möchte), sondern vielmehr erst die Kollegen wirklich wertvoll. Gerade meine Doktorbrüder- und -schwestern hatten immer ein offenes Ohr und waren extrem hilfsbereit. So wurde ich beispielsweise gegen Ende meiner Arbeit, schon bevor ich so weit war, diverse Male gefragt,

ob sie mir helfen können oder meine Arbeit Korrektur lesen sollen. Für das Lesen nochmal besten Dank an Florian, Maria, Miran, Sebastian und Rui!
Noch mehr als die inner- habe ich unsere außeruniversitären Aktivitäten genossen. Ich gebe zu, nicht ganz vorurteilsfrei mit dem Bild eines steifen, in sich gekehrten Mathematikers (von dem ich mittlerweile weiß, dass es auf Gegenseitigkeit beruhte) angefangen zu haben, doch dieses Bild hat sich schnell verflüchtigt. Von Barbarabar und Boot fahren über Heidepark, Jumphouse und Minigolf bis hin zu Pulverfass und Weihnachtsmarkt (die Liste ist noch erheblich länger) haben wir eigentlich so ziemlich alles mitgenommen, was Spaß macht. Nicht nur für diese ganzen Erlebnisse, sondern auch für den regen Austausch möchte ich euch sehr danken. Ihr wart mit den gleichen Herausforderungen konfrontiert und so dicht dran an meinem Alltag wie kaum jemand anderes. Es hat mir sehr geholfen, meine Empfindungen und Erfahrungen mit euch teilen zu können, und ich würde mich glücklich schätzen, in Zukunft mit auch nur annähernd so guten Kollegen zusammenarbeiten zu dürfen.

Einen so großen Platz die Uni in den letzten Jahren in meinem Leben auch eingenommen hat, habe ich immer den Ausgleich bei meiner Familie, meinen Freunden und beim Sport genossen.
Im manchmal stressigen Arbeitsalltag war es mir wichtig, von Zeit zu Zeit abzuschalten, mich auszupowern und auf andere Gedanken zu kommen. Von großer Bedeutung war dabei für mich der Sport beim Fußball und insbesondere beim Badminton. In welcher Verfassung ich auch in die Halle gegangen bin, bin ich danach eigentlich immer mit einem besseren Gefühl wieder rausgegangen. Nach mittlerweile über 22 Jahren genieße ich das Gefühl, unter Freunden für- und miteinander zu trainieren und zu kämpfen, mehr denn je und möchte für die schöne Zeit sowie die Nachsicht, wenn ich in den letzten Jahren mal die ein oder andere Sache schleifen lassen habe, danken. Gerade Fabian musste sich im letzten Jahr nicht nur sportlich, sondern als mein Mitbewohner auch persönlich mit mir beschäftigen und ließ trotz seiner knappen Zeit keine Möglichkeit ungenutzt, mich mental (und glücklicherweise auch kulinarisch) zu unterstützen.
Stellvertretend für meine Freunde möchte ich der PPC danken. Es ist nicht selbstverständlich und ich weiß es sehr zu schätzen, dass wir uns in dieser großen Gruppe seit zum Teil über zwanzig Jahren derart nah stehen, dass man nicht nur über alles reden, sondern auch zusammenwohnen und in den Urlaub fahren kann.

Vor dem letzten Absatz möchte ich allen Personen, die mich auf meinem bisherigen Weg begleitet und unterstützt haben und die hier noch nicht erwähnt worden sind, bspw. meiner Mathematiklehrerin Kerstin Kuhn, die mich erst zum Mathematik- bzw. Wirtschaftsmathematikstudium bewegt hat, meinen Dank aussprechen.

Als letzten und wichtigsten Punkt möchte ich meiner Familie danken. Ich habe großen Respekt vor dem, was ihr geleistet habt, seien es meine Großeltern, meine Geschwister Nora und Fabian oder meine Eltern. Dass ich von euch stets bedingungslose Wertschätzung und Unterstützung erfahren habe, hat mir immer geholfen und ich hatte großes Glück, in diese Familie hineingeboren worden zu sein. Ihr habt mir all das hier ermöglicht und ich weiß,

dass ich zu jeder Zeit zu euch kommen kann. Dafür möchte ich euch von ganzem Herzen danken.

Contents

List of Figures

List of Tables

Notations and Conventions

Although all of the following notations will be introduced when they arise for the first time, already a selection of those with a universal meaning throughout the whole thesis is given in this Section. Corresponding estimators of the listed parameters and functions will be marked with a "ˆ" and are omitted here for the sake of brevity.

Notations

$\mathbb{N}$	set of natural numbers
$\mathbb{Q}$	set of rational numbers
$\mathbb{R}$	set of real numbers
Y	dependent variable, real valued
X	independent variable, $\mathbb{R}^{d_X}$-valued for some $d_X \in \mathbb{N}$
ε	error term, real valued, centred and independent of X
g	regression function
σ^2	variance (function) of the errors
h	nonparametric transformation function
Θ	transformation parameter set, subset of $\mathbb{R}^{d_\Theta}$ for some $d_\Theta \in \mathbb{N}$
θ	transformation parameter
Λ_θ	parametric transformation function
B	regression parameter set, subset of $\mathbb{R}^{d_B}$ for some $d_B \in \mathbb{N}$
β	regression parameter
g_β	parametric regression function
$F_{Y\|X}$	distribution function of Y conditional on X
$F_{Y\|X}^{-1}$	conditional quantile function of Y conditional on X
$f_{Y\|X}$	density function of Y conditional on X

v, w	weight functions
K	univariate kernel function in the context of kernel estimation
$\mathbf{K}$	multivariate kernel function in the context of kernel estimation
$\mathcal{K}$	integrated kernel function
h_y, h_x	bandwidths
μ	finite measure
$\xrightarrow{\mathcal{D}}$	convergence in distribution
$\rightsquigarrow$	weak convergence
o_p	term, that converges to zero in probability
$\mathcal{O}_p$	term, that is bounded in probability

Conventions

Let d be a natural number.

- Let $f : \mathbb{R}^{d_X} \to \mathbb{R}$ be some real valued function. Then, $\int f(x)\,dx$ is defined as $\int_{\mathbb{R}^{d_X}} f(x)\,dx$.
- Let $g_\beta : \mathbb{R}^{d_X} \to \mathbb{R}$ be a real valued function, which depends on some parameter $\beta \in B$. Then, $D_\beta g_\beta$ denotes the derivative with respect to β. The corresponding Hessian matrix is denoted by $\operatorname{Hess} g_\beta$.

Introduction

Arguably one of the most important contributions of mathematical statistics consists in the prediction of some variable Y, which is based on the realisations of some explanatory variable X. A powerful tool for predicting this so called dependent or response variable Y by the covariate or regressor X are regression models. It is difficult to date back the beginnings of such models, but they can be traced back even to Francis Galton and Karl Pearson, see Stanton (2001).
Nowadays, much attention is still concentrated on linear regression models. From a statistical point of view, Y and X are viewed as (possibly multivariate) random variables and the linear regression model can be written as

$$Y = \beta^t X + \varepsilon. \tag{1}$$

Here, β is called the regression coefficient and ε is an unobservable random variable, which is independent of X and fulfils $E[\varepsilon] = 0$. Sometimes, ε is assumed to be normally distributed. At a first glance, the linear regression model (1) seems to be attractive when analysing some given data set, since it is easy to implement and nicely interpretable. Nevertheless, the model relies on very restrictive assumptions such as additivity, homoscedasticity and sometimes normality of the error term ε and of course linearity of the relationship between Y and X. These problems regarding applicability of model (1) were already observed by Box and Cox (1964) and motivated them to introduce the parametric transformation model

$$\Lambda_{\theta_0}(Y) = \beta^t X + \varepsilon. \tag{2}$$

Here, the response variable Y is transformed by a transformation function Λ_{θ_0} before fitting it to the linear regression model. The unknown function Λ_{θ_0} is assumed to belong to some parametric class of strictly increasing functions $\{\Lambda_\theta : \theta \in \Theta\}$ for some finite dimensional parameter space Θ. Box and Cox (1964) presented a parametric class of power transformations, the famous Box-Cox-transformations, which contains as special cases the identity and the logarithm. Their transformation class was enlarged by Yeo and Johnson (2000), but there are meanwhile various classes of transformation functions, see Zellner and Revankar (1969), John and Draper (1980), Bickel and Doksum (1981) or Jones and Pewsey (2009) for further examples. An alternative motivation for transformation models is the perspective of first transforming the data in order to make somehow "simpler" procedures from nonparametric regression applicable.
Although the parametric transformation model generalizes the linear regression model by far, the basic assumption of a linear regression function remains. Moreover, this selection

problem carries over to the choice of the transformation function, since model (2) relies on the assumption that the experimenter chooses an appropriate transformation class. As a further extension, model (2) can be generalized by allowing nonparametric transformation, regression and variance functions h, g and σ^2. The resulting nonparametric transformation model

$$h(Y) = g(X) + \sigma(X)\varepsilon \tag{3}$$

with $\varepsilon \perp X, E[\varepsilon] = 0$ and $\mathrm{Var}(\varepsilon) = 1$ will be the central object of this thesis. Usually, h is assumed to be strictly increasing and some smoothness assumptions on h, g and σ are made.

There are many reasons for considering nonparametric and heteroscedastic transformation models. First, the analysis of a general model helps to understand the links between all of its components better. Additionally, some results, e.g., on identifiability, which will be explained below, can be carried over to simpler models. Second, hypothesis tests, which are based on the comparison of parametric and nonparametric estimators, can be constructed so that general models allow inferences to be drawn about the type of the relationship between Y and X. Two examples for such tests will be provided in Chapters 2 and 5. Third, previous knowledge or assumptions on the relationship between Y and X can make the application of parametric or homoscedastic models inappropriate.

Box and Cox (1964) applied their transformation functions to data on survival times of intoxicated animals and on the duration of worsted yarn before it gets broken. Transformation models are used frequently in such duration models, see Gørgens and Horowitz (1999) or Van den Berg (2001). John and Draper (1980) analysed the ability of expert inspectors in assessing the thickness of certain types of piping. They found that the usual Box-Cox transformations describe their data rather badly and adjusted them by taking the absolute value of Y and changing the sign afterwards. Carroll and Ruppert (1984) applied the Box-Cox transforms to spawner recruit and chemical reaction data. The Michaelis Menten equation (Michaelis and Menten (1913)) is often used in such contexts. Ruppert, Cressie, and Carroll (1989) considered the estimation of Michaelis-Menten parameters and pointed out that a wrong transformation of the model may lead to heteroscedastic errors. They also applied the Box-Cox transforms to study the reproduction of the sockeye salmon and some enzyme kinetics. Horowitz (2009) examined the influence of the economic activity on the duration of contract strikes and mentioned hedonic pricing as a further application. The latter was done by Wen, Bu, and Zhang (2013), who modelled the house prices in Hangzhou City. Another research field, where transformation models are applied and which is related to duration models, is the field of survival analysis. To mention only one application, Cheng, Wei, and Ying (1995) analysed the influence of a patient's age on his or her survival time. When considering transformation models as in (2) or (3), there are basically three categories, into which questions can be classified. The first group contains the more probability theoretical questions regarding solvability of the model or uniqueness of its components. Second, when faced with an underlying model probably every statistician is interested in estimating its components. Third, the price of nonparametric modelling often consists in a decreasing performance of the estimators of its components. Hence, it may be desirable to test whether some simple parametric model like (2) holds. Indeed, all

of the four main chapters 2–5 of this thesis can be classified into one of these categories.
One question, which belongs to the first group and which was already mentioned above, is that of identifiability of a model. A model is identified if its components are uniquely determined by the joint distribution of (Y, X). To illustrate that this uniqueness in general does not hold without further assumptions, let $\alpha > 0$ and $\beta \in \mathbb{R}$ be some constants. Multiplying both sides of (3) by α and adding β to both sides afterwards leads to

$$\alpha h(Y) + \beta = \alpha g(X) + \beta + \alpha\sigma(X)\varepsilon,$$

that is

$$\tilde{h}(Y) = \tilde{g}(X) + \tilde{\sigma}(X)\varepsilon \tag{4}$$

for $\tilde{h} = \alpha h + \beta, \tilde{g} = \alpha g + \beta, \tilde{\sigma} = \alpha\sigma$. Even if the transformation function is assumed to be strictly increasing, any triple h, g, σ in (3) leads to infinitely many other solutions $\tilde{h}, \tilde{g}, \tilde{\sigma}$ in (4). Various identification results in different models were provided, e.g., by Horowitz (1996), Ekeland, Heckman, and Nesheim (2004), Linton, Sperlich, and Van Keilegom (2008), Chiappori, Komunjer, and Kristensen (2015) and Vanhems and Van Keilegom (2019), see Chapter 3 for details. Most of these results show that it suffices for each of the corresponding transformation models to fix the parameters α and β from above to ensure identifiability of the model. Usually, this is done by so called location and scale constraints, e.g., like $h(0) = 0$ and $h(1) = 1$. Nevertheless, further assumptions are necessary to ensure identifiability in the context of heteroscedastic models, see Remark 3.4.1. In Chapter 3, identifiability of the nonparametric heteroscedastic model (3) will be proven. So far, such a general result has not been provided in the literature.
Several approaches of estimating the transformation function in various models have been discussed in the past. The fully parametric models mentioned above assume normality of the error ε and apply maximum likelihood estimators. Klaaßen, Kück, and Spindler (2017) analysed a fully parametric model in the context of high dimensional data. For some estimating approaches in models with a parametric regression function, but a nonparametric transformation function, see Horowitz (1996), Chen (2002), Zhou, Lin, and Johnson (2009) and Jochmans (2013). A summary was given by Horowitz (2009). There are only a few estimators in models with a parametric transformation function and a nonparametric regression function. Linton, Chen, Wang, and Härdle (1997) considered a model with an additive regression function and suggested to estimate the transformation parameter by an instrumental variable approach or a pseudo-likelihood method. Linton et al. (2008) used a profile likelihood approach and ideas on minimum distance estimators (see Chen, Linton, and Van Keilegom (2003), Chapter 5 of Koul (2002) or Section 3.2 of Van der Vaart and Wellner (1996)) to develop a "profile likelihood estimator" and a "mean square distance from independence estimator" in their seminal paper. Colling and Van Keilegom (2018) introduced a third estimator for the transformation parameter. See Section 1.3 for details on these estimators. Fully nonparametric, but homoscedastic transformation models have been treated by Chiappori et al. (2015) and Colling and Van Keilegom (2019), see Section 1.4 for details. Heteroscedastic, but semiparametric models have been considered by Zhou et al. (2009), Neumeyer, Noh, and Van Keilegom (2016) and Wang and Wang (2018). While Zhou et al. (2009) assumed the regression function to be linear and the variance function to

be known, Neumeyer et al. (2016) extended the results of Linton et al. (2008) to semiparametric transformation models with parametric transformation functions and nonparametric regression and variance functions. Wang and Wang (2018) considered a similar model to Zhou et al. (2009), but allowed censored data.

Every special case (e.g. the homoscedastic model) of (3) may induce the need to test for its validity. Hence, there are various different objectives to test for in the context of transformation models. There already exist a couple of hypothesis tests concerning parametric assumptions on the regression or transformation function or the significance of the components of the covariate X, that is, if all of the covariate's components are necessary to describe Y. Goodness of fit tests for the regression function were developed by Colling and Van Keilegom (2016, 2017) and Kloodt and Neumeyer (2019), while Allison, Hušková, and Meintanis (2018) and Kloodt and Neumeyer (2019) examined the significance of the components of the covariate X, see Section 2.2 for details. Tests for the hypothesis of a parametric transformation function were provided by Neumeyer et al. (2016), Hušková, Meintanis, Neumeyer, and Pretorius (2018), Hušková, Meintanis, and Pretorius (2019) and Szydłowski (2017), see the introduction of Chapter 5 for details.

This thesis is structured as follows. First, some essentials on kernel estimators, goodness of fit tests in semiparametric transformation models as well as some results on the estimation of the transformation function in semiparametric and nonparametric models are given. Then, a goodness of fit test for the regression function in a nonparametric and homoscedastic model is developed in Chapter 2. Chapter 3 addresses the issue of identifiability in the nonparametric and heteroscedastic model (3). The results obtained there are in turn used in Chapter 4 to construct estimators for the transformation function h in model (3). Moreover, uniform convergence results are presented. In Chapter 5, a hypothesis test for the null hypothesis of a parametric transformation function in the nonparametric and homoscedastic transformation model is given. The test is based on the ideas of Colling and Van Keilegom (2018). The asymptotic behaviour under the null hypothesis as well as under (local) alternatives is analysed. Furthermore, relevant hypotheses are considered. Finally, the results of this theses are summarized and discussed in Chapter 6 and possible ideas for future research are mentioned.

1

Essentials

This thesis treats several aspects of nonparametric transformation models. Especially, goodness of fit tests for parametric assumptions on the regression and transformation function are developed in Chapters 2 and 5, respectively. Therefore, a brief overview about goodness of fit tests in regression models as well as some previous results on parametric and nonparametric estimation of the transformation function are given in Sections 1.2, 1.3 and 1.4. Moreover, kernel estimators will be introduced in Section 1.1, since they will be used as the main tool for nonparametric estimation in this thesis.

1.1 Kernel Estimators

Kernel estimation is arguably one of the most frequently applied methods in nonparametric estimation. See the book of Wand and Jones (1995) for a well written examination. In this section, only a limited selection of results on kernel estimators, which are used in almost all of the following chapters, is presented. While doing so, the main framework will remain the same: Independent and identically distributed random pairs $(Y, X), (Y_1, X_1), ..., (Y_n, X_n)$ with joint distribution function $F_{Y,X}$ and density $f_{Y,X}$ are given. Y is assumed to be real valued, while X is assumed to be $\mathbb{R}^{d_X}$-valued for some $d_X \in \mathbb{N}$.
Mostly in this thesis, the notations from Chiappori et al. (2015) are adopted. In particular, let f_X denote the marginal density of X and define $p(y, x) = \int_{-\infty}^{y} f_{Y,X}(u, x)\, du$. Partial derivatives with respect to y or some component x_j are marked with a lower y and x_j, respectively, e.g., $p_y(y, x) = \frac{\partial}{\partial y} p(y, x)$. When considering f_X, the random variable in the index will be omitted sometimes, that is $f(x) = f_X(x), f_{x_j}(x) = f_{X_{x_j}}(x) = \frac{\partial}{\partial x_j} f_X(x)$ and the abbreviation $f_x = f_{x_1}$ is used.
To define the kernel estimators of these quantities, let $K : \mathbb{R} \to \mathbb{R}$ denote a kernel function, which means $\int K(x)\, dx = 1$ here, and let $\mathbf{K} : \mathbb{R}^{d_X} \to \mathbb{R}$ denote the corresponding product kernel on $\mathbb{R}^{d_X}$. Mostly, a continuous kernel with bounded support will be considered in this thesis, but together with some chapter specific assumptions these conditions will be listed in each of the chapters separately. Moreover, let $h_y \searrow 0$ and $h_x \searrow 0$ be some bandwidth sequences. Define

$$K_{h_y}(y) = \frac{1}{h_y} K\left(\frac{y}{h_y}\right), \quad \mathbf{K}_{h_x}(x) = \frac{1}{h_x^{d_X}} \mathbf{K}\left(\frac{x}{h_x}\right), \quad \mathcal{K}_{h_y}(y) = \int_{-\infty}^{y} K_{h_y}(u)\, du, \tag{1.1}$$

$$\hat{f}_X(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{K}_{h_x}(x - X_i), \quad \hat{f}_x(x) = \frac{1}{nh_x^{d_X+1}}\sum_{i=1}^{n} \frac{\partial}{\partial v_1}\mathbf{K}(v)\bigg|_{v=\frac{x-X_i}{h_x}}, \tag{1.2}$$

$$p(y,x) = \int_{-\infty}^{y} f_{Y,X}(u,x)\,du, \quad p_y(y,x) = f_{Y,X}(y,x), \quad p_x(y,x) = \int_{-\infty}^{y} \frac{\partial}{\partial x_1} f_{Y,X}(u,x)\,du, \tag{1.3}$$

$$\Phi(y,x) = \frac{p(y,x)}{f_X(x)}, \quad \Phi_y(y,x) = \frac{p_y(y,x)}{f_X(x)}, \quad \Phi_x(y,x) = \frac{p_x(y,x)}{f_X(x)} - \frac{p(y,x)f_x(x)}{f_X^2(x)} \tag{1.4}$$

and

$$\hat{p}(y,x) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{K}_{h_y}(y - Y_i)\mathbf{K}_{h_x}(x - X_i) \tag{1.5}$$

$$\hat{p}_y(y,x) = \frac{1}{n}\sum_{i=1}^{n} K_{h_y}(y - Y_i)\mathbf{K}_{h_x}(x - X_i) \tag{1.6}$$

$$\hat{p}_x(y,x) = \frac{1}{nh_x^{d_X+1}}\sum_{i=1}^{n} \mathcal{K}_{h_y}(y - Y_i)\frac{\partial}{\partial v_1}\mathbf{K}(v)\bigg|_{v=\frac{x-X_i}{h_x}}. \tag{1.7}$$

Φ from (1.4) is an alternative expression for the conditional distribution function of Y conditioned on X. In the following, it will be referred to the estimators in (1.2) and (1.5)–(1.7) as the kernel estimators for f_X, f_x, p, p_y and p_x. The index of the bandwidths h_x and h_y refers to the components which they are used for. h_x and h_y do not depend on specific values of $x \in \mathbb{R}^{d_X}, y \in \mathbb{R}$.
As for most estimators, the expected quadratic errors of these estimates can be divided into systematic and random errors. To handle the systematic error or bias of the estimators, higher order kernels are used (Wand and Jones (1995, p. 32)). When doing so, a kernel $K : \mathbb{R} \to \mathbb{R}$ is said to have order $q \in \mathbb{N}$, if

$$\int |x|^j K(x)\,dx = 0 \text{ for all } j = 1, ..., q-1 \quad \text{and} \quad \int x^q K(x)\,dx < \infty.$$

Basically, this together with integration by substitution and Taylor expansions will ensure negligibility of the systematic error.

Lemma 1.1.1 *Let K be of order $q \in \mathbb{N}$ and $h_y^q, h_x^q = o\big(n^{-\frac{1}{2}}\big)$. Further, let Y, X be real and $\mathbb{R}^{d_X}$-valued random variables with joint density $f_{Y,X}$, which is q-times partially continuously differentiable with bounded derivatives of order q. Then,*

$$E[\mathcal{K}_{h_y}(u - Y)\mathbf{K}_{h_x}(X - x)] = \int_{-\infty}^{u} f_{Y,X}(z,x)\,dz + o\Big(\frac{1}{\sqrt{n}}\Big) = p(u,x) + o\Big(\frac{1}{\sqrt{n}}\Big)$$

uniformly in $u \in \mathbb{R}, x \in \mathbb{R}^d$.

The proof is given in Section 1.6.1. Once tools for bounding the bias of an estimator have been introduced, the question of how to treat the random errors arises. Although dependent data is considered there, the ideas of Hansen (2008) will be applied for this purpose. Since his results have to be fit to the context of each chapter separately, the details are not presented in this Section. Nevertheless, note that under the conditions mentioned there Theorem 2 of Hansen (2008) already yields

$$\sup_{x\in\mathcal{K}} \left|\hat{f}_X(x) - E[\hat{f}_X(x)]\right| = \mathcal{O}_p\left(\sqrt{\frac{\log(n)}{nh^{d_X}}}\right)$$

for all compact sets $\mathcal{K} \in \mathbb{R}^{d_X}$. By the same reasoning as in the proof of Lemma 1.1.1, this can be extended to

$$\sup_{x\in\mathcal{K}} |\hat{f}_X(x) - f_X(x)| = \mathcal{O}_p\left(\sqrt{\frac{\log(n)}{nh^{d_X}}}\right).$$

The nonparametric estimators for the transformation function of Chiappori et al. (2015) and Colling and Van Keilegom (2019) are based on the idea of expressing the transformation function via Φ_x and Φ_y from equation (1.4). When plugging in the estimators from (1.2) and (1.5)–(1.7) into (1.4) to obtain estimators $\hat{\Phi}_x$ and $\hat{\Phi}_y$ for Φ_x and Φ_y, the uniform convergence results can be extended to $\hat{\Phi}_x$ and $\hat{\Phi}_y$ rather easily. The following Lemma was taken directly from Chiappori et al. (2015).

Lemma 1.1.2 *Let $a, b, \hat{a}, \hat{b} \in \mathbb{R}, b, \hat{b} \neq 0$. Then,*

$$\frac{\hat{a}}{\hat{b}} - \frac{a}{b} = \frac{1}{b}(\hat{a} - a) - \frac{a}{b^2}(\hat{b} - b) - \frac{\hat{b} - b}{\hat{b}b}\left(\hat{a} - a - \frac{a(\hat{b} - b)}{b}\right). \tag{1.8}$$

Replacing $\hat{a}$ and $\hat{b}$ by $\hat{p}$ and $\hat{f}_X$ leads to a uniform bound for the difference $|\hat{\Phi} - \Phi|$.

1.2 Goodness of Fit Tests in Mean Regression Models

The idea of justifying the application of a specific model by applying a corresponding goodness of fit test previously came up in the beginning of the twentieth century and has attracted more and more attention in the context of regression models since the early 1990s (González-Manteiga and Crujeiras (2013)). A huge variety of procedures testing, e.g., for a parametric regression function in the model

$$Y = g(X) + \varepsilon, \tag{1.9}$$

where X is $\mathbb{R}^{d_X}$-valued and Y and ε are real valued with $E[\varepsilon|X] \equiv 0$, can be found in the literature. A thorough review of such tests was given by González-Manteiga and Crujeiras (2013). They distinguished between smoothing based tests and tests that are based on empirical regression processes. Since an exhaustive presentation would go beyond the scope of this thesis, only some goodness of fit tests of both categories, which already have been extended to transformation models, are described in the following.

The approaches of Bierens (1982) and Stute (1997) were extended by Colling and Van Keilegom (2016) to the context of semiparametric transformation models and belong to the second class of goodness of fit tests. Consider a parametric class of regression functions $\{g_\beta : \beta \in B\}$ for some parameter space $B \subseteq \mathbb{R}^{d_B}$ with some $d_B \in \mathbb{N}$ and n independent and identically distributed observations $(Y_i, X_i), i = 1, .., n$, from the nonparametric regression model (1.9). Both of the mentioned papers made use of the observation that the parametric residuals $\varepsilon_i(\beta) := Y_i - g_\beta(X_i)$ fulfil $E[\varepsilon_i(\beta)|X] \equiv 0$, which means that model (1.9) is fulfilled with $g = g_\beta$, if and only if one has

$$E[\varepsilon_i(\beta)w(X_i, \gamma)] = 0$$

for an appropriate weight function w, an appropriate parameter space $\Gamma \subseteq \mathbb{R}^{d_\Gamma}$ with some $d_\Gamma \in \mathbb{N}$ and all $\gamma \in \Gamma$, but differ in the applied weight function. While Bierens (1982) used (as a special case)

$$w(x, \gamma) = \exp(ix^t\gamma)$$

and a compact multidimensional interval as a parameter space Γ, Stute (1997) applied

$$w(x, \gamma) = I_{\{x \leq \gamma\}}$$

for one dimensional x. Colling and Van Keilegom (2017) considered the generalization with multidimensional $x \in \mathbb{R}^{d_X}$, componentwise indicator functions and a parameter space $\Gamma = \mathbb{R}^{d_X}$ in a model with parametric transformations. Under the null hypothesis of $g = g_{\beta_0}$ for some $\beta_0 \in B$, Bierens (1982) and Stute (1997) showed weak convergence of the empirical process

$$R_n(\gamma) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Y_i - g_{\hat{\beta}}(X_i))w(X_i, \gamma), \quad \gamma \in \Gamma,$$

where $\hat{\beta}$ is an appropriate estimator of β_0, to some Gaussian processes for both choices of w and Γ. Afterwards, a Kolmogorov-Smirnov type (Stute (1997)) and a Cramér-von-Mises type (Bierens (1982)) test statistic were defined for testing the null hypothesis of a parametric regression function and the asymptotic distribution was derived from the weak convergence of the corresponding process.

The tests of Härdle and Mammen (1993) and Zheng (1996) are representatives of the smoothing based procedures. Both were extended to semiparametric transformation models by Kloodt and Neumeyer (2019). Consider again independent and identically distributed observations $(Y_i, X_i), i = 1, .., n$, from model (1.9). Let K and h_x be a kernel function and bandwidth sequence, respectively. Recall definition (1.1) and let $\{g_\beta : \beta \in B\}$ be a class of parametric regression functions to test for. The test of Zheng (1996) is based on the fact that

$$E\big[(Y - g_{\beta_0}(X))E[(Y - g_{\beta_0}(X))|X]f_X(X)\big] = E\big[E[(Y - g_{\beta_0}(X))|X]^2 f_X(X)\big]$$

is equal to zero if and only if $g \equiv g_{\beta_0}$ holds for some $\beta_0 \in B$. Here, f_X denotes the density function of X. The test statistic can be written as

$$V_n = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \mathbf{K}_{h_x}(X_i - X_j)(Y_i - g_{\hat{\beta}}(X_i))(Y_j - g_{\hat{\beta}}(X_j))$$

for some parametric estimator $\hat{\beta}$ of the minimizer $\beta_0 = \underset{\beta \in B}{\arg\min}\, E[(Y - g_\beta(X))^2]$.

To describe the test of Härdle and Mammen (1993), let

$$\hat{g}(x) = \frac{\sum_{i=1}^{n} \mathbf{K}_{h_x}(x - X_i)Y_i}{\sum_{i=1}^{n} \mathbf{K}_{h_x}(x - X_i)}$$

denote the Nadaraya Watson estimator of g and define for every $g_\beta, \beta \in B$, the smoothed version

$$\tilde{g}_\beta(x) = \frac{\sum_{i=1}^n \mathbf{K}_{h_x}(x - X_i) g_\beta(X_i)}{\sum_{i=1}^n \mathbf{K}_{h_x}(x - X_i)}.$$

The approach is based on a comparison of the nonparametric estimator $\hat{g}$ and the smoothed versions $\tilde{g}_\beta$ of the parametric regression functions g_β. To be precise, let w be an appropriate weight function and define the test statistic

$$T_n = nh^{\frac{d_X}{2}} \int \left(\hat{g}(x) - \tilde{g}_{\hat{\beta}}\right)^2 w(x)\, dx,$$

where $\hat{\beta}$ is an appropriate estimator of the minimizer $\beta_0 = \arg\min_{\beta \in B} E[(Y - g_\beta(X))^2]$. Under the null hypothesis of a parametric regression function, Härdle and Mammen (1993) proved weak convergence of the appropriately standardized test statistic T_n to some normally distributed random variable.

Although the test provided in Section 2.3 is based on the estimated conditional quantile function instead of the estimated regression function, there are many parallels between the test of Härdle and Mammen (1993) and the approach developed there. Especially, the parameters of the asymptotic normal distribution and the rather slow convergence of the test statistic to this distribution for finite sample sizes are similar.

1.3 Semiparametric Transformation Models

Since transformation models usually are applied to the data to avoid misspecification of the underlying model or to induce desired properties such as homoscedasticity, additivity or normality of the error terms (Box and Cox (1964)), it is worthwhile to generalize model (2) further by for example considering a nonparametric regression function g. The consequent model

$$\Lambda_{\theta_0}(Y) = g(X) + \varepsilon \tag{1.10}$$

for independent ε and X, for some true transformation parameter $\theta_0 \in \Theta$, some parameter space $\Theta \subseteq \mathbb{R}^{d_\Theta}$ with some $d_\Theta \in \mathbb{N}$ and a class of transformation functions $\{\Lambda_\theta : \theta \in \Theta\}$ has been studied extensively in the literature. Linton et al. (2008) introduced two estimating approaches, which will be described later in this section, for θ_0 in their seminal paper. These ideas were revisited among others by Neumeyer et al. (2016) and Vanhems and Van Keilegom (2019) to extend the approaches to heteroscedastic errors and endogenous regressors, respectively. Recently, Colling and Van Keilegom (2018) developed two further estimators by minimizing appropriate distances between the parametric class of transformation functions and the nonparametric estimator by Colling and Van Keilegom (2019). In the following, some of the approaches mentioned above are described briefly. Note that the class of transformation functions $\{\Lambda_\theta : \theta \in \Theta\}$ has to fulfil some conditions to ensure uniqueness of the true transformation parameter. This issue will be discussed later in more detail.

The Profile Likelihood Estimator

Linton et al. (2008) introduced two methods for estimating the true transformation parameter θ_0 in model (1.10), the profile likelihood estimator and the mean square distance from independence estimator. While the latter one is discussed in the subsection below, the profile likelihood method will be explained in the following.
Let $\theta \in \Theta$ and define $g_\theta(\cdot) = E[\Lambda_\theta(Y)|X = \cdot]$ as well as $\varepsilon(\theta) = \Lambda_\theta(Y) - g_\theta(X)$. Denote the conditional distribution and density function of Y conditional on X by $F_{Y|X}$ and $f_{Y|X}$, respectively. Then, the conditional distribution function can be written as

$$F_{Y|X}(y|x) = P(\Lambda_\theta(Y) \leq \Lambda_\theta(y)|X = x) = P(\varepsilon(\theta) \leq \Lambda_\theta(y) - g_\theta(x)|X = x).$$

For the true transformation parameter $\theta = \theta_0$, this results due to the independence of ε and X in

$$F_{Y|X}(y|x) = F_\varepsilon(\Lambda_\theta(y) - g_\theta(x))$$

and

$$f_{Y|X}(y|x) = f_\varepsilon(\Lambda_\theta(y) - g_\theta(x))\Lambda_\theta'(y),$$

where F_ε and f_ε are the distribution function and density of $\varepsilon = \varepsilon(\theta_0)$ and Λ_θ' denotes the derivative of Λ_θ with respect to y. This can be used to apply techniques from maximum likelihood estimation. Let $(Y_i, X_i), i = 1, ..., n$, be independent and identically distributed observations from model (1.10). Then, the profile likelihood estimator $\hat{\theta}_{PL}$ is defined as

$$\hat{\theta}_{PL} = \underset{\theta \in \theta}{\arg\max} \sum_{i=1}^{n} \big(\log(\hat{f}_{\varepsilon(\theta)}(\Lambda_\theta(Y_i) - \hat{g}_\theta(X_i))) + \log(\Lambda_\theta'(Y_i))\big),$$

where $f_{\varepsilon(\theta)}$ denotes the density of $\varepsilon(\theta)$ and $\hat{f}_{\varepsilon(\theta)}$ and $\hat{g}_\theta$ are some nonparametric estimators for $f_{\varepsilon(\theta)}$ and g_θ. By using the Kullback Leibler divergence similarly to Neumeyer et al. (2016), it can be shown that the true transformation parameter θ_0 minimizes the function

$$\theta \mapsto E[\log(f_{\varepsilon(\theta)}(\Lambda_\theta(Y) - g_\theta(X))) + \log(\Lambda_\theta'(Y))],$$

so that $\hat{\theta}_{PL}$ is a meaningful estimate of θ_0. Linton et al. (2008) were able to show asymptotic normality of $\sqrt{n}(\hat{\theta}_{PL} - \theta_0)$ under the assumptions mentioned in their paper.

The Mean Square Distance from Independence Estimator

The second estimator introduced by Linton et al. (2008) was the mean square distance from independence estimator (MDE). Let g_θ and $\varepsilon(\theta)$ be defined as in the subsection above and let $\hat{\varepsilon}(\theta)$ be some estimator of $\varepsilon(\theta)$. The idea behind the MDE is that $\varepsilon(\theta)$ is independent of X if and only if $\theta = \theta_0$. Hence, the joint distribution function of X and $\varepsilon(\theta)$ can be written as the product of the marginal distribution functions if and only if $\theta = \theta_0$. To define the estimator, let $(Y_i, X_i), i = 1, ..., n$, be independent and identically distributed observations from (1.10) and define the empirical distribution functions

$$\hat{F}_X(x) = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i \leq x\}}, \quad \hat{F}_{\varepsilon(\theta)}(e) = \frac{1}{n}\sum_{i=1}^{n} I_{\{\hat{\varepsilon}_i(\theta) \leq e\}}, \quad \hat{F}_{X,\varepsilon(\theta)}(x, e) = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i \leq x\}} I_{\{\hat{\varepsilon}_i(\theta) \leq e\}}.$$

Then, the MDE is defined as

$$\hat{\theta}_{MD} = \operatorname*{arg\,min}_{\theta \in \theta} \int \left(\hat{F}_X(x) \hat{F}_{\varepsilon(\theta)}(e) - \hat{F}_{X,\varepsilon(\theta)}(x,e) \right)^2 d\mu(x,e)$$

for some appropriate measure μ. Under the assumptions mentioned there, it was shown in Linton et al. (2008) that $\sqrt{n}(\hat{\theta}_{MD} - \theta_0)$ is asymptotically normal. In their simulations, Linton et al. (2008) observed that $\hat{\theta}_{PL}$ seems to outperform $\hat{\theta}_{MD}$.
An estimator, which is related to the MDE, is used later in Chapter 4 to estimate some component of a nonprametric estimator for the transformation h in model (3) with heteroscedastic errors. Roughly speaking, the reason for adapting the MDE instead of the PLE approach there is that the estimation of f_ε requires good estimates of ε on the whole set of real numbers, while the exact value of ε does not influence the indicator function $I_{\{\varepsilon \leq e\}}$, as long as it exceeds some boundary. This issue will be discussed in detail in Chapter 4.

Comparing the Transformation Class to a Nonparametric Estimator

The remaining two approaches for estimating θ_0 in model (1.10) that were mentioned above, were developed by Colling and Van Keilegom (2018). Both of the procedures are based on a comparison of the parametric transformation class and a nonparametic estimator of the transformation function. Colling and Van Keilegom (2019) considered the nonparametric model

$$h(Y) = g(X) + \varepsilon, \tag{1.11}$$

where h is assumed to be strictly increasing, $E[\varepsilon] = 0$ holds and X and ε are independent. Their estimator will be denoted by $\hat{h}$ in the following. Note that the validity of the model in (1.10) is unaffected by linear transforms, that is, the model still holds when Λ_{θ_0}, g and ε are replaced by

$$\tilde{\Lambda}(y) = a\Lambda_{\theta_0}(y) + b, \quad \tilde{g}(x) = ag(x) + b \quad \text{and} \quad \tilde{\varepsilon} = a\varepsilon$$

for any constants $a > 0, b \in \mathbb{R}$. Therefore, so called identification constraints are necessary to fix a and b and to induce uniqueness of the true transformation function. The nonparametric estimator of Colling and Van Keilegom (2018) fulfils $\hat{h}(0) = 0$ and $\hat{h}(1) = 1$. To make the nonparametric estimator comparable to the parametric class of transformation functions $\{\Lambda_\theta : \theta \in \Theta\}$, the same identification constraints have to be applied. Thus, some distance between

$$y \mapsto (\Lambda_\theta(1) - \Lambda_\theta(0))\hat{h}(y) + \Lambda_\theta(0)$$

and Λ_θ is used to construct an estimator $\hat{\theta}$ for θ_0, since both functions attain the same values at $y = 0$ and $y = 1$ then. To be precise, a quadratic distance is used and $\hat{\theta}$ is defined as

$$\hat{\theta} = \operatorname*{arg\,min}_{\theta \in \Theta} \sum_{i=1}^{n} \left(\hat{h}(Y_i)(\Lambda_\theta(1) - \Lambda_\theta(0)) + \Lambda_\theta(0) - \Lambda_\theta(Y_i) \right)^2. \tag{1.12}$$

The factors $c_1(\theta) := (\Lambda_\theta(1) - \Lambda_\theta(0))$ and $c_2(\theta) := \Lambda_\theta(0)$, that are necessary to fit the underlying identification constraints of $\hat{h}$ to those of $\{\Lambda_\theta : \theta \in \Theta\}$, are uniquely determined.

Nevertheless, it might be sensible to minimize in (1.12) with respect to c_1 and c_2 as well. The corresponding estimator can be written as

$$\tilde{\theta} = \operatorname*{arg\,min}_{\theta\in\Theta, c_1\in\mathbb{R}^+, c_2\in\mathbb{R}} \sum_{i=1}^{n} \big(\hat{h}(Y_i)c_1 + c_2 - \Lambda_\theta(Y_i)\big)^2.$$

Indeed, Colling and Van Keilegom (2018) showed that $\tilde{\theta}$ outperforms $\hat{\theta}$ in simulations. Additionally, they proved weak convergence of $\sqrt{n}(\hat{\theta}-\theta_0)$ and $\sqrt{n}(\tilde{\theta}-\theta_0)$ to some normally distributed random variables.

A goodness of fit test for a parametric transformation function, which makes use of the idea to compare a nonparametric estimator of a transformation function to a parametric transformation class, will be presented in Chapter 5.

1.4 Nonparametric Transformation Models

In this section, previous results on the nonparametric estimation of the transformation function are presented. Compared to the previous section, this is again a further step towards modelling the relationship between two random variables as flexible as possible. While Chapter 3 will allow heteroscedasticity, the homoscedastic model (1.11) will be considered here. As in the previous section, any triple (h, g, F_ε), where F_ε denotes the distribution function of ε, can only be unique up to linear transformations, which again leads to the question of identifiability of model (1.11). In the case of a linear regression function g, Horowitz (1996) developed a single index approach to identify and estimate the transformation function h. Chiappori et al. (2015) and Colling and Van Keilegom (2019) applied related methods, but estimated h in the general framework of a nonparametric g.

The basic idea of all of three approaches was to write the conditional distribution function $F_{Y|X}$ of Y conditioned on X as

$$F_{Y|X}(y|x) = F_\varepsilon(h(y) - g(x)).$$

The conditional distribution $F_{Y|X}$ can be alternatively expressed as Φ from (1.4). Let f_ε be the density of ε and denote the derivative of h by h'. Then, the derivatives of $F_{Y|X}$ with respect to y and some component x_i of x can be written as

$$\frac{\partial}{\partial y} F_{Y|X}(y|x) = f_\varepsilon(h(y) - g(x))h'(y) \quad \text{and} \quad \frac{\partial}{\partial x_i} F_{Y|X}(y|x) = -f_\varepsilon(h(y) - g(x)) \frac{\partial}{\partial x_i} g(x).$$

In the following, the derivative with respect to x_1 is considered w.l.o.g. Consequently, the quotient of both derivatives can be written as

$$\lambda(y|x) := \frac{\frac{\partial F_{Y|X}(y|x)}{\partial y}}{\frac{\partial F_{Y|X}(y|x)}{\partial x_1}} = -\frac{1}{\frac{\partial}{\partial x_1} g(x)} h'(y). \tag{1.13}$$

Note that the equation above only holds if

$$f_\varepsilon(h(y) - g(x)) \neq 0 \quad \text{and} \quad \frac{\partial}{\partial x_1} g(x) \neq 0. \tag{1.14}$$

Horowitz (1996), Chiappori et al. (2015) and Colling and Van Keilegom (2019) ensured validity of (1.14) by appropriate assumptions, but this issue is not discussed further here. Integrating equation (1.13) with respect to y leads to

$$\int_{y_0}^{y} \lambda(u|x)\,du = -\frac{1}{\frac{\partial}{\partial x_1} g(x)}(h(y) - h(y_0))$$

for every $y_0 \in \mathbb{R}$. Several kinds of identification constraints are conceivable to make any solution h to this equation unique. Chiappori et al. (2015) and Colling and Van Keilegom (2019) required for example

$$h(0) = 0 \quad \text{and} \quad h(1) = 1,$$

which leads to

$$h(y) = \frac{\int_0^y \lambda(u|x)\,du}{\int_0^1 \lambda(u|x)\,du}.$$

Although the right hand side does not depend on x, the performance of the estimator defined later increases when integrating with respect to x. To make this precise, let v be some weight function with $\int v(x)\,dx = 1$. Then, it holds that

$$h(y) = \underset{q \in \mathbb{R}}{\arg\min} \int v(x) \left(\frac{\int_0^y \lambda(u|x)\,du}{\int_0^1 \lambda(u|x)\,du} - q \right)^2 dx$$

as well as

$$h(y) = \underset{q \in \mathbb{R}}{\arg\min} \int v(x) \left| \frac{\int_0^y \lambda(u|x)\,du}{\int_0^1 \lambda(u|x)\,du} - q \right| dx. \tag{1.15}$$

Applying the square loss as in the first equation results in

$$h(y) = \int v(x) \frac{\int_0^y \lambda(u|x)\,du}{\int_0^1 \lambda(u|x)\,du}\,dx, \tag{1.16}$$

while applying the absolute loss leads to some kind of median.
Let $(Y_i, X_i), i = 1, ..., n$, be independent and identically distributed observations of (1.11). Then, an estimator for the transformation function h can be obtained by first inserting the estimators given in (1.2) and (1.5)–(1.7) into (1.4) and (1.13) to get estimators $\hat{\Phi}_y, \hat{\Phi}_x$ and $\hat{\lambda}$ for Φ_y, Φ_x and λ and by plugging in these estimators into (1.15) and (1.16) afterwards. Actually, Chiappori et al. (2015) and Colling and Van Keilegom (2019) used a smoothed version

$$\hat{h}(y) = \underset{q \in \mathbb{R}}{\arg\min} \int v(x) \left(\frac{\hat{s}_1(y,x)}{\hat{s}_1(1,x)} - q \right) \left(2L_b\left(\frac{\hat{s}_1(y,x)}{\hat{s}_1(1,x)} - q \right) - 1 \right) dx$$

instead of (1.15), where $\hat{s}_1(y,x)$ is defined as $\int_0^y \hat{\lambda}(u|x)\,du$, $b = b_n$ is some bandwidth sequence and $L_b(\cdot) = L\left(\frac{\cdot}{b}\right)$ for an appropriate distribution function L.
An equation similar to (1.16) is used in Chapter 4 to construct an estimator for h in the context of heteroscedastic errors.

Basing the Estimator on Pretransformed Data

Colling and Van Keilegom (2019) observed that the estimator of Chiappori et al. (2015) might perform badly if the distribution of Y is somehow inappropriate. For example, if the distribution of Y is very asymmetric or highly skewed, it seems that the procedure works rather badly. To overcome this problem, Colling and Van Keilegom (2019) first applied a pretransformation to the data. To make this precise, consider again independent and identically distributed random variables $(Y_i, X_i), i = 1, ..., n$, fulfilling model (1.11). Then, denote the distribution function of Y by F_Y and define

$$\mathcal{T}(y) = \frac{F_Y(y) - F_Y(0)}{F_Y(1) - F_Y(0)} \quad \text{and} \quad U_i = \mathcal{T}(Y_i).$$

For $Q := h \circ \mathcal{T}^{-1}$, model (1.11) can be expressed as

$$Q(U) = g(X) + \varepsilon.$$

After replacing U_i by the estimates

$$\hat{U}_i = \hat{\mathcal{T}}(Y_i) \quad \text{with} \quad \hat{\mathcal{T}}(y) = \frac{\hat{F}_Y(y) - \hat{F}_Y(0)}{\hat{F}_Y(1) - \hat{F}_Y(0)},$$

where $\hat{F}_Y$ denotes the empirical distribution function of $Y_1, ..., Y_n$, an estimator

$$\hat{Q}(u) = \arg\min_{q \in \mathbb{R}} \int v(x) \left(\frac{\tilde{s}_1(u,x)}{\tilde{s}_1(1,x)} - q \right) \left(2L_b \left(\frac{\tilde{s}_1(u,x)}{\tilde{s}_1(1,x)} - q \right) - 1 \right) dx, \tag{1.17}$$

Q can be obtained as described above. Here $\tilde{s}_1(y,x)$ is defined as $\int_0^u \tilde{\lambda}(z|x)\, dz$ and

$$\tilde{\lambda}(u|x) := \frac{\frac{\partial \hat{F}_{U|X}(u|x)}{\partial u}}{\frac{\partial \hat{F}_{U|X}(u|x)}{\partial x_1}}.$$

The corresponding estimator for h is $\hat{h}(y) = \hat{Q}(\hat{\mathcal{T}}(y))$. Using the notations from Section 1.1, define (compare Colling and Van Keilegom (2019))

$$D_{p,0}(u,x) = \frac{\Phi_u(x,u) f_x(x)}{\Phi_1^2(u,x) f^2(x)}, \qquad D_{p,u}(u,x) = \frac{1}{f(x)\Phi_1(u,x)},$$

$$D_{p,1}(u,x) = -\frac{\Phi_u(u,x)}{f(x)\Phi_1^2(u,x)}, \qquad D_{f,0}(u,x) = -\frac{\Phi_u(u,x)\Phi(u,x) f_x(x)}{\Phi_1^2(u,x) f^2(x)},$$

$$D_{f,1}(u,x) = \frac{\Phi_u(u,x)\Phi(u,x)}{\Phi_1^2(u,x) f(x)}.$$

Moreover, define

$$s_1(u,x) := \int_0^u \frac{\frac{\partial F_{U|X}(r|x)}{\partial u}}{\frac{\partial F_{U|X}(r|x)}{\partial x_1}}\, dr, \quad \tilde{v}_1(u_0,x) = \frac{v(x)}{s_1(u_0,x)}, \quad \tilde{v}_2(u_0,x) = \frac{v(x) s_1(u_0,x)}{s_1(1,x)^2}$$

and (for $\tilde{v} \in \{\tilde{v}_1, \tilde{v}_2\}$)

$$\delta_j^{\tilde{v}}(u_0,u) = \int_{\max(0,U_j)}^{\max(u,U_j)} \left(\tilde{v}(u_0,X_j) D_{p,0}(r,X_j) - \frac{\partial}{\partial x_1} \left(\tilde{v}(u_0,x) D_{p,1}(r,x) \right) \Big|_{x=X_j} \right) dr$$

$$
\begin{aligned}
&+\int_0^u \left(\tilde{v}(u_0,X_j)D_{f,0}(r,X_j)-\frac{\partial}{\partial x_1}\big(\tilde{v}(u_0,X_j)D_{f,1}(r,x)\big)\Big|_{x=X_j}\right)dr\\
&+\big(I_{\{U_j\le u\}}-I_{\{U_j\le 0\}}\big)\tilde{v}(u_0,X_j)D_{p,u}(U_j,X_j)\\
&+\int_0^u \left(\frac{I_{\{U_j\le u\}}-I_{\{U_j\le 0\}}}{F_U(1)-F_U(0)}-r\right)\\
&\quad\int_{\mathcal{X}}\left(\left(\tilde{v}(u_0,x)D_{p,0}(r,x)+\frac{\partial}{\partial x_1}\big(\tilde{v}(u_0,x)D_{p,1}(r,x)\big)\right)\right.\\
&\quad\left. f_{U,X}(r,x)+\tilde{v}(u_0,x)D_{p,u}(r,x)\frac{\partial}{\partial r}f_{U,X}(r,x)\right)dx\,dr\\
&-\left(\frac{I_{\{U_j\le 1\}}-I_{\{U_j\le 0\}}}{F_U(1)-F_U(0)}-1\right)\int_0^u r\int_{\mathcal{X}}\left(\tilde{v}(u_0,x)D_{p,0}(r,x)\right.\\
&\quad\left.-\tilde{v}(u_0,x)\frac{\partial}{\partial r}D_{p,u}(r,x)+\frac{\partial}{\partial x_1}\big(\tilde{v}(u_0,x)D_{p,1}(r,x)\big)\right)f_{U,X}(r,x)\,dx\,dr\\
&\quad-\left(\frac{I_{\{U_j\le 1\}}-I_{\{U_j\le 0\}}}{F_U(1)-F_U(0)}-1\right)u\int_{\mathcal{X}}\tilde{v}(u_0,x)D_{p,u}(u,x)f_{U,X}(u,x)\,dx.
\end{aligned}
$$

Then, Colling and Van Keilegom (2019) proved for the estimator in (1.17) as well as for another estimator, which is based on similar thoughts as in (1.16),

$$
\hat{h}(y)-h(y)=\frac{1}{n}\sum_{i=1}^n\psi(U_i,X_i,\mathcal{T}(y))+o_p\left(\frac{1}{\sqrt{n}}\right)
$$

for

$$
\begin{aligned}
\psi(U_j,X_j,u)=\delta_j^{\tilde{v}_1}(1,u)-\delta_j^{\tilde{v}_2}(u,1)+\frac{Q'(u)}{F_U(1)-F_U(0)}\big(I_{\{U_j\le u\}}-I_{\{U_j\le 0\}}-F_U(u)+F_U(0)\big)\\
-Q'(u)\frac{F_U(u)-F_U(0)}{(F_U(1)-F_U(0))^2}\big(I_{\{U_j\le 1\}}-I_{\{U_j\le 0\}}-F_U(1)+F_U(0)\big).
\end{aligned}
$$

The approaches developed in Chapters 2 and 5 can be applied for both of the estimators of Chiappori et al. (2015) and Colling and Van Keilegom (2019), but the latter one is used in the simulation studies there.

1.5 Miscellaneous

Finally, a technical Lemma is presented, which given any sequence of random variables $Z_n=o_p(1)$, yields the existence of a deterministic sequence $\delta_n\searrow 0$, such that $Z_n=o_p(\delta_n)$. This Lemma will be applied in Sections 2.8.7 and 4.6.3.

Lemma 1.5.1 *Let $(Z_n)_{n\in\mathbb{N}}$ be a sequence of random variables such that $Z_n=o_p(1)$. Then, there exists a deterministic and monotonic null sequence $(\delta_n)_{n\in\mathbb{N}}$ such that*

$$
Z_n=o_p(\delta_n).
$$

The proof is given in Section 1.6.2.

1.6 Proofs

Here, the proofs of the lemmas stated in Sections 1.1 and 1.5 are given.

1.6.1 Proof of Lemma 1.1.1

This lemma can be proven by basic calculations. More precisely, recall that $\mathbf{K}$ is a kernel of order q and write

$$\begin{aligned}
E[\mathcal{K}_{h_y}(u-Y)\mathbf{K}_{h_x}(X-x)] &= \int\int \mathcal{K}_{h_y}(u-z)\mathbf{K}_{h_y}(v-x)f_{Y,X}(z,v)\,dz\,dv \\
&= \int\int\int_{-\infty}^{\frac{u-z}{h_y}} K(w)\,dw\,\mathbf{K}_{h_y}(v-x)f_{Y,X}(z,v)\,dz\,dv \\
&= \int\int\int_{-\infty}^{u-h_yw} f_{Y,X}(z,v)\,dz\,K(w)\,dw\,\mathbf{K}_{h_y}(v-x)\,dv \\
&= \int\int p(u-h_yw,v)\,K(w)\,dw\,\mathbf{K}_{h_y}(v-x)\,dv \\
&= \int p(u,v)\,\mathbf{K}_{h_y}(v-x)\,dv + \mathcal{O}\big(h_y^q\big) \\
&= p(u,x) + \mathcal{O}\big(h_y^q + h_x^q\big) \\
&= p(u,x) + o\bigg(\frac{1}{\sqrt{n}}\bigg)
\end{aligned}$$

by using a Taylor expansion of p.

1.6.2 Proof of Lemma 1.5.1

Let $(\varepsilon_m)_{m\in\mathbb{N}}, (\tau_i)_{i\in\mathbb{N}}$ be decreasing null sequences. Define

$$\tilde{m}_{N,\tau} := \sup\Big\{m\in\mathbb{N} : \sup_{n\geq N} P(|Z_n|>\varepsilon_m)\leq\tau\Big\},$$

where

$$\begin{aligned}
\tilde{m}_{N,\tau} &= \infty, && \text{if } \sup_{n\geq N} P(|Z_n|>\varepsilon_m)\leq\tau \text{ for all } m\in\mathbb{N} \\
\tilde{m}_{N,\tau} &= 0, && \text{if } \sup_{n\geq N} P(|Z_n|>\varepsilon_m)>\tau \text{ for all } m\in\mathbb{N}.
\end{aligned}$$

If $\tilde{m}_{N,\tau}<\infty$ define $m_{N,\tau}=\tilde{m}_{N,\tau}$. If the case $\tilde{m}_{N,\tau}=\infty$ occurs for some $N\in\mathbb{N}$ (and consequently for all $\tilde{N}\geq N$ as well), define $m_{N,\tau}$ such that $(m_{N,\tau})_{N\in\mathbb{N}}$ is a monotonic sequence in $\mathbb{N}$ converging to ∞. Therefore, the sequence $(\varepsilon_{m_{N,\tau}})_{N\in\mathbb{N}}$ with $\varepsilon_0=\varepsilon_1$ is monotonically decreasing and converging to 0 for all $\tau\in(0,1)$.
Moreover, define recursively

$$\begin{aligned}
k_1 &= 0 \\
\tilde{\delta}_1 &= \varepsilon_{m_{1,\tau_1}}
\end{aligned}$$

$$k_{i+1} = \min\left\{k \geq k_i + 1 : \varepsilon_{m_{k,\tau_{i+1}}} \leq \frac{\tilde{\delta}_i}{2}\right\}, \qquad i \geq 1$$

$$\tilde{\delta}_{i+1} = \varepsilon_{m_{k_{i+1},\tau_{i+1}}} \qquad i \geq 1.$$

An appropriate sequence $(\delta_n)_{n\in\mathbb{N}}$ can be defined via

$$(\delta_n)_{n\in\mathbb{N}} = \Big(\underbrace{\sqrt{\tilde{\delta}_1}, ..., \sqrt{\tilde{\delta}_1}}_{(k_2-k_1)-times}, \underbrace{\sqrt{\tilde{\delta}_2}, ..., \sqrt{\tilde{\delta}_2}}_{(k_3-k_2)-times}, ...\Big).$$

Then, $\delta_n \to 0$ by construction and for all $C > 0$ one has

$$\begin{aligned}
\limsup_{n\to\infty} P\left(\frac{|Z_n|}{\delta_n} > C\right) &\leq \limsup_{n\to\infty} P(|Z_n| > \delta_n^2) \\
&= \lim_{N\to\infty} \sup_{n\geq N} P(|Z_n| > \delta_n^2) \\
&\leq \lim_{j\to\infty} \sup_{i\geq j} \sup_{k\geq k_i} P\big(|Z_k| > \varepsilon_{m_{k_i,\tau_i}}\big) \\
&\leq \lim_{j\to\infty} \sup_{i\geq j} \tau_i \\
&= \lim_{j\to\infty} \tau_j \\
&= 0,
\end{aligned}$$

that is $Z_n = o_p(\delta_n)$.

2

Testing for a Parametric Regression Function in Nonparametric Transformation Models - A Quantile Approach

Let $\mathcal{G}_B = \{g_\beta : \beta \in B\}$ be a class of regression functions indexed by some finite dimensional regression parameter $\beta \in B \subseteq \mathbb{R}^{d_B}$. Consider the simple regression model

$$Y = g(X) + \varepsilon \tag{2.1}$$

with $E[\varepsilon] = 0$ and ε independent of X. While there is a large variety of goodness of fit tests for the null hypothesis of g belonging to $\mathcal{G}_B$

$$H_0 : \quad g \in \mathcal{G}_B, \tag{2.2}$$

so far only a handful of them have been extended to the semiparametric transformation case and, to the author's knowledge, none of those has been extended to nonparametric transformation models.
In this chapter, methods known from the estimation of conditional quantiles are used to develop a test for the hypothesis of the conditional mean fulfilling (2.2). Therefore, first a brief overview of some nonparametric estimation techniques for conditional quantile functions and of some tests, which already have been generalized to semiparametric transformation models, is given in Sections 2.1 and 2.2, respectively. Afterwards, a new testing approach is presented in Section 2.3 and this test is extended to nonparametric transformation models in Section 2.4. Some thoughts on the asymptotic behaviour of the provided test are postponed to Section 2.5.

2.1 Nonparametric Conditional Quantile Estimation

There is a huge variety of literature concerning the issue of estimating the quantiles of a real valued random variable Y given a (possibly multidimensional) covariate X. Before a

brief insight into some of the methods is provided recall the definitions of Section 1.1. Just like there, assume that independent and identically distributed observations $(Y_i, X_i), i = 1, ..., n$, of a joint distribution $P^{(Y,X)}$ are given. An overview not only of the nonparametric estimation of conditional quantiles, but also of some hypothesis tests for model assumptions can be found in the dissertation of Guhlich (2013).
The arguably most common approach uses the so called "check-function"

$$\rho_\tau(u) = u(\tau - I_{\{u<0\}}) \tag{2.3}$$

to estimate the conditional τ-quantile of the distribution of Y conditioned on $X = x$, for example by

$$\arg\min_{q\in\mathbb{R}} \frac{1}{n}\sum_{i=1}^{n} \rho_\tau(Y_i - q)\mathbf{K}_{h_x}(x - X_i).$$

This approach is motivated by the fact that the true quantile can be expressed as the minimizer

$$\arg\min_{q\in\mathbb{R}} E[\rho_\tau(Y_1 - q)|X_1 = x].$$

See the book of Koenker (2005) for a detailed examination of this estimator and several adjustments in various contexts. Some basic results had already been provided by Stone (1977) and Chaudhuri (1991). The last papers also mentioned local polynomial extensions, which are also considered by Yu and Jones (1997). Horowitz and Lee (2005, 2007) introduced a procedure based on instrumental variables to estimate the conditional quantile function nonparametrically. Mu and He (2007) applied check-functions to estimate a transformation parameter in parametric transformation models on the one hand and a goodness of fit test for the model itself on the other hand.
A second type of quantile estimators can be classified as the inverting estimators. There, the basic idea is to estimate the conditional distribution function appropriately and to invert this estimator afterwards at some level $\tau \in (0, 1)$. Since nonparametric estimation of the conditional distribution function itself is a topic of large interest and thus there exist various approaches for such an estimation, these approaches provide various methods to estimate quantiles of the conditional distribution function as well. See for example Hall, Wolff, and Yao (1999) for some ideas on estimating the conditional distribution function. One class of such estimators for the conditional distribution function are the Nadaraya-Watson-type estimators referring to the papers of Nadaraya (1964) and Watson (1964). Some convergence results can be found in the paper of Devroye (1981). While Hall et al. (1999) provided a monotonically growing estimator of the conditional distribution function, Dette and Volgushev (2008) used related techniques to obtain non crossing estimators of the conditional quantile curves, that is, a with respect to τ monotonically growing quantile function. In this chapter, a smoothed version of the Nadaraya-Watson-type estimator, which was treated for example by Hansen (2004) and which was also applied by Chiappori et al. (2015), is used. With (1.1) the estimator for the conditional distribution function can be written similarly to (1.4) as

$$\hat{\Phi}(y, x) = \hat{F}_{Y|X}(y|x) = \frac{\sum_{i=1}^{n} \mathcal{K}_{h_y}(y - Y_i)\mathbf{K}_{h_x}(x - X_i)}{\sum_{i=1}^{n} \mathbf{K}_{h_x}(x - X_i)}.$$

Although $\hat{\Phi}$ and $\hat{F}_{Y|X}$ denote the same quantity, the latter notation is used throughout this chapter. The corresponding estimator of the quantile function can be written as

$$\hat{F}^{-1}_{Y|X}(\tau|x) = \min\{y \in \mathbb{R} : \hat{F}_{Y|X}(y|x) \geq \tau\}. \tag{2.4}$$

It is conjectured that after some adjustments the theory of this chapter can be applied to other estimating approaches as well. Nevertheless, the usage of (2.4) is accompanied with some synergy effects reducing the complexity of the (anyway quite technical) proofs of this section, especially when considering transformation models in Section 2.4.

2.2 Previous Tests in Semiparametric Models

Here, a small insight into model specification testing in transformation models is given. Note that many of those tests are strongly linked to the estimation approaches presented in Sections 1.3 and 1.4. Up to now, there is no test which allows nonparametric estimation of the transformation and regression functions at the same time. Therefore, only tests in the semiparametric model

$$\Lambda_{\theta_0}(Y) = g(X) + \varepsilon, \tag{2.5}$$

where $\{\Lambda_\theta : \theta \in \Theta\}$ is a class of transformation functions indexed by a finite dimensional transformation parameter θ and θ_0 denotes the true transformation parameter, are mentioned here. Specification tests in models like (2.5) in general aim to justify some reduction of the model complexity, which may result in faster or more precise estimators of the model components. With respect to the regression function g such a reduction may consist in a reduction of the dimension of the covariate or even in a parametric assumption.

Allison et al. (2018) and Kloodt and Neumeyer (2019) provided tests for the significance of components of the covariate X in semiparametric transformation models, that is, if all of the covariate's components are necessary to describe Y. While Allison et al. (2018) extended the approaches of Bierens (1982) and Hlávka, Hušková, Kirch, and Meintanis (2017), Kloodt and Neumeyer (2019) developed a test, which is based on the ideas of Lavergne, Maistre, and Patilea (2015). Neither the first nor the second approach outperforms the other procedure. The approaches of Allison et al. (2018) detect local alternatives with parametric rates, which is not the case for the test of Kloodt and Neumeyer (2019). Kloodt and Neumeyer (2019) in turn supplied a test statistic with an asymptotic distribution, that is independent of the estimation of the transformation parameter, which in general does not hold for the procedures of Allison et al. (2018), and introduced a fast bootstrap algorithm, which performs as good as that of Allison et al. (2018). The independence of the asymptotic behaviour of the test statistic and the estimation of the transformation function is a desirable property that will be fulfilled by the statistic presented in this Section as well.

In the context of testing the null hypothesis of a parametric regression function in semiparametric transformation models, Colling and Van Keilegom (2016, 2017) developed two classes of tests. One the one hand, they extended the approaches of Van Keilegom, González-Manteiga, and Sánchez Sellero (2008) and compared the empirical distribution function of the semiparametrically estimated errors to that of the parametrically estimated errors (Colling and Van Keilegom (2016)). On the other hand, they generalized the ideas of Bierens (1982), Stute (1997) and Escanciano (2006) to develop procedures which were called "integrated approaches" by them (Colling and Van Keilegom (2017)). The main idea of these integrated approaches consists in summing appropriately weighted estimated residuals for some weighting functions that depend on some weighting parameter first (which forms an empirical process with respect to the weighting parameter) and integrating the square of this sum with respect to the weighting parameter. All of the procedures of Colling and Van Keilegom (2016, 2017) detect local alternatives with parametric rates, but require a quite sophisticated and computationally demanding bootstrap algorithm. The methods of Kloodt and Neumeyer (2019) are based on the ideas of Härdle and Mammen (1993) and Zheng (1996) and suffer from a slower convergence rate of detected local alternatives, but provide a test statistic with an asymptotic distribution, that is independent of the estimation of the transformation parameter. Furthermore, they introduced a fast bootstrap procedure, which is competitive to those of Colling and Van Keilegom (2016, 2017). Mu and He (2007) introduced a goodness of fit test for a parametric transformation (quantile regression) model as a whole.

2.3 Testing for a Parametric Regression Function via the Conditional Quantile Function

In Section 1.2, some approaches of how to test for a parametric regression function were presented. Afterwards, the main idea in Section 2.2 was to take these approaches and to modify them in order to obtain valid tests in semiparametric transformation models. Although the presented approach will follow the same spirit, the tools used in this section slightly differ from those in 2.2. Basically, the influence of estimating the transformation parameter in semiparametric models is described by an appropriate Taylor expansion, where asymptotic negligibility of higher terms of the expansion is ensured by appropriate integrability conditions on the parametric transformation function and its derivatives.
In the infinite dimensional nonparametric setting, one has to proceed differently since on the one hand Taylor expansions can not be applied as simply as for parametric transformations and on the other hand the available estimators of the transformation in general only yield satisfying uniform convergence rates on compact sets. See for example the results of Chiappori et al. (2015) and Colling and Van Keilegom (2019). Therefore, a new testing approach, which is extended to nonparametric transformation models in part 2.4, is presented in this section.
Although aiming to test for a parametric regression function, the method provided here is related to testing for a parametric quantile function. Already Chiappori et al. (2015) suggested the estimation of conditional quantiles. See Zheng (1998), Bierens and Ginter (2001), Horowitz and Spokoiny (2002), He and Zhu (2003) and Horowitz and Lee (2009) for some testing approaches in the context of quantile regression or Zheng (2000) for the related question of testing for a parametric conditional distribution function. The test in this section uses a Cramér-von-Mises-type test statistic based on the inverse function of a kernel estimator of the conditional distribution function. In this regard, the testing approach differs from the tests mentioned above and, to the author's knowledge, also from other

tests in the literature. Hence, the asymptotic behaviour is examined in detail in Subsection 2.3.2.

2.3.1 The Test Statistic

From now on, the regression function is allowed to have an arbitrary intercept under the null hypothesis. Usually in regression models, the intercept is estimated as a part of the regression function anyway, so that this is does not reduce the generality of the model severely. Although possible as well, instead of assuming $\mathcal{G}_B$ in (2.2) to be closed with respect to addition of constants the adjusted null hypothesis

$$H_0: \quad g \in \mathcal{G}_B + \mathbb{R} = \{x \mapsto g_\beta(x) + c : \beta \in B, c \in \mathbb{R}\} \tag{2.6}$$

will be considered for reasons of comprehensibility. Here, β and c are identified under assumption (**A7**) from Section 2.7, which will be introduced and discussed later.
Let $(Y_i, X_i), i = 1, ..., n$, be realisations of model (2.1) and let $\tau \in (0,1)$. Let F_ε be the distribution function of ε and denote the τ-quantile of the conditional distribution of a random variable Z (given $X = x$) by $F_Z^{-1}(\tau)$ and $F_{Z|X}^{-1}(\tau|x)$, respectively. Due to

$$F_{Y|X}^{-1}(\tau|x) = E[Y|X = x] + F_\varepsilon^{-1}(\tau) = g(x) + F_\varepsilon^{-1}(\tau), \tag{2.7}$$

there is a strong connection between the conditional τ-quantile and the conditional expectation. Many Cramér-von-Mises-type tests like that of Härdle and Mammen (1993) take advantage of the fact that $g \in \mathcal{G}_B$ is equivalent to $(E[Y|X = x] - g_{\beta_0}(x))^2 = 0$ for all $x \in \mathbb{R}^{d_X}$ and some $\beta_0 \in B$. Referring to (2.7), another condition, which is equivalent to (2.2), is

$$(F_{Y|X}^{-1}(\tau|x) - g_{\beta_0}(x) - F_\varepsilon^{-1}(\tau))^2 = 0 \quad \text{for all } x \in \mathbb{R}^{d_X} \text{ and some } \beta_0 \in B. \tag{2.8}$$

This condition can be translated to the context of (2.6) as

$$(F_{Y|X}^{-1}(\tau|x) - g_{\beta_0}(x) - c)^2 = 0 \quad \text{for all } x \in \mathbb{R}^{d_X} \text{ and some } \beta_0 \in B, c \in \mathbb{R}. \tag{2.9}$$

Let v be a weighting function with compact support in $\mathbb{R}^{d_X}$, such that for all $\tau \in (0,1)$ condition (2.9) and

$$v(x)(F_{Y|X}^{-1}(\tau|x) - g_{\beta_0}(x) - c)^2 = 0 \quad \text{for all } x \in \mathbb{R}^{d_X} \text{ and some } \beta_0 \in B, c \in \mathbb{R} \tag{2.10}$$

are equivalent. Thanks to (2.7), for all $\tau, \chi \in (0,1)$ the function $x \mapsto F_{Y|X}^{-1}(\tau|x) - F_{Y|X}^{-1}(\chi|x)$ is constant, so that equation (2.10) can be extended to multiple quantiles. For this purpose, let μ be a finite measure with compact support in $(0,1)$. Then, (2.10) is equivalent to

$$\min_{c \in \mathbb{R}} \sup_{x \in \mathbb{R}^{d_X}} v(x)(F_{Y|X}^{-1}(\tau|x) - g_{\beta_0}(x) - c)^2 = 0 \quad \text{for all } \tau \in (0,1) \text{ and some } \beta_0 \in B,$$

so that

$$\min_{\beta \in B} \int \min_{c \in \mathbb{R}} \int v(x)(F_{Y|X}^{-1}(\tau|x) - g_\beta(x) - c)^2 \, dx \, \mu(d\tau) = 0. \tag{2.11}$$

Equation (2.11) will be the base of the test statistic. Recall the definitions of Section 1.1 and let K, h_x and h_y be some kernel functions and some bandwidths, respectively, and define $K_{h_y}(y) = \frac{1}{h_y} K\big(\frac{y}{h_y}\big)$ as well as

$$\mathcal{K}(y) = \int_{-\infty}^{y} K(u)\, du, \qquad \mathcal{K}_{h_y}(y) = \int_{-\infty}^{y} K_{h_y}(u)\, du,$$

$$\mathbf{K}(x_1, ..., x_{d_X}) = \prod_{i=1}^{d_X} K(x_i), \qquad \mathbf{K}_{h_x}(x_1, ..., x_{d_X}) = \prod_{i=1}^{d_X} K_{h_x}(x_i)$$

and

$$\begin{aligned}
\hat{p}(y,x) &= \frac{1}{n}\sum_{i=1}^{n} \mathcal{K}_{h_y}(y - Y_i)\mathbf{K}_{h_x}(x - X_i), \\
\hat{f}_X(x) &= \frac{1}{n}\sum_{i=1}^{n} \mathbf{K}_{h_x}(x - X_i), \\
\hat{F}_{Y|X}(y|x) &= \frac{\hat{p}(y,x)}{\hat{f}_X(x)}.
\end{aligned} \tag{2.12}$$

Now, estimate the conditional τ-quantile

$$F_{Y|X}^{-1}(\tau|x) = g(x) + F_{\varepsilon}^{-1}(\tau)$$

via $\hat{F}_{Y|X}^{-1}(\tau|x)$ and define the test statistic as

$$T_n = nh_x^{\frac{d_X}{2}} \min_{\beta\in B} \int \min_{c\in\mathbb{R}} \int v(x)\big(\hat{F}_{Y|X}^{-1}(\tau|x) - g_\beta(x) - c\big)^2\, dx\, \mu(d\tau). \tag{2.13}$$

Here and in the following, $F_{Y|X}^{-1}$ is assumed to be the quantile function if the inverse function of $F_{Y|X}$ does not exist.

Remark 2.3.1 *The inner minimization in (2.13) can be done analytically by solving*

$$\frac{\partial}{\partial c}\int v(x)\big(\hat{F}_{Y|X}^{-1}(\tau|x) - g_\beta(x) - c\big)^2\, dx = -2\int v(x)\big(\hat{F}_{Y|X}^{-1}(\tau|x) - g_\beta(x) - c\big)\, dx = 0,$$

that is

$$\begin{aligned}
T_n &= nh_x^{\frac{d_X}{2}} \min_{\beta\in B} \int\int v(x)\big(\hat{F}_{Y|X}^{-1}(\tau|x) - g_\beta(x) - \hat{c}_{\beta,\tau}\big)^2\, dx\, \mu(d\tau) \\
&= nh_x^{\frac{d_X}{2}} \min_{\beta\in B} \int\int v(x)\left(\frac{\int v(w)\big(\hat{F}_{Y|X}^{-1}(\tau|x) - g_\beta(x) - (\hat{F}_{Y|X}^{-1}(\tau|w) - g_\beta(w))\big)\, dw}{\int v(w)\, dw}\right)^2 \\
&\qquad dx\, \mu(d\tau)
\end{aligned}$$

with

$$\hat{c}_{\beta,\tau} = \frac{\int v(w)(\hat{F}_{Y|X}^{-1}(\tau|w) - g_\beta(w))\, dw}{\int v(w)\, dw}. \tag{2.14}$$

□

2.3.2 Asymptotic Behaviour of the Test Statistic

In the following, the asymptotic behaviour of T_n is examined. Consider the local alternatives

$$H_{1,n}: \quad g(x) = g_{\beta_0}(x) + c_0 + c_n\Delta_n(x) \quad \text{for all } x \in \operatorname{supp}(v) \tag{2.15}$$

and some fixed $\beta_0 \in B, c_0 \in \mathbb{R}$ and define

$$Y_0 = g_{\beta_0}(X) + c_0 + \varepsilon.$$

Here, $c_n = n^{-\frac{1}{2}} h_x^{-\frac{d_X}{4}}$ and Δ_n is assumed to be uniformly bounded in x and n. Similarly to (2.14), define

$$c_{\beta,\tau} = \frac{\int v(x)(F^{-1}_{Y_0|X}(\tau|x) - g_\beta(x))\,dx}{\int v(x)\,dx}. \tag{2.16}$$

As before, the conditional distribution function of Y_0 given X, the (integrated) joint density of (Y_0, X) etc. are denoted by $F_{Y_0|X}$, $f_{Y_0,X}$ (p_0) etc. Let D_β and Hess denote the derivative and the Hessian of a function with respect to β. Let

$$\Omega = \int v(x)\left(\frac{\int v(w)\big(D_\beta g_{\beta_0}(x) - D_\beta g_{\beta_0}(w)\big)\,dw}{\int v(w)\,dw}\right)^t \left(\frac{\int v(w)\big(D_\beta g_{\beta_0}(x) - D_\beta g_{\beta_0}(w)\big)\,dw}{\int v(w)\,dw}\right) dx \tag{2.17}$$

be positive definite (this will be the assumption (**A7**) in Section 2.7). Moreover, define

$$\kappa(x,\tau) = \frac{v(x)}{f_{Y_0|X}(F^{-1}_{Y_0|X}(\tau|x)|x)^2 f_X(x)^2}. \tag{2.18}$$

Before the main result is presented, an auxiliary lemma is given. The assumptions are listed in Section 2.7.

Lemma 2.3.2 *Let $\tau \in (0,1)$. Assume model (2.1) under (**A1**),(**A3**)–(**A6**) from Section 2.7. Further, let*

$$h_x^{3-\frac{d_X}{2}} \to 0, \quad h_y^2 h_x^{-\frac{d_X}{2}} \to 0, \quad h_y h_x^{1-\frac{d_X}{2}} \to 0 \tag{2.19}$$

or let $\int \boldsymbol{K}(w)^2 w\,dw := \big(\int \boldsymbol{K}(w)^2 w_j\,dw\big)_{j=1,\dots,d_X} = 0 \in \mathbb{R}^{d_X}$ and

$$h_x^{3-\frac{d_X}{2}} \to 0, \quad h_y^2 h_x^{-\frac{d_X}{2}} \to 0, \quad h_y h_x^{2-\frac{d_X}{2}} \to 0. \tag{2.20}$$

Then, one has

$$n h_x^{\frac{d_X}{2}} \int\int v(x)\big(\hat{F}^{-1}_{Y_0|X}(\tau|x) - F^{-1}_{Y_0|X}(\tau|x)\big)^2\,dx\,\mu(d\tau) - b \xrightarrow{\mathcal{D}} Z \tag{2.21}$$

with $Z \sim \mathcal{N}(0, V)$,

$$V = 2\int\left(\int \boldsymbol{K}(x)\boldsymbol{K}(x+s)\,dx\right)^2 ds \int \frac{v(w)^2}{f_X(w)^2}\,dw \int_0^1\int_0^1 \left(\int \frac{\big(I_{\{u_1\le\tau\}} - \tau\big)\big(I_{\{u_2\le\tau\}} - \tau\big)}{f_\varepsilon(F_\varepsilon^{-1}(\tau))^2}\,\mu(d\tau)\right)^2 du_1\,du_2 \tag{2.22}$$

and

$$b = h_x^{-\frac{d_X}{2}} \int \boldsymbol{K}(w)^2\,dw \int\int \kappa(x,\tau) p_0(F_{Y_0|X}^{-1}(\tau|x),x)\Big(1 - \frac{p_0(F_{Y_0|X}^{-1}(\tau|x),x)}{f_X(x)}\Big)\,dx\,\mu(d\tau)$$

$$- h_y h_x^{-\frac{d_X}{2}} \int \boldsymbol{K}(w)^2\,dw \int 2z\mathcal{K}(z)K(z)\,dz$$

$$\int\int \kappa(x,\tau) f_{Y_0,X}(F_{Y_0|X}^{-1}(\tau|x),x)\Big(1 - \frac{p_0(F_{Y_0|X}^{-1}(\tau|x),x)}{f_X(x)}\Big)\,dx\,\mu(d\tau)$$

$$+ h_x^{1-\frac{d_X}{2}} \int\int \kappa(x,\tau)\frac{\partial}{\partial u}\Big[p_0(F_{Y_0|X}^{-1}(\tau|x),u)\Big(1 - \frac{p_0(F_{Y_0|X}^{-1}(\tau|x),u)}{f_X(u)}\Big)\Big]_{u=x}\,dx\,\mu(d\tau)$$

$$\int \boldsymbol{K}(w)^2 w\,dw + h_x^{2-\frac{d_X}{2}} \int \boldsymbol{K}(w)^2 w^t \int\int \kappa(x,\tau)\Big[\frac{\partial^2}{\partial u^2} p_0(F_{Y_0|X}^{-1}(\tau|x),u)\Big|_{u=x}$$

$$- 2\frac{p_0(F_{Y_0|X}^{-1}(\tau|x),x)}{f_X(x)}\frac{\partial^2}{\partial u^2} p_0(F_{Y_0|X}^{-1}(\tau|x),u)\Big|_{u=x}$$

$$+ \frac{p_0(F_{Y_0|X}^{-1}(\tau|x),x)^2}{f_X(x)^2}\frac{\partial^2}{\partial x^2} f_X(x)\Big]\,dx\,\mu(d\tau)\,w\,dw.$$

Here, $\frac{\partial^2}{\partial x^2} f_X(x)$ *denotes the Hessian of* f_X*. If*

$$h_x^{1-\frac{d_X}{2}} \to 0, \quad h_y h_x^{-\frac{d_X}{2}} \to 0 \tag{2.23}$$

or $\int \boldsymbol{K}(w)^2 w\,dw = 0$ *and*

$$h_x^{2-\frac{d_X}{2}} \to 0, \quad h_y h_x^{-\frac{d_X}{2}} \to 0, \tag{2.24}$$

b simplifies to

$$b = h_x^{-\frac{d_X}{2}} \int \boldsymbol{K}(w)^2\,dw \int \frac{v(x)}{f_X(x)}\,dx \int \frac{\tau(1-\tau)}{f_\varepsilon(F_\varepsilon^{-1}(\tau))^2}\,\mu(d\tau) + o(1). \tag{2.25}$$

The proof is given in Section 2.8.2.

Remark 2.3.3 *1. Condition (2.19) requires* $3 - \frac{d_X}{2} > 0$*, that is* $d_X \le 5$*, (2.24) leads to* $d_X \le 3$ *and (2.23) even to* $d_X = 1$*. Principally, b can alternatively be defined without any of these conditions as*

$$b = h_x^{\frac{d_X}{2}} \int\int \kappa(x,\tau) E\Big[\Big(\mathcal{K}_{h_y}(F_{Y_0|X}^{-1}(\tau|x) - g_{\beta_0}(X_1) - c_0 - \varepsilon_1) - \frac{p_0(F_{Y_0|X}^{-1}(\tau|x),x)}{f_X(x)}\Big)^2$$

$$\boldsymbol{K}_{h_x}(x - X_1)^2\Big]\,dx\,\mu(d\tau). \tag{2.26}$$

There is a trade off between how simple b is and how strict the bandwidth conditions are.

2. Let $\alpha \in \mathbb{R}^{d_B}$*. Then,*

$$\alpha^t \Omega \alpha = \int v(x)\Big(\Big(D_\beta g_{\beta_0}(x) - \frac{\int v(w) D_\beta g_{\beta_0}(w)\,dw}{\int v(w)\,dw}\Big)\alpha\Big)^2\,dx$$

$$= \int v(x)\Big(D_\beta g_{\beta_0}(x)\alpha - \frac{\int v(w)D_\beta g_{\beta_0}(w)\alpha\,dw}{\int v(w)\,dw}\Big)^2\,dx,$$

*that is, positive definiteness of Ω is only violated if there exists some $\alpha \in \mathbb{R}^{d_B}, \alpha \neq 0$, such that the map $x \mapsto D_\beta g_{\beta_0}(x)\alpha$ is constant with respect to x. For example, this is the case, if $\mathcal{G}_B$ already contains constant functions (e.g. polynomial functions with intercepts). Such a case will be excluded by assumption (**A7**). However, the test can be applied for the general class as well (see the explanation below).*

In the following, the assumptions (2.19) and (2.20) are relaxed and expression (2.26) is used for b. The minimization with respect to c may cause the existence of multiple minimizing parameters $\beta \in B$, especially if $\mathcal{G}_B$ is already closed with respect to addition of constants. To derive the asymptotic behaviour, it will be assumed that

$$\int\int v(x)(F^{-1}_{Y_0|X}(y|x) - g_\beta(x) - c_{\beta,\tau})^2\,dx\,\mu(d\tau) > 0 \quad \text{for all} \quad \beta_0 \neq \beta \in B,$$

that is, the class $\mathcal{G}_B$ is possibly shrunk to avoid multiple solutions β of the outer minimization. Nevertheless, since the value of the test statistic is not influenced by this shrinkage, the resulting test can be applied in the general case as well.

Theorem 2.3.4 *Assume model (2.1). Further, let (**A1**)–(**A8**) from Section 2.7 hold and let b, V and Z be defined as in Lemma 2.3.2. Then,*

$$T_n - b - \delta_n \xrightarrow{\mathcal{D}} Z,$$

where

$$\delta_n = \mu([0,1])\int v(x)\Bigg(\Delta_n(x) - \frac{\int v(w_1)\Delta_n(w_1)\,dw_1}{\int v(w_2)\,dw_2} - \Big(D_\beta g_{\beta_0}(x) - \frac{\int D_\beta g_{\beta_0}(w_3)\,dw_3}{\int v(w_4)\,dw_4}\Big)$$

$$\Omega^{-1}\Big(\int v(w_5)\Delta_n(w_5)\Big(D_\beta g_{\beta_0}(w_5) - \frac{\int D_\beta g_{\beta_0}(w_6)\,dw_6}{\int v(w_7)\,dw_7}\Big)\,dw_5\Big)^t\Bigg)^2\,dx.$$

Under H_0 (that is $\Delta_n \equiv 0$ and thus $\delta_n = 0$), this leads to $T_n - b \xrightarrow{\mathcal{D}} Z$.

The proof can be found in Section 2.8.3. Later, a hypothesis test will be deduced from Theorem 2.3.4. To see whether δ_n lies above some threshold $\delta > 0$, that is, a test based on the asymptotic distribution of T_n would detect the local alternative, define

$$\tilde{\Delta}_n(x) = \Delta_n(x) - \frac{\int v(w_1)\Delta_n(w_1)\,dw_1}{\int v(w_2)\,dw_2}$$

as well as

$$\tilde{D}(x) = D_\beta g_{\beta_0}(x) - \frac{\int D_\beta g_{\beta_0}(w_1)\,dw_1}{\int v(w_2)\,dw_2}.$$

Then,

$$\Omega = \int v(x)\tilde{D}(x)^t\tilde{D}(x)\,dx.$$

Moreover, it can be shown by similar arguments as in the proof of Remark 2.3.5 below that

$$\delta_n = \mu([0,1])\int v(x)\Big(\tilde{\Delta}_n(x) - \Big(\int v(w)\tilde{\Delta}_n(w)\tilde{D}(w)\,dw\Big)\Omega^{-1}\tilde{D}(x)^t\Big)^2\,dx.$$

Let $\tilde{\beta}$ be the minimizer

$$\tilde{\beta} = \arg\min_{\beta \in B} \int v(x)\big(\tilde{\Delta}_n(x) - \tilde{D}(x)\beta\big)^2 \, dx.$$

By standard calculations, it can be shown that

$$\tilde{\beta} = \Omega^{-1} \int v(x)\tilde{\Delta}_n(x)\tilde{D}(x) \, dx$$

and

$$\delta_n = \int v(x)\big(\tilde{\Delta}_n(x) - \tilde{D}(x)\tilde{\beta}\big)^2 \, dx,$$

that is, δ_n is greater than zero, if $\tilde{\Delta}_n$ as a function is linearly independent of the components of $\tilde{D}$. For some fixed $\Delta_n \equiv \Delta$, $\tilde{\beta}$ and δ_n are independent of n, so that

$$\delta_n = \delta = \int v(x)\big(\tilde{\Delta}(x) - \tilde{D}(x)\tilde{\beta}\big)^2 \, dx > 0,$$

if $\tilde{\Delta}$ is linearly independent of $\tilde{D}$. Such an orthogonality condition is quite intuitive and is often assumed explicitly, e.g., by Härdle and Mammen (1993). See the proof of the following remark for more details.

Remark 2.3.5 *1. δ_n can alternatively be expressed as*

$$\begin{aligned}
\delta_n = {} & \mu([0,1]) \int v(x)\Delta_n(x)^2 \, dx - \mu([0,1]) \frac{\big(\int v(x)\Delta_n(x)\, dx\big)^2}{\int v(w)\, dw} \\
& - \mu([0,1]) \Big(\int v(x)\Delta_n(x) \Big(D_\beta g_{\beta_0}(x) - \frac{\int D_\beta g_{\beta_0}(w_1)\, dw_1}{\int v(w_2)\, dw_2} \Big) dx \Big) \\
& \Omega^{-1} \Big(\int v(x)\Delta_n(x) \Big(D_\beta g_{\beta_0}(x) - \frac{\int D_\beta g_{\beta_0}(w_1)\, dw_1}{\int v(w_2)\, dw_2} \Big) dx \Big)^t \\
= {} & \delta_{1,n} + \delta_{2,n} + \delta_{3,n}.
\end{aligned}$$

2. Let model (2.15) hold for some sequence β_n with $||\beta_n - \beta_0|| = c_n$ and $\Delta_n(x) = \frac{g_{\beta_n}(x) - g_{\beta_0}(x)}{c_n}$. Then, $\delta_n = o(1)$, that is, the asymptotic behaviour of the test statistic is the same as for $g(x) = g_{\beta_0}(x) + c_0$. This is consistent with $g = g_{\beta_0} + c_0 + c_n\Delta_n = g_{\beta_n} + c_0 \in \mathcal{G}_B + \mathbb{R}$. The proof can be found on page 71.

The easiest way to construct a test with asymptotic level α for a given $\alpha \in (0,1)$ may consist in estimating b and V by some estimators $\hat{b}$ and $\hat{V}$ and to reject H_0 if $T_n > \hat{b} + \sqrt{\hat{V}} u_{1-\alpha}$, where $u_{1-\alpha}$ denotes the $(1-\alpha)$-quantile of the standard normal distribution. The corresponding test looks like

$$\Phi(Y_1, X_1, ..., Y_n, X_n) = I_{\{T_n > \hat{b} + \sqrt{\hat{V}} u_{1-\alpha}\}}. \tag{2.27}$$

See Section 2.5 for more details.

Theorem 2.3.6 *1. Assume **(A1)**,**(A3)**,**(A5)** from Section 2.7 as well as*

$$\inf_{\tau \in \text{supp}(\mu)} f_\varepsilon(F_\varepsilon^{-1}(\tau)) \inf_{x \in \text{supp}(v)} f_X(x) > 0. \tag{2.28}$$

Moreover, let f_X, f_ε and g_{β_0} be uniformly continuous and assume

$$h_x, h_y, \frac{\log(n)}{nh_x^{d_X}}, \frac{\log(n)}{nh_y} \to 0.$$

If $\hat{b}$ and $\hat{V}$ are some estimators of b and V, such that

$$\hat{b} = o_p\big(nh_x^{\frac{d_X}{2}}\big) \quad \text{and} \quad \sqrt{\hat{V}} = o_p\big(nh_x^{\frac{d_X}{2}}\big),$$

one has

$$P(\Phi(Y_1, X_1, ..., Y_n, X_n) = 1) \to 1$$

*under fixed alternatives. Especially, the test is consistent under (**A1**)–(**A8**) with fixed c_n, Δ_n.*

2. *Assume (**A1**)–(**A8**) for $\Delta_n \equiv 0$ and let $\hat{b}, \hat{V}$ be some estimators of b and V with*

$$\hat{b} - b = o_p(1) \quad \text{and} \quad \hat{V} - V = o_p(1). \tag{2.29}$$

Then,

$$P(\Phi(Y_1, X_1, ..., Y_n, X_n) = 1) \to \alpha.$$

The proof is given in Section 2.8.5.

Remark 2.3.7 *When considering a quantile regression model*

$$Y = g(X) + \tilde{\varepsilon}$$

for some fixed $\tau \in (0,1)$ with $\tilde{\varepsilon}$ not necessarily independent of X and $F^{-1}_{\tilde{\varepsilon}|X}(\tau|X) = 0$ almost surely, the test (without minimizing with respect to c and with μ being the Dirac measure in τ) can still be applied to test for the null hypothesis

$$\tilde{H}_0: \quad g \in \mathcal{G}_B.$$

*It is supposed that after replacing the product density of (X, ε) by the joint density $f_{X,\tilde{\varepsilon}}$ in (**A1**)–(**A8**) and assuming*

$$\inf_{x \in \mathrm{supp}(v)} f_{X,\tilde{\varepsilon}}(x, 0) > 0,$$

the presented results remain valid for testing $\tilde{H}_0$, although in general for different b and V.

□

2.4 Extending the Test to Nonparametric Transformation Models

In this section, the nonparametric transformation model

$$h(Y) = g(X) + \varepsilon \tag{2.30}$$

is considered. Denote the conditional distribution function of $h(Y)$ conditioned on $X = x$ by $F^h_{Y|X}(y|x) = F_{h(Y)|X}(y|x) = P(h(Y) \leq y|X = x)$. Let $\hat{h}$ be a monotonic estimator of h that fulfils

$$\hat{h}(y) - h(y) = \frac{1}{n}\sum_{k=1}^{n} \psi(Y_k, X_k, y) + o_p\Big(\frac{1}{\sqrt{n}}\Big) = \mathcal{O}_p\Big(\frac{1}{\sqrt{n}}\Big) \tag{2.31}$$

uniformly on compact sets, where ψ fulfils (**A9**) in Section 2.7. Examples for such estimators are those of Chiappori et al. (2015) and Colling and Van Keilegom (2019). The main advantage of basing the test statistic on the conditional quantile (instead of the conditional mean) consists in the ability to transfer the convergence rate of the nonparametric transformation estimator to the estimated quantile, which is in general not possible for the conditional mean. To illustrate what is meant by this consider the problem of estimating a specific quantile $(F^h_{Y|X})^{-1}(\tau|x)$ for some $\tau \in (0,1)$ of the conditional distribution function of $h(Y)$ conditioned on $X = x$. If the estimator (2.12) for the conditional distribution function in the untransformed model and its inverse are modified by replacing Y with $\hat{h}(Y)$, this leads to the estimator $(\hat{F}^{\hat{h}}_{Y|X})^{-1}(\tau|x)$, where

$$\hat{F}^{\hat{h}}_{Y|X}(y|x) = \frac{\hat{p}^{\hat{h}}(y,x)}{\hat{f}_X(x)} \quad \text{and} \quad \hat{p}^{\hat{h}}(y,x) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{K}_{h_y}(y - \hat{h}(Y_i))\mathbf{K}_{h_x}(x - X_i).$$

Since K was assumed to have a compact support, for example $[c_1, c_2] \subseteq \mathbb{R}$ for some $c_1 < c_2 \in \mathbb{R}$, the value of $\hat{p}^{\hat{h}}(y,x)$ for some $x \in \mathbb{R}^{d_X}, y \in \mathbb{R}$ only depends on those values of $\hat{h}$ that belong to the interval $[y + h_y c_1, y + h_y c_2]$. This implies that $\hat{h}$ (for this illustration assumed to be strictly increasing) only needs to be evaluated at $[\hat{h}^{-1}(y + h_y c_1), \hat{h}^{-1}(y + h_y c_2)]$, which indeed is a compact set. Translating this to $y \in [y_1, y_2]$, $\hat{h}$ needs to be evaluated at $[\hat{h}^{-1}(y_1 + h_y c_1), \hat{h}^{-1}(y_2 + h_y c_2)]$. In the proof of Theorem 2.4.1 the randomness of this interval is taken into account as well, but together with the fast convergence rate of $\hat{h}$ to h this will be the main reason for the asymptotic negligibility of the transformation estimation.

In the following, quantities which depend on the estimator $\hat{h}$ of the transformation function are marked with an upper $\hat{h}$ (e.g. $T^{\hat{h}}_n, \hat{F}^{\hat{h}}_{Y|X}, \hat{p}^{\hat{h}}$), whereas quantities which depend on the true transformation function h are marked with an upper h (e.g. $T^h_n, \hat{F}^h_{Y|X}, \hat{p}^h, F^h_{Y|X}, p^h$).

Theorem 2.4.1 *Assume model (2.30) under (**A1**)–(**A9**) for $\Delta_n \equiv 0$ and with $F^{-1}_{Y|X}$ replaced by $(F^h_{Y|X})^{-1}$. Then,*

$$\begin{aligned}
T^{\hat{h}}_n &= n h_x^{\frac{d_X}{2}} \min_{\beta \in B} \int \min_{c \in \mathbb{R}} \int v(x)\big((\hat{F}^{\hat{h}}_{Y|X})^{-1}(\tau|x) - g_\beta(x) - c\big)^2 \, dx \, \mu(d\tau) \\
&= n h_x^{\frac{d_X}{2}} \min_{\beta \in B} \int \min_{c \in \mathbb{R}} \int v(x)\big((\hat{F}^{h}_{Y|X})^{-1}(\tau|x) - g_\beta(x) - c\big)^2 \, dx \, \mu(d\tau) + o_p(1) \\
&= T^h_n + o_p(1).
\end{aligned}$$

The proof can be found in Section 2.8.6. The asymptotic behaviour of $T^{\hat{h}}_n$ follows directly from Theorem 2.3.4 and Theorem 2.4.1.

Corollary 2.4.2 *Assume model (2.30) under (**A1**)–(**A9**) from Section 2.7 for $\Delta_n \equiv 0$ as well as (2.23). Then,*

$$T_n^{\hat{h}} - b \xrightarrow{\mathcal{D}} Z$$

with $Z \sim \mathcal{N}(0, V)$,

$$V = 2\int\Big(\int \boldsymbol{K}(x)\boldsymbol{K}(x+s)\,dx\Big)^2 ds \int \frac{v(w)^2}{f_X(w)^2}\,dw$$
$$\int_0^1\int_0^1\Big(\int \frac{\big(I_{\{u_1\le\tau\}}-\tau\big)\big(I_{\{u_2\le\tau\}}-\tau\big)}{f_\varepsilon(F_\varepsilon^{-1}(\tau))^2}\,\mu(d\tau)\Big)^2 du_1\,du_2$$

and

$$b = h_x^{-\frac{d_X}{2}}\int \boldsymbol{K}(w)^2\,dw\int\frac{v(x)}{f_X(x)}\,dx\int\frac{\tau(1-\tau)}{f_\varepsilon(F_\varepsilon^{-1}(\tau))^2}\,\mu(d\tau)+o(1).$$

As in the nontransformation case, a corresponding test can be defined via

$$\Phi^h(Y_1, X_1, ..., Y_n, X_n) = I_{\left\{T_n^{\hat{h}} > \hat{b} + \sqrt{\hat{V}}u_{1-\alpha}\right\}}. \tag{2.32}$$

Theorem 2.4.3 *1. Let $\tau \in (0,1)$ and assume (**A1**),(**A3**),(**A5**) from Section 2.7 as well as (2.28). Further, let f_X, f_ε and g be uniformly continuous and assume*

$$h_x, h_y, \frac{\log(n)}{nh_x^{d_X}}, \frac{\log(n)}{nh_y} \to 0.$$

Let $\hat{h}$ be uniformly consistent over compact sets. If $\hat{b}$ and $\hat{V}$ are some estimators of b and V, such that

$$\hat{b} = o_p\big(nh_x^{\frac{d_X}{2}}\big) \quad \text{and} \quad \sqrt{\hat{V}} = o_p\big(nh_x^{\frac{d_X}{2}}\big),$$

it holds that

$$P(\Phi^h(Y_1, X_1, ..., Y_n, X_n) = 1) \to 1$$

*under fixed alternatives. Especially, the test is consistent under (**A1**),(**A3**)–(**A9**) with fixed alternatives.*

*2. Assume (**A1**)–(**A9**) for $\Delta_n \equiv 0$ and let $\hat{b}, \hat{V}$ be some estimators fulfilling (2.29). Then,*

$$P(\Phi^h(Y_1, X_1, ..., Y_n, X_n) = 1) \to \alpha.$$

The proof is given in Section 2.8.7.

Remark 2.4.4 *1. Note that the asymptotic distribution of $T_n^{\hat{h}}$ is not only independent of the estimation of the transformation function, but also independent of the transformation function h itself, since V and b only depend on the distributions of X and ε.*

2. In principle, if equation (2.31) is valid under (2.15), it should be possible after minor adjustments to extend Theorem 2.4.1 to local alternatives like in (2.15) as well. Since to the author's knowledge there are no results regarding the convergence of a *nonparametric estimator $\hat{h}$ to h (especially no results as (2.31)) in the context of local alternatives with respect to g so far, this idea is not pursued further. Note that this question heads in the same direction as Lemma 5.8.1, where local alternatives with respect to the transformation function are considered.*

□

2.5 Some Thoughts on the Behaviour for Finite Sample Sizes

In this section, a brief overview of the behaviour of the test statistic for small sample sizes is given. Not claiming to present a perfectly elaborated analysis of the tests given in (2.27) and (2.32) the aim is to point out some advantages and disadvantages the given tests might be accompanied with and to give the author's view of possible adjustments worth to consider during future research.
As stated in Theorems 2.3.6 and 2.4.3, some estimators of b and V fulfilling (2.29) are needed before the tests in (2.27) and (2.32) can be applied. Although conceivable for general μ, only the Dirac-Measure $\mu = \delta_\tau$ for $\tau = \frac{1}{2}$ is treated here. Consequently, equations (2.22) and (2.25) lead to

$$V = \frac{2}{f_\varepsilon(F_\varepsilon^{-1}(\frac{1}{2}))^4} \int \Big(\int \mathbf{K}(x)\mathbf{K}(x+s)\,dx \Big)^2 ds \int \frac{v(w)^2}{f_X(w)^2}\,dw \bigg(\int_0^1 \Big(I_{\{u \le \frac{1}{2}\}} - \frac{1}{2} \Big)^2 du \bigg)^2$$

$$= \frac{1}{8 f_\varepsilon(F_\varepsilon^{-1}(\frac{1}{2}))^4} \int \Big(\int \mathbf{K}(x)\mathbf{K}(x+s)\,dx \Big)^2 ds \int \frac{v(w)^2}{f_X(w)^2}\,dw$$

and

$$b = \frac{1}{4 f_\varepsilon(F_\varepsilon^{-1}(\frac{1}{2}))^2 h_x^{\frac{d_x}{2}}} \int \mathbf{K}(w)^2\,dw \int \frac{v(x)}{f_X(x)}\,dx.$$

Some of the components in the expressions above can be calculated explicitly. For example, if K is the Epanechnikow kernel, one has

$$\int \Big(\int K(x)K(x+s)\,dx \Big)^2 ds = \frac{167}{385} \quad \text{and} \quad \int K(x)^2\,dx = \frac{3}{5}.$$

Hence, only $f_\varepsilon(F_\varepsilon^{-1}(\frac{1}{2}))$, $\int \frac{v(w)}{f_X(w)}\,dw$ and $\int \frac{v(w)^2}{f_X(w)^2}\,dw$ are unknown and need to be estimated. For the integrals this can be done by

$$\frac{1}{n}\sum_{i=1}^n \frac{v(X_i)}{\hat{f}_X(X_i)^2} \approx \int \frac{v(w)}{f_X(w)}\,dw \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^n \frac{v(X_i)^2}{\hat{f}_X(X_i)^3} \approx \int \frac{v(w)^2}{f_X(w)^2}\,dw,$$

where $\hat{f}_X$ is the estimator defined in (1.2). $f_\varepsilon(F_\varepsilon^{-1}(\frac{1}{2}))$ in turn can be estimated by

$$\frac{\frac{1}{n}\sum_{i=1}^n v(X_i) K_{h_\varepsilon}\big(\hat{F}_{Y|X}^{-1}(\frac{1}{2}|X_i) - Y_i\big)}{\frac{1}{n}\sum_{i=1}^n v(X_i)}$$

in model (2.1) and in model (2.30) by

$$\frac{\frac{1}{n}\sum_{i=1}^n v(X_i) K_{h_\varepsilon}\big((\hat{F}_{Y|X}^{\hat{h}})^{-1}(\frac{1}{2}|X_i) - \hat{h}(Y_i)\big)}{\frac{1}{n}\sum_{i=1}^n v(X_i)}$$

for some appropriate bandwidth h_ε.

Theorem 2.5.1 *Assume model (2.30) and* ***(A1)****–****(A9)*** *as well as*

$$nh_x^{2d_X}h_\varepsilon^4 \to \infty, \quad n^r h_x^{2d_X}h_\varepsilon^{4(r+1)} \to \infty \quad \text{and} \quad h_\varepsilon^q = o\big(h_x^{\frac{d_X}{2}}\big).$$

Then,

$$\frac{1}{n}\sum_{i=1}^n \frac{v(X_i)}{\hat f_X(X_i)^2} = \int \frac{v(w)}{f_X(w)}\,dw + o\big(h_x^{\frac{d_X}{2}}\big), \tag{2.33}$$

$$\frac{1}{n}\sum_{i=1}^n \frac{v(X_i)^2}{\hat f_X(X_i)^3} = \int \frac{v(w)^2}{f_X(w)^2}\,dw + o\big(h_x^{\frac{d_X}{2}}\big), \tag{2.34}$$

$$\frac{\frac{1}{n}\sum_{i=1}^n v(X_i)K_{h_\varepsilon}\big((\hat F_{Y|X}^{\hat h})^{-1}(\frac{1}{2}|X_i) - \hat h(Y_i)\big)}{\frac{1}{n}\sum_{i=1}^n v(X_i)} = f_\varepsilon\left(F_\varepsilon^{-1}\left(\frac{1}{2}\right)\right) + o\big(h_x^{\frac{d_X}{2}}\big), \tag{2.35}$$

so that $\hat b$ and $\hat V$ are consistent estimators of b and V, respectively. The assertion also holds, if the transformation function h is known, that is, under model (2.1).

The proof is given in Section 2.8.8.

Remark 2.5.2 *1. Theorems 2.5.1, 2.3.6 and 2.4.3 ensure that after plugging the estimators from above into the expressions for b and V the tests presented in (2.27) and (2.32), respectively, are indeed consistent level-α-tests. Note that in addition to the relatively strict bandwidth conditions used in the last Theorem, the expression for b from (2.25) itself requires (2.24) and thus $d_X \leq 3$.*

2. With respect to the asymptotic behaviour of the test statistic and the examination of this behaviour, there are many similarities to the approach of Härdle and Mammen (1993). Since some of the asymptotically negligible terms converge to zero at quite a slow rate, they advise against using the asymptotic distribution in their model to construct tests as is (2.27) or (2.32) via some plug-in-approach. Possibly, some of their ideas can be transferred to the context here to improve the behaviour of the tests for finite sample sizes.

This theoretical background in mind, some observations $(Y_1, X_1), ..., (Y_n, X_n)$ are simulated from the underlying models

$$Y = \gamma\Delta(X) + X - \frac{1}{2} + \varepsilon, \tag{2.36}$$

$$h(Y) = \gamma\Delta(X) + X - \frac{1}{2} + \varepsilon \tag{2.37}$$

in the following, where $X \sim \mathcal{U}([-2,2])$ and $\varepsilon \sim \mathcal{N}(0,1)$ in both cases. In (2.37) h is chosen to be the Yeo-Johnson transformation with parameter equal to $\theta = 1$ (see (5.3)), that is the identity. Note that in contrast to model (2.36), the transformation function will be estimated when model (2.37) is mentioned. The parametric class of regression functions to test for is chosen to be $\mathcal{G}_1 = \{x \mapsto bx + a : a, b \in \mathbb{R}\}$, that is, in terms of the null hypothesis in (2.6), it will be tested for $\mathcal{G}_B + \mathbb{R} = \mathcal{G}_1$ with $\mathcal{G}_B = \{x \mapsto bx : b \in \mathbb{R}\}$. $\Delta(X) = X^2$ and

$\Delta(X) = \exp(X)$ are used as deviation functions, in both cases with $\gamma = 0, 0.5, 1, 1.5, 2, 2.5, 3$. Further,

$$K(x) = \frac{3}{4}(1 - x^2) I_{\{|x| \leq 1\}} \quad \text{and} \quad v(x) = \frac{1}{2} I_{\{|x| \leq 1\}}$$

are used as a kernel and weighting function, respectively. For the implementation, the language R as well as the already implemented functions *integrate* and *h.select* (with *method='cv'*) for integration and bandwidth calculation were used. For the sample size of $n = 100$ the normal reference rule (Silverman (1986)) was applied since the cross validation bandwidths tend to be too small to ensure $\hat{f}_X > 0$ on the support of v.
Table 2.1 shows some rejection probabilities of the test in (2.27) for sample sizes of $n = 100, 200, 1000$ with $m = 200$ repetitions to estimate the rejection probability. First, the asymptotic test seems to be very conservative since under the null hypothesis the test does not even reject once during all of the repetitions in each of the scenarios. The reason for this can be seen in table 2.2. Although only the case $n = 200$ and $\gamma = 0$ is considered there, the same conclusions can be drawn for the other scenarios as well. Table 2.2 lists the means of $\hat{b}$ and $\hat{V}$ as well as the theoretical quantities from the asymptotic distribution and the empirical mean and variance of the test statistic. Since b depends on the random bandwidth h_x, the mean over all $m = 200$ corresponding values of b was taken. Although the estimation is very sensitive with respect to estimation errors in $f_\varepsilon(F_\varepsilon^{-1}(\frac{1}{2}))$, the approximations of b and V are reasonable. However, mean and variance of T seem to be far below their asymptotic counterparts, which results in an extremely conservative test. It is not clear, if estimating the more complex expression for b given in Lemma 2.3.2 would lead to better results. Applying this expression would have been much more computationally demanding on the one hand and does not even necessarily improve the test's performance on the other hand since these estimators might be accompanied with new estimation errors and the estimated variance would be unaffected by those adjustments. Second, this phenomenon not only lowers the rejection probabilities under the null hypothesis, but those under the alternatives as well resulting at least in some models in rather small power.
Some rejection probabilities for the transformation model (2.37) are listed in Table 2.3. Due to the choice of the transformation parameter, model (2.37) coincides with model (2.36), that is, the only difference of the tests consists in the estimation of the transformation function. While for $\Delta(X) = \exp(X)$ the test Φ outperforms Φ^h by far the latter seems to perform slightly better for $\Delta(X) = X^2$. Nevertheless, note that both tests in general test different null hypotheses, which makes a comparison difficult.
Due to the mentioned drawbacks it is not advisable to apply the asymptotic tests directly. Instead, at least the estimators of b and V have to be adjusted. A similar conclusion was drawn by Härdle and Mammen (1993) in their scenario. The author would suggest to apply some bootstrap procedure to mimic the behaviour of the test statistic as a whole instead of only the mean and the variance. Because of the similar structure of the given test to that in Section 5 it is conjectured that an algorithm similar to that of (5.4.1) might work. Nevertheless, although due to the dependence of b on $f_\varepsilon(F_\varepsilon^{-1}(\tau))^2$ some procedures like the wild bootstrap introduced by Liu (1988) and applied for example by Härdle and Mammen (1993) might not lead to consistent estimates of the critical values, already simpler algorithms, probably based on smoothed residual bootstrap like in the paper of Neumeyer

et al. (2016), might work. Since the convergence of nonparametric estimators of the transformation function h in model (2.30) is limited to compact sets so far it is conjectured that applying these approaches might require to estimate the residuals parametrically.

Sample Size		$n = 100$		$n = 200$		$n = 1000$	
Alternative\Level		$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$
$\Delta(X) = X^2$	$\gamma = 0$	0.000	0.000	0.000	0.000	0.000	0.000
	$\gamma = 0.5$	0.020	0.040	0.040	0.060	0.715	0.805
	$\gamma = 1$	0.240	0.320	0.535	0.600	1.000	1.000
	$\gamma = 1.5$	0.640	0.775	0.955	0.980	1.000	1.000
	$\gamma = 2$	0.875	0.940	1.000	1.000	1.000	1.000
	$\gamma = 2.5$	0.970	0.995	1.000	1.000	1.000	1.000
	$\gamma = 3$	0.980	0.990	1.000	1.000	1.000	1.000
$\Delta(X) = \exp(X)$	$\gamma = 0$	0.000	0.000	0.000	0.000	0.000	0.000
	$\gamma = 0.5$	0.005	0.020	0.015	0.020	0.075	0.140
	$\gamma = 1$	0.055	0.070	0.085	0.145	0.865	0.905
	$\gamma = 1.5$	0.160	0.210	0.355	0.420	1.000	1.000
	$\gamma = 2$	0.390	0.465	0.585	0.670	1.000	1.000
	$\gamma = 2.5$	0.560	0.620	0.810	0.850	1.000	1.000
	$\gamma = 3$	0.710	0.755	0.905	0.950	1.000	1.000

Table 2.1: Rejection probabilities for the nontransformation model (2.36) and sample sizes of $n = 100, n = 200$ and $n = 1000$.

Quantity	b	V
mean of the true asymptotic values	5.09	17.04
mean of the estimators	5.21	21.5
empirical mean and variance	1.34	1.87

Table 2.2: Estimated b and V compared to the true asymptotic values and the simple empirical mean and variance in model (2.36) with $n = 200$ and $\gamma = 0$.

Sample Size		$n = 200$		$n = 500$		$n = 1000$	
Level		$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$
$\Delta(X) = X^2$	$\gamma = 0$	0.000	0.000	0.000	0.000	0.000	0.000
	$\gamma = 0.5$	0.040	0.075	0.205	0.310	0.795	0.835
	$\gamma = 1$	0.715	0.820	1.000	1.000	1.000	1.000
	$\gamma = 1.5$	0.990	0.995	1.000	1.000	1.000	1.000
	$\gamma = 2$	1.000	1.000	1.000	1.000	1.000	1.000
	$\gamma = 2.5$	1.000	1.000	1.000	1.000	1.000	1.000
	$\gamma = 3$	1.000	1.000	1.000	1.000	1.000	1.000
$\Delta(X) = \exp(X)$	$\gamma = 0$	0.000	0.000	0.000	0.000	0.000	0.000
	$\gamma = 0.5$	0.000	0.005	0.005	0.010	0.025	0.045
	$\gamma = 1$	0.000	0.005	0.025	0.045	0.175	0.280
	$\gamma = 1.5$	0.010	0.015	0.085	0.115	0.430	0.530
	$\gamma = 2$	0.000	0.000	0.140	0.185	0.570	0.660
	$\gamma = 2.5$	0.005	0.015	0.130	0.185	0.535	0.625
	$\gamma = 3$	0.015	0.025	0.135	0.205	0.490	0.580

Table 2.3: Rejection probabilities for the transformation model (2.37) and sample sizes of $n = 200, n = 500$ and $n = 1000$.

2.6 Discussion

A procedure for testing the null hypothesis of a parametric regression function and an extension to nonparametric transformation models have been developed. Although the testing approach itself might be of interest as well, the main contribution consists in providing a test that allows nonparametric estimation of the transformation function. To the author's knowledge, this has not been done before in the literature and might open the door to a new general class of goodness of fit tests.

There are mainly two aspects of potential improvement. On the one hand, as was mentioned in Section 2.5 an appropriate bootstrap procedure could be applied to mimic the asymptotic behaviour of T_n and $T_n^{\hat{h}}$ better than when using the asymptotic distribution. Such a phenomenon was already observed by Härdle and Mammen (1993). On the other hand, the detection of local alternatives can be possibly improved to detecting parametric rates by changing the test statistic. One way could consist in applying empirical process approaches like He and Zhu (2003) and Horowitz and Lee (2009) did. For example, consider

$$\hat{Z}(\tau, x, e) = \frac{1}{n}\sum_{i=1}^{n} \Xi\big(\hat{\varepsilon}_i(\tau), X_i, x, e\big) \tag{2.38}$$

with $\hat{\varepsilon}_i(\tau) = \hat{h}(Y_i) - (\hat{F}_{Y|X}^{\hat{h}})^{-1}(\tau|X_i)$ and some appropriate function Ξ. If Ξ is chosen to be a weighted indicator function of the form $\Xi(\hat{\varepsilon}(\tau), X, x, e) = v(X_i)I_{\{\hat{\varepsilon}_i(\tau)\leq e\}}$, this leads to a weighted residual process

$$\hat{Z}(\tau, x, e) = \frac{1}{n}\sum_{i=1}^{n} v(X_i) I_{\{\hat{h}(Y_i)-(\hat{F}_{Y|X}^{\hat{h}})^{-1}(\tau|X_i)\leq e\}} \approx \frac{1}{n}\sum_{i=1}^{n} v(X_i) I_{\{\varepsilon_i \leq e + F_\varepsilon^{-1}(\tau)\}} \tag{2.39}$$

similar to that treated in Akritas and Van Keilegom (2001). Perhaps some of the techniques applied there are helpful for the asymptotic analysis of (2.39). In the context of quantile regression, it is more common to use residuals of the form $\tilde{\varepsilon}_i(\tau) = I_{\{\hat{h}(Y_i)-g_{\hat{\beta}}(X_i)\}} - \tau$ for some parametric estimator $g_{\hat{\beta}}$ instead of $\hat{\varepsilon}_i$. Nevertheless, processes as in (2.38) or (2.39) can be possibly used to construct a hypothesis test that is sensitive to faster converging local alternatives.

Note that the general idea to replace the estimated conditional mean by the estimated conditional distribution function is not limited to testing for a parametric regression function. Tests of similar type as Φ and Φ^h to test for significance of components of the covariate X are conceivable as well. Moreover, other nonparametric estimators of the transformation function can be applied as long as they fulfil certain properties, e.g., as (2.31). Especially, it should be possible to obtain an even more general test in the context of heteroscedastic transformation models by applying the estimator, which will be presented in Chapter 4.

2.7 Assumptions

Let $q, r \in \mathbb{N}$.

(**A1**) Depending on which model is considered, models (2.1) or (2.30) hold with ε independent of X. Let $(Y_i, X_i), i = 1, ..., n$, be independent and identically distributed observations from the corresponding model.

(**A2**) One has $g(x) = g_{\beta_0}(x) + c_0 + c_n \Delta_n(x)$ with $c_n = n^{-\frac{1}{2}} h_x^{-\frac{d_X}{4}}$ and some uniformly bounded and continuous function Δ_n (both uniformly in x and n).

(**A3**) The kernel K is r-times continuously differentiable with bounded support. Moreover, one has $\int K(z)\, dz = 1, \int z^l K(z)\, dz = 0$ for all $l = 1, ..., q-1$ and $\int |z^q K(z)|\, dz < \infty$.

(**A4**) The bandwidths h_x and h_y fulfil

$$\frac{n h_x^{2d_X}}{\log(n)^2} \to \infty, \tag{2.40}$$

$$n h_x^{\frac{d_X(r+1)}{2r}} h_y^{\frac{2(r+1)}{r}} \to \infty, \tag{2.41}$$

$$n \log(n)^{-\frac{1}{2}} h_x^{d_X} h_y^{\frac{3}{2}} \to \infty, \tag{2.42}$$

$$n h_x^{\frac{d_X r}{2r-1}} h_y^{\frac{4(r+1)}{2r-1}} \to \infty, \tag{2.43}$$

$$n \log(n)^{-\frac{2}{3}} h_x^{\frac{5d_X}{3}} h_y^2 \to \infty, \tag{2.44}$$

$$\sqrt{n} h_x^q \to 0,$$

$$\sqrt{n} h_y^q \to 0.$$

(**A5**) v is uniformly continuous with compact support.

(**A6**) f_X and g_{β_0} are q-times continuously differentiable. f_ε is q-times continuously differentiable with bounded derivatives. One has $\frac{\partial}{\partial e} f_\varepsilon(e) \overset{|e|\to\infty}{\longrightarrow} 0$, $\frac{\partial^2}{\partial e^2} f_\varepsilon(e) \overset{|e|\to\infty}{\longrightarrow} 0$ and

$$\inf_{\tau \in \text{supp}(\mu)} f_\varepsilon(F_\varepsilon^{-1}(\tau)) \inf_{x \in \text{supp}(v)} f_X(x) > 0.$$

(**A7**) $B \subseteq \mathbb{R}^{d_B}$ is a compact parameter set and the function $(\beta, x) \mapsto g_\beta(x)$ is two times continuously differentiable on $B \times \operatorname{supp}(v)$ with respect to both components. Moreover, it holds that $\int \int v(x)(F^{-1}_{Y_0|X}(y|x) - g_\beta(x) - c_{\beta,\tau})^2\, dx\, \mu(d\tau) > 0$ for all $\beta_0 \neq \beta \in B$.

(**A8**) Ω is positive definite.

(**A9**) The estimator $\hat{h}$ of h is strictly monotone and fulfils (2.31). ψ fulfils $E[\psi(Y, X, y)] = 0$ for all $y \in \mathbb{R}$. For all compact sets $\mathcal{C} \subseteq \mathbb{R}$ the function value $\psi(y_1, x_1, y)$ is uniformly bounded in $(y_1, x_1, y) \in \mathbb{R}^{d_X+1} \times \mathcal{C}$.

Remark 2.7.1 *(i) The standard assumptions $h_x, h_y = o(1)$, $nh_x^{d_X}, nh_y \to \infty$ are implied by (**A4**).*

(ii) Together with $c_n = n^{-\frac{1}{2}} h_x^{-\frac{d_X}{4}}$ the equations (2.41)–(2.44) imply

$$\frac{\sqrt{n}c_n^{r+1}}{h_y^{r+1}} \to 0, \quad \frac{c_n^4 \log(n)}{h_x^{d_X} h_y^3} \to 0, \quad \frac{n^{\frac{1}{4}} c_n^r}{h_y^{r+1}} \to 0 \quad \text{and} \quad \frac{c_n^2 \log(n)}{n^{\frac{1}{4}} h_x^{d_X} h_y^3} \to 0.$$

*(iii) (**A3**) and (**A4**) are fulfilled for $K(u) = \frac{945}{512}(1-u^2)^3\big(1 - \frac{11}{3}u^2\big) I_{[-1,1]}(u)$ (that is $q = 4$ and $r = 2$), $d_X \in \{1, 2\}$ and $h_x = h_y = n^{-\frac{1}{6}}$ (see Hansen (2009)).*

*(iv) Due to $F^{-1}_{Y_0|X}(\tau|x) = g_{\beta_0}(x) + c_0 + F_\varepsilon^{-1}(\tau)$, (**A6**) implies*

$$\inf_{x \in \operatorname{supp}(v),\, \tau \in \operatorname{supp}(\mu)} f_X(x) f_\varepsilon(F^{-1}_{Y_0|X}(\tau|x) - g_{\beta_0}(x) - c_0) > 0.$$

$\square$

2.8 Proofs

Before proving the main results of the Sections 2.3 and 2.4 an auxiliary lemma is given.

2.8.1 An Auxiliary Result

The following Lemma yields an asymptotic expansion for the difference of the conditional quantile function and its estimator $\hat{F}^{-1}_{Y|X}(\tau|x)$. Recall model (2.1) under the local alternatives $H_{1,n}$ in (2.15) and $Y_0 = g_{\beta_0}(X) + c_0 + \varepsilon$.

Lemma 2.8.1 *Assume (**A1**)–(**A5**). Then,*

$$\begin{aligned}
&F^{-1}_{Y_0|X}(\tau|x) - \hat{F}^{-1}_{Y|X}(\tau|x) \\
&= \frac{1}{f_{Y_0|X}(F^{-1}_{Y_0|X}(\tau|x)|x)} \bigg(\frac{1}{f_X(x)} \hat{p}(F^{-1}_{Y_0|X}(\tau|x), x) - \frac{p_0(F^{-1}_{Y_0|X}(\tau|x), x)}{f_X(x)^2} \hat{f}_X(x) \bigg) + o_p\Big(\frac{1}{\sqrt{n}}\Big) \\
&= o_p\big(n^{-\frac{1}{4}}\big), \\
&F^{-1}_{Y_0|X}(\tau|x) - \hat{F}^{-1}_{Y_0|X}(\tau|x)
\end{aligned}$$

$$= \frac{1}{f_{Y_0|X}(F^{-1}_{Y_0|X}(\tau|x)|x)} \left(\frac{1}{f_X(x)} \hat{p}_0(F^{-1}_{Y_0|X}(\tau|x), x) - \frac{p_0(F^{-1}_{Y_0|X}(\tau|x), x)}{f_X(x)^2} \hat{f}_X(x) \right) \tag{2.45}$$

$$+ o_p\left(\frac{1}{\sqrt{n}}\right)$$

$$= o_p(n^{-\frac{1}{4}})$$

and

$$\hat{F}^{-1}_{Y|X}(\tau|x) - \hat{F}^{-1}_{Y_0|X}(\tau|x)$$

$$= \frac{c_n}{n f_{Y_0|X}(F^{-1}_{Y_0|X}(\tau|x)|x) f_X(x)} \sum_{i=1}^{n} K_{h_y}(F^{-1}_{Y_0|X}(\tau|x) - g_{\beta_0}(X_i) - c_0 - \varepsilon_i)$$

$$\boldsymbol{K}_{h_x}(x - X_i)\Delta_n(X_i) + o_p\left(\frac{1}{\sqrt{n}}\right) \tag{2.46}$$

uniformly in $x \in \operatorname{supp}(v)$ *and* $\tau \in \operatorname{supp}(\mu)$.

Proof: Denote the j-th derivatives of K and f_ε by $K^{(j)}$ and $f_\varepsilon^{(j)}$, respectively. For appropriate $y_i^* \in \mathbb{R}$ one has

$$\hat{p}(y, x) - \hat{p}_0(y, x)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \big(\mathcal{K}_{h_y}(y - g_{\beta_0}(X_i) - c_0 - \varepsilon_i - c_n\Delta_n(X_i)) - \mathcal{K}_{h_y}(y - g_{\beta_0}(X_i) - c_0 - \varepsilon_i) \big)$$

$$\mathbf{K}_{h_x}(x - X_i)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=0}^{r-1} K^{(j)}\left(\frac{y - g_{\beta_0}(X_i) - c_0 - \varepsilon_i}{h_y} \right) \mathbf{K}_{h_x}(x - X_i) \frac{(-1)^{j+1} c_n^{j+1} \Delta_n(X_i)^{j+1}}{h_y^{j+1}(j+1)!}$$

$$+ \frac{1}{n} \sum_{i=1}^{n} K^{(r)}(y_i^*) \mathbf{K}_{h_x}(x - X_i) \frac{(-1)^{r+1} c_n^{r+1} \Delta_n(X_i)^{r+1}}{h_y^{r+1}(r+1)!}.$$

Thanks to (**A2**) and (2.41), one has (for an appropriate constant $C > 0$)

$$\left| \frac{1}{n} \sum_{i=1}^{n} K^{(r)}(y_i^*) \mathbf{K}_{h_x}(x - X_i) \frac{(-1)^{r+1} c_n^{r+1} \Delta_n(X_i)^{r+1}}{h_y^{r+1}(r+1)!} \right| \le \frac{C c_n^{r+1}}{h_y^{r+1}} \frac{1}{n} \sum_{i=1}^{n} |\mathbf{K}_{h_x}(x - X_i)|$$

$$= o_p\left(\frac{1}{\sqrt{n}}\right).$$

Moreover, integration by parts yields

$$E\left[\frac{1}{h_y^{j+1}} K^{(j)}\left(\frac{y - g_{\beta_0}(X_1) - c_0 - \varepsilon_i}{h_y} \right) \mathbf{K}_{h_x}(x - X_1) \Delta_n(X_1)^{j+1} \right]$$

$$= \int \int \frac{1}{h_y^{j+1}} K^{(j)}\left(\frac{y - g_{\beta_0}(w) - c_0 - e}{h_y} \right) \mathbf{K}_{h_x}(x - w) \Delta_n(w)^{j+1} f_\varepsilon(e)\, de\, f_X(w)\, dw$$

$$= \int \Bigg(\bigg[-\frac{1}{h_y^j} K^{(j-1)}\left(\frac{y - g_{\beta_0}(w) - c_0 - e}{h_y} \right) f_\varepsilon(e) \bigg]_{-\infty}^{\infty}$$

$$
\begin{aligned}
&+\int \frac{1}{h_y^j} K^{(j-1)}\Big(\frac{y-g_{\beta_0}(w)-c_0-e}{h_y}\Big) f_\varepsilon^{(1)}(e)\,de\Big) \mathbf{K}_{h_x}(x-w)\Delta_n(w)^{j+1} f_X(w)\,dw \\
&= \quad ... \\
&\vdots \\
&= \int \frac{1}{h_y} K\Big(\frac{y-g_{\beta_0}(w)-c_0-e}{h_y}\Big) f_\varepsilon^{(j)}(e)\,de\, \mathbf{K}_{h_x}(x-w)\Delta_n(w)^{j+1} f_X(w)\,dw \\
&= \int K(e) f_\varepsilon^{(j)}(y-g_{\beta_0}(w)-c_0-h_y e)\,de\, \mathbf{K}_{h_x}(x-w)\Delta_n(w)^{j+1} f_X(w)\,dw \\
&= f_\varepsilon^{(j)}(y-g_{\beta_0}(w)-c_0)\Delta_n(x)^{j+1} f_X(x)+o(1)
\end{aligned}
$$

uniformly in $y\in\mathbb{R}, x\in\operatorname{supp}(v)$ for all $j=1,...,r-1$. (2.41) and (2.42) imply

$$
\frac{c_n^{2(j+1)}\log(n)}{h_x^{d_X} h_y^{2j+1}} \to 0
$$

for all $j=1,...,r-1$, so that

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n K^{(j)}\Big(\frac{y-g_{\beta_0}(X_i)-c_0-\varepsilon_i}{h_y}\Big)\mathbf{K}_{h_x}(x-X_i)\frac{(-1)^{j+1}c_n^{j+1}\Delta_n(X_i)^{j+1}}{h_y^{j+1}(j+1)!} \\
&= \frac{(-1)^{j+1}c_n^{j+1}}{nh_y^{j+1}(j+1)!}\sum_{i=1}^n\Bigg(K^{(j)}\Big(\frac{y-g_{\beta_0}(X_i)-c_0-\varepsilon_i}{h_y}\Big)\mathbf{K}_{h_x}(x-X_i)\Delta_n(X_i)^{j+1} \\
&\quad - E\Big[K^{(j)}\Big(\frac{y-g_{\beta_0}(X_1)-c_0-\varepsilon_1}{h_y}\Big)\mathbf{K}_{h_x}(x-X_1)\Delta_n(X_1)^{j+1}\Big]\Bigg)+o_p\Big(\frac{1}{\sqrt{n}}\Big) \\
&= \frac{c_n^{j+1}}{h_y^j}\mathcal{O}_p\Bigg(\sqrt{\frac{\log(n)}{nh_x^{d_X}h_y}}\Bigg)+o_p\Big(\frac{1}{\sqrt{n}}\Big) \\
&= o_p\Big(\frac{1}{\sqrt{n}}\Big)
\end{aligned}
$$

for all $j=1,...,r-1$ and uniformly with respect to $x\in\operatorname{supp}(v)$ and with respect to y in some compact set, where the second to last equality follows from the results of Hansen (2008) (see section 1.1). Hence,

$$
\begin{aligned}
\hat{p}(y,x)-\hat{p}_0(y,x) &= -\frac{c_n}{n}\sum_{i=1}^n K_{h_y}(y-g_{\beta_0}(X_i)-c_0-\varepsilon_i)\mathbf{K}_{h_x}(x-X_i)\Delta_n(X_i)+o_p\Big(\frac{1}{\sqrt{n}}\Big) \\
&= \mathcal{O}_p(c_n) \\
&= o_p\big(n^{-\frac{1}{4}}\big). \qquad (2.47)
\end{aligned}
$$

uniformly on compact sets with respect to y and uniformly in $x\in\operatorname{supp}(v)$. Since (2.43) and (2.44) imply

$$
n\log(n)^{-\frac{2}{2j+1}} h_x^{\frac{d_X(j+4)}{2j+1}} h_y^2 \to \infty \ \Big(\Rightarrow \frac{c_n^{2j}\log(n)}{n^{\frac{1}{4}}h_x^{d_X}h_y^{2j+1}}\to 0\Big)
$$

for all $j = 1, ..., r-1$, a similar reasoning leads to

$$\hat{f}_{Y|X}(y|x) - f_{Y_0|X}(y|x) = o_p\big(n^{-\frac{1}{4}}\big)$$

uniformly on compact sets. These asymptotic expressions will be used to obtain a similar expression for $\hat{F}^{-1}_{Y|X}(\tau|x) - F^{-1}_{Y|X}(\tau|x)$. Again, Theorem 2 of Hansen (2008) or more precisely the adjustments discussed later in the proof of Lemma 4.2.12 combined with (**A4**) ensure that

$$\hat{p}_0(y,x) - p_0(y,x) = o_p\big(n^{-\frac{1}{4}}\big), \quad \hat{f}_X(x) - f_X(x) = o_p\big(n^{-\frac{1}{4}}\big) \quad \text{and} \quad \frac{\partial}{\partial y}\hat{f}_{Y|X}(y|x) = \mathcal{O}_p(1)$$

uniformly on compact sets, so that Lemma 1.1.2 leads to

$$\begin{aligned}
&\hat{F}_{Y_0|X}(y|x) - F_{Y_0|X}(y|x)\\
&= \frac{\hat{p}_0(y,x)}{\hat{f}_X(x)} - \frac{p_0(y,x)}{f_X(x)}\\
&= \frac{1}{f_X(x)}(\hat{p}_0(y,x) - p_0(y,x)) - \frac{p_0(y,x)}{f_X(x)^2}(\hat{f}_X(x) - f_X(x))\\
&\quad - \frac{\hat{f}_X(x) - f_X(x)}{\hat{f}_X(x) f_X(x)}\left(\hat{p}_0(y,x) - p_0(y,x) - \frac{p_0(y,x)(\hat{f}_X(x) - f_X(x))}{f_X(x)}\right)\\
&= \frac{1}{f_X(x)}(\hat{p}_0(y,x) - p_0(y,x)) - \frac{p_0(y,x)}{f_X(x)^2}(\hat{f}_X(x) - f_X(x)) + o_p\left(\frac{1}{\sqrt{n}}\right)\\
&= o_p\big(n^{-\frac{1}{4}}\big).
\end{aligned}$$

and

$$\begin{aligned}
\hat{F}_{Y|X}(y|x) - F_{Y_0|X}(y|x) &= \frac{1}{f_X(x)}(\hat{p}(y,x) - p_0(y,x)) - \frac{p_0(y,x)}{f_X(x)^2}(\hat{f}_X(x) - f_X(x))\\
&\quad + o_p\left(\frac{1}{\sqrt{n}}\right)\\
&= o_p\big(n^{-\frac{1}{4}}\big).
\end{aligned}$$

Since for an appropriate y^* between $\hat{F}^{-1}_{Y|X}(\tau|x)$ and $F^{-1}_{Y_0|X}(\tau|x)$

$$\begin{aligned}
0 &= \hat{F}_{Y|X}(\hat{F}^{-1}_{Y|X}(\tau|x)|x) - F_{Y_0|X}(F^{-1}_{Y_0|X}(\tau|x)|x)\\
&= \hat{F}_{Y|X}(F^{-1}_{Y_0|X}(\tau|x)|x) + \hat{f}_{Y|X}(F^{-1}_{Y_0|X}(\tau|x)|x)\big(\hat{F}^{-1}_{Y|X}(\tau|x) - F^{-1}_{Y_0|X}(\tau|x)\big)\\
&\quad + \frac{\partial}{\partial y}\hat{f}_{Y|X}(y^*|x)\big(\hat{F}^{-1}_{Y|X}(\tau|x) - F^{-1}_{Y_0|X}(\tau|x)\big)^2 - F_{Y_0|X}(F^{-1}_{Y_0|X}(\tau|x)|x)\\
&= \hat{F}_{Y|X}(F^{-1}_{Y_0|X}(\tau|x)|x) - F_{Y_0|X}(F^{-1}_{Y_0|X}(\tau|x)|x)\\
&\quad + f_{Y|X}(F^{-1}_{Y_0|X}(\tau|x)|x)\big(\hat{F}^{-1}_{Y|X}(\tau|x) - F^{-1}_{Y_0|X}(\tau|x)\big)\\
&\quad + \big(\hat{f}_{Y|X}(F^{-1}_{Y_0|X}(\tau|x)|x) - f_{Y|X}(F^{-1}_{Y_0|X}(\tau|x)|x)\big)\big(\hat{F}^{-1}_{Y|X}(\tau|x) - F^{-1}_{Y_0|X}(\tau|x)\big)
\end{aligned}$$

$$+\frac{\partial}{\partial y}\hat{f}_{Y|X}(y^*|x)\big(\hat{F}^{-1}_{Y|X}(\tau|x)-F^{-1}_{Y_0|X}(\tau|x)\big)^2$$

due to the continuity of $\hat{F}_{Y|X}$ and $F_{Y_0|X}$, it holds that $\hat{F}^{-1}_{Y|X}(\tau|x)-F^{-1}_{Y_0|X}(\tau|x)=o_p\big(n^{-\frac{1}{4}}\big)$ uniformly in $x\in\operatorname{supp}(v)$ and $\tau\in\operatorname{supp}(\mu)$. Moreover, note that

$$f_{Y|X}(y|x)-f_{Y_0|X}(y|x)=\mathcal{O}(c_n)$$

uniformly on compact sets and that $c_n n^{-\frac{1}{4}}=o\big(n^{-\frac{1}{2}}\big)$. Hence,

$$\begin{aligned}
0&=\hat{F}_{Y|X}(\hat{F}^{-1}_{Y|X}(\tau|x)|x)-F_{Y_0|X}(F^{-1}_{Y_0|X}(\tau|x)|x)\\
&=\hat{F}_{Y|X}(F^{-1}_{Y_0|X}(\tau|x)|x)-F_{Y_0|X}(F^{-1}_{Y_0|X}(\tau|x)|x)\\
&\quad+f_{Y|X}(F^{-1}_{Y_0|X}(\tau|x)|x)\big(\hat{F}^{-1}_{Y|X}(\tau|x)-F^{-1}_{Y_0|X}(\tau|x)\big)+\mathcal{O}_p\Big(\big(\hat{F}^{-1}_{Y|X}(\tau|x)-F^{-1}_{Y_0|X}(\tau|x)\big)^2\Big)\\
&\quad+\mathcal{O}_p\Big(\big(\hat{f}_{Y|X}(F^{-1}_{Y_0|X}(\tau|x))-f_{Y_0|X}(F^{-1}_{Y_0|X}(\tau|x))\big)\big(\hat{F}^{-1}_{Y|X}(\tau|x)-F^{-1}_{Y_0|X}(\tau|x)\big)\Big)+o_p\Big(\frac{1}{\sqrt{n}}\Big)\\
&=\hat{F}_{Y|X}(F^{-1}_{Y_0|X}(\tau|x)|x)-F_{Y_0|X}(F^{-1}_{Y_0|X}(\tau|x)|x)\\
&\quad+f_{Y_0|X}(F^{-1}_{Y_0|X}(\tau|x)|x)\big(\hat{F}^{-1}_{Y|X}(\tau|x)-F^{-1}_{Y_0|X}(\tau|x)\big)+o_p\Big(\frac{1}{\sqrt{n}}\Big)
\end{aligned}$$

uniformly in $x\in\operatorname{supp}(\mathrm{v})$ and $\tau\in\operatorname{supp}(\mu)$. Due to (**A6**), this in turn implies

$$\begin{aligned}
&F^{-1}_{Y_0|X}(\tau|x)-\hat{F}^{-1}_{Y|X}(\tau|x)\\
&=\frac{\hat{F}_{Y|X}(F^{-1}_{Y_0|X}(\tau|x)|x)-F_{Y_0|X}(F^{-1}_{Y_0|X}(\tau|x)|x)}{f_{Y|X}(F^{-1}_{Y_0|X}(\tau|x)|x)}+o_p\Big(\frac{1}{\sqrt{n}}\Big)\\
&=\frac{1}{f_{Y_0|X}(F^{-1}_{Y_0|X}(\tau|x)|x)}\Big(\frac{1}{f_X(x)}\hat{p}(F^{-1}_{Y_0|X}(\tau|x),x)-\frac{p_0(F^{-1}_{Y_0|X}(\tau|x),x)}{f_X(x)^2}\hat{f}_X(x)\Big)+o_p\Big(\frac{1}{\sqrt{n}}\Big).
\end{aligned}$$

The same expression can be obtained for $F^{-1}_{Y_0|X}(\tau|x)-\hat{F}^{-1}_{Y_0|X}(\tau|x)$ when replacing $\hat{p}$ by $\hat{p}_0$, so that (see (2.47))

$$\begin{aligned}
&\hat{F}^{-1}_{Y|X}(\tau|x)-\hat{F}^{-1}_{Y_0|X}(\tau|x)\\
&=\frac{1}{f_{Y_0|X}(F^{-1}_{Y_0|X}(\tau|x)|x)f_X(x)}\big(\hat{p}_0(F^{-1}_{Y_0|X}(\tau|x),x)-\hat{p}(F^{-1}_{Y_0|X}(\tau|x),x)\big)+o_p\Big(\frac{1}{\sqrt{n}}\Big)\\
&=\frac{c_n}{f_{Y_0|X}(F^{-1}_{Y_0|X}(\tau|x)|x)f_X(x)n}\sum_{i=1}^n K_{h_y}(F^{-1}_{Y_0|X}(\tau|x)-g_{\beta_0}(X_i)-c_0-\varepsilon_i)\\
&\qquad\mathbf{K}_{h_x}(x-X_i)\Delta_n(X_i)+o_p\Big(\frac{1}{\sqrt{n}}\Big).
\end{aligned}$$

□

2.8.2 Proof of Lemma 2.3.2

Thanks to Lemma 2.8.1 the difference $\hat{F}^{-1}_{Y_0|X}(\tau|x)-F^{-1}_{Y_0|X}(\tau|x)$ can be written as

$$\hat{F}^{-1}_{Y_0|X}(\tau|x)-F^{-1}_{Y_0|X}(\tau|x)$$

$$= \frac{1}{f_{Y_0|X}(F_{Y_0|X}^{-1}(\tau|x)|x)} \frac{1}{n} \sum_{i=1}^{n} \bigg(\frac{1}{f_X(x)} \mathcal{K}_{h_y}(F_{Y_0|X}^{-1}(\tau|x) - g_{\beta_0}(X_i) - c_0 - \varepsilon_i)$$
$$- \frac{p_0(F_{Y_0|X}^{-1}(\tau|x), x)}{f_X(x)^2} \bigg) \mathbf{K}_{h_x}(x - X_i) + o_p\Big(\frac{1}{\sqrt{n}}\Big)$$

uniformly in $x \in \operatorname{supp}(v)$, so that

$$n h_x^{\frac{d_X}{2}} \int v(x) \big(\hat{F}_{Y_0|X}^{-1}(\tau|x) - F_{Y_0|X}^{-1}(\tau|x)\big)^2 \, dx$$
$$= n h_x^{\frac{d_X}{2}} \int\int v(x) \bigg(\frac{1}{f_{Y_0|X}(F_{Y_0|X}^{-1}(\tau|x)|x)} \frac{1}{n} \sum_{i=1}^{n} \bigg(\frac{1}{f_X(x)} \mathcal{K}_{h_y}(F_{Y_0|X}^{-1}(\tau|x) - g_{\beta_0}(X_i) - c_0 - \varepsilon_i)$$
$$- \frac{p_0(F_{Y_0|X}^{-1}(\tau|x), x)}{f_X(x)^2} \bigg) \mathbf{K}_{h_x}(x - X_i) + o_p\Big(\frac{1}{\sqrt{n}}\Big) \bigg)^2 \, dx \, \mu(d\tau)$$
$$= n h_x^{\frac{d_X}{2}} \int\int v(x) \bigg(\frac{1}{f_{Y_0|X}(F_{Y_0|X}^{-1}(\tau|x)|x)} \frac{1}{n} \sum_{i=1}^{n} \bigg(\frac{1}{f_X(x)} \mathcal{K}_{h_y}(F_{Y_0|X}^{-1}(\tau|x) - g_{\beta_0}(X_i) - c_0 - \varepsilon_i)$$
$$- \frac{p_0(F_{Y_0|X}^{-1}(\tau|x), x)}{f_X(x)^2} \bigg) \mathbf{K}_{h_x}(x - X_i) \bigg) \bigg)^2 \, dx \, \mu(d\tau) + \int\int o_p\Big(\sqrt{n} h_x^{\frac{d_X}{2}}\Big) v(x)$$
$$\bigg(\frac{1}{f_{Y_0|X}(F_{Y_0|X}^{-1}(\tau|x)|x)} \frac{1}{n} \sum_{i=1}^{n} \bigg(\frac{1}{f_X(x)} \mathcal{K}_{h_y}(F_{Y_0|X}^{-1}(\tau|x) - g_{\beta_0}(X_i) - c_0 - \varepsilon_i)$$
$$- \frac{p_0(F_{Y_0|X}^{-1}(\tau|x), x)}{f_X(x)^2} \bigg) \mathbf{K}_{h_x}(x - X_i) \bigg) \, dx \, \mu(d\tau) + o_p(1).$$

Recall

$$\kappa(x, \tau) = \frac{v(x)}{f_{Y_0|X}(F_{Y_0|X}^{-1}(\tau|x)|x)^2 f_X(x)^2}.$$

Because of Hölder's inequality it suffices to show the assertion for

$$n h_x^{\frac{d_X}{2}} \int\int \frac{v(x)}{f_{Y_0|X}(F_{Y_0|X}^{-1}(\tau|x)|x)^2 f_X(x)^2} \bigg(\frac{1}{n} \sum_{i=1}^{n} \bigg(\mathcal{K}_{h_y}(F_{Y_0|X}^{-1}(\tau|x) - g_{\beta_0}(X_i) - c_0 - \varepsilon_i)$$
$$- \frac{p_0(F_{Y_0|X}^{-1}(\tau|x), x)}{f_X(x)} \bigg) \mathbf{K}_{h_x}(x - X_i) \bigg)^2 \, dx \, \mu(d\tau)$$
$$= \frac{h_x^{\frac{d_X}{2}}}{n} \sum_{i=1}^{n} \int\int \kappa(x, \tau) \bigg(\mathcal{K}_{h_y}(F_{Y_0|X}^{-1}(\tau|x) - g_{\beta_0}(X_i) - c_0 - \varepsilon_i) - \frac{p_0(F_{Y_0|X}^{-1}(\tau|x), x)}{f_X(x)} \bigg)^2$$
$$\mathbf{K}_{h_x}(x - X_i)^2 \, dx \, \mu(d\tau) + \frac{h_x^{\frac{d_X}{2}}}{n} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \int\int \kappa(x, \tau)$$
$$\bigg(\mathcal{K}_{h_y}(F_{Y_0|X}^{-1}(\tau|x) - g_{\beta_0}(X_i) - c_0 - \varepsilon_i) - \frac{p_0(F_{Y_0|X}^{-1}(\tau|x), x)}{f_X(x)} \bigg) \mathbf{K}_{h_x}(x - X_i)$$
$$\bigg(\mathcal{K}_{h_y}(F_{Y_0|X}^{-1}(\tau|x) - g_{\beta_0}(X_j) - c_0 - \varepsilon_j) - \frac{p_0(F_{Y_0|X}^{-1}(\tau|x), x)}{f_X(x)} \bigg) \mathbf{K}_{h_x}(x - X_j) \, dx \, \mu(d\tau)$$

$$= T_1 + T_2.$$

Asymptotic Behaviour of T_1

First, note that using integration by parts and Lemma 1.1.1, one has

$$\int \mathcal{K}_{h_y}(F^{-1}_{Y_0|X}(\tau|x) - z)^2 f_{Y_0,X}(z,x)\,dz$$

$$= \int 2\mathcal{K}_{h_y}(F^{-1}_{Y_0|X}(\tau|x) - z)K_{h_y}(F^{-1}_{Y_0|X}(\tau|x) - z)p_0(z,x)\,dz$$

$$= \int 2\mathcal{K}(z)K(z)p_0(F^{-1}_{Y_0|X}(\tau|x) - h_y z, x)\,dz$$

$$= \int 2\mathcal{K}(z)K(z)\,dz\,p_0(F^{-1}_{Y_0|X}(\tau|x), x) - h_y \int 2z\mathcal{K}(z)K(z)\,dz\,f_{Y_0,X}(F^{-1}_{Y_0|X}(\tau|x), x) + O\big(h_y^2\big)$$

as well as

$$\int 2\mathcal{K}(z)K(z)\,dz = \lim_{u\to\infty}\int_{-\infty}^{u} 2\mathcal{K}(z)K(z)\,dz = \lim_{u\to\infty}\mathcal{K}(u)^2 = 1$$

and

$$\int \mathcal{K}_{h_y}(F^{-1}_{Y_0|X}(\tau|x) - z)f_{Y_0,X}(z,x)\,dz = p_0(F^{-1}_{Y_0|X}(\tau|x), x) + o\Big(\frac{1}{\sqrt{n}}\Big)$$

uniformly in $x \in \operatorname{supp}(v)$ and $\tau \in \operatorname{supp}(\mu)$, so that

$$\int \bigg(\mathcal{K}_{h_y}(F^{-1}_{Y_0|X}(\tau|x) - z) - \frac{p_0(F^{-1}_{Y_0|X}(\tau|x))}{f_X(x)}\bigg)^2 f_{Y_0,X}(z,x)\,dz$$

$$= p_0(F^{-1}_{Y_0|X}(\tau|x), x) - h_y\int 2z\mathcal{K}(z)K(z)\,dz\,f_{Y_0,X}(F^{-1}_{Y_0|X}(\tau|x), x)$$

$$- \frac{p_0(F^{-1}_{Y_0|X}(\tau|x), x)^2}{f_X(x)} + o\Big(\frac{1}{\sqrt{n}}\Big) + O\big(h_y^2\big)$$

$$= p_0(F^{-1}_{Y_0|X}(\tau|x), x)\bigg(1 - \frac{p_0(F^{-1}_{Y_0|X}(\tau|x), x)}{f_X(x)}\bigg) - h_y\int 2z\mathcal{K}(z)K(z)\,dz\,f_{Y_0,X}(F^{-1}_{Y_0|X}(\tau|x), x)$$

$$+ o\Big(\frac{1}{\sqrt{n}}\Big) + O\big(h_y^2\big)$$

uniformly in $x \in \operatorname{supp}(v)$ and $\tau \in \operatorname{supp}(\mu)$. Similar calculations yield

$$\int \bigg(\mathcal{K}_{h_y}(F^{-1}_{Y_0|X}(\tau|x) - z) - \frac{p_0(F^{-1}_{Y_0|X}(\tau|x))}{f_X(x)}\bigg)^2 \frac{\partial}{\partial x} f_{Y_0,X}(z,x)\,dz$$

$$= \frac{\partial}{\partial u}p_0(F^{-1}_{Y_0|X}(\tau|x), u)\Big|_{u=x} - 2\frac{p_0(F^{-1}_{Y_0|X}(\tau|x), x)}{f_X(x)}\frac{\partial}{\partial u}p_0(F^{-1}_{Y_0|X}(\tau|x), u)\Big|_{u=x}$$

$$+ \frac{p_0(F^{-1}_{Y_0|X}(\tau|x), x)^2}{f_X(x)^2}\frac{\partial}{\partial x}f_X(x) - h_y\int 2z\mathcal{K}(z)K(z)\,dz\,\frac{\partial}{\partial x}f_{Y_0,X}(y,x)\Big|_{y=F^{-1}_{Y_0|X}(\tau|x)}$$

$$+ o\Big(\frac{1}{\sqrt{n}}\Big) + O\big(h_y^2\big)$$

$$= \frac{\partial}{\partial u}\left(p_0(F^{-1}_{Y_0|X}(\tau|x), u)\left(1 - \frac{p_0(F^{-1}_{Y_0|X}(\tau|x), u)}{f_X(u)}\right)\right)\Big|_{u=x}$$

$$- h_y \int 2z\mathcal{K}(z)K(z)\,dz\,\frac{\partial}{\partial x} f_{Y_0,X}(y,x)\Big|_{y=F^{-1}_{Y_0|X}(\tau|x)} + o\Big(\frac{1}{\sqrt{n}}\Big) + O\big(h_y^2\big)$$

and

$$\int \left(\mathcal{K}_{h_y}(F^{-1}_{Y_0|X}(\tau|x) - z) - \frac{p_0(F^{-1}_{Y_0|X}(\tau|x))}{f_X(x)}\right)^2 \frac{\partial^2}{\partial x^2} f_{Y_0,X}(z,x)\,dz$$

$$= \frac{\partial^2}{\partial u^2} p_0(F^{-1}_{Y_0|X}(\tau|x), u)\Big|_{u=x} - 2\frac{p_0(F^{-1}_{Y_0|X}(\tau|x), x)}{f_X(x)}\frac{\partial^2}{\partial u^2} p_0(F^{-1}_{Y_0|X}(\tau|x), u)\Big|_{u=x}$$

$$+ \frac{p_0(F^{-1}_{Y_0|X}(\tau|x), x)^2}{f_X(x)^2}\frac{\partial^2}{\partial x^2} f_X(x) - h_y \int 2z\mathcal{K}(z)K(z)\,dz\,\frac{\partial^2}{\partial u^2} f_{Y_0,X}(F^{-1}_{Y_0|X}(\tau|x), u)\Big|_{u=x}$$

$$+ o\Big(\frac{1}{\sqrt{n}}\Big) + O\big(h_y^2\big)$$

uniformly in $x \in \operatorname{supp}(v)$ and $\tau \in \operatorname{supp}(\mu)$. Here, $\frac{\partial}{\partial x} f_X(x)$ and $\frac{\partial^2}{\partial x^2} f_X(x)$ are the derivative and the Hessian of f_X, that is a vector and a matrix. The expectation of T_1 can be written as

$$E[T_1]$$

$$= h_x^{\frac{d_X}{2}} \int\int \kappa(x,\tau) E\Bigg[\left(\mathcal{K}_{h_y}(F^{-1}_{Y_0|X}(\tau|x) - g_{\beta_0}(X_i) - c_0 - \varepsilon_i) - \frac{p_0(F^{-1}_{Y_0|X}(\tau|x), x)}{f_X(x)}\right)^2$$

$$\mathbf{K}_{h_x}(x - X_i)^2\Bigg]\,dx\,\mu(d\tau)$$

$$= h_x^{\frac{d_X}{2}} \int\int \kappa(x,\tau) \int \left(\mathcal{K}_{h_y}(F^{-1}_{Y_0|X}(\tau|x) - z) - \frac{p_0(F^{-1}_{Y_0|X}(\tau|x), x)}{f_X(x)}\right)^2$$

$$\int \mathbf{K}_{h_x}(x - w)^2 f_{Y_0,X}(z,w)\,dw\,dz\,dx\,\mu(d\tau)$$

$$= h_x^{-\frac{d_X}{2}} \int\int \kappa(x,\tau) \int \left(\mathcal{K}_{h_y}(F^{-1}_{Y_0|X}(\tau|x) - z) - \frac{p_0(F^{-1}_{Y_0|X}(\tau|x), x)}{f_X(x)}\right)^2$$

$$\int \mathbf{K}(w)^2 f_{Y_0,X}(z, x - h_x w)\,dw\,dz\,dx\,\mu(d\tau)$$

$$= h_x^{-\frac{d_X}{2}} \int\int \kappa(x,\tau) \int \left(\mathcal{K}_{h_y}(F^{-1}_{Y_0|X}(\tau|x) - z) - \frac{p_0(F^{-1}_{Y_0|X}(\tau|x), x)}{f_X(x)}\right)^2$$

$$\bigg(\int \mathbf{K}(w)^2\,dw\, f_{Y_0,X}(z,x) + h_x \frac{\partial}{\partial u} f_{Y_0,X}(z,u)\Big|_{u=x} \int \mathbf{K}(w)^2 w\,dw$$

$$+ h_x^2 \int \mathbf{K}(w)^2 w^t \frac{\partial^2}{\partial x^2} f_{Y_0,X}(z,x) w\,dw + \mathcal{O}\big(h_x^3\big)\bigg)\,dz\,dx\,\mu(d\tau)$$

$$= h_x^{-\frac{d_X}{2}} \int \mathbf{K}(w)^2\,dw \int\int \kappa(x,\tau) p_0(F^{-1}_{Y_0|X}(\tau|x), x)\left(1 - \frac{p_0(F^{-1}_{Y_0|X}(\tau|x), x)}{f_X(x)}\right)dx\,\mu(d\tau)$$

$$
\begin{aligned}
&- h_y h_x^{-\frac{d_X}{2}} \int \mathbf{K}(w)^2\,dw \int 2z\mathcal{K}(z)K(z)\,dz \int\int \kappa(x,\tau) f_{Y_0,X}(F_{Y_0|X}^{-1}(\tau|x),x)\,dx\,\mu(d\tau) \\
&+ h_x^{1-\frac{d_X}{2}} \int\int \kappa(x,\tau) \\
&\frac{\partial}{\partial u}\bigg[p_0(F_{Y_0|X}^{-1}(\tau|x),u)\Big(1-\frac{p_0(F_{Y_0|X}^{-1}(\tau|x),u)}{f_X(u)}\Big)\bigg]_{u=x}\,dx\,\mu(d\tau) \int \mathbf{K}(w)^2 w\,dw \\
&+ h_x^{2-\frac{d_X}{2}} \int \mathbf{K}(w)^2 w^t \int\int \kappa(x,\tau)\bigg[\frac{\partial^2}{\partial u^2} p_0(F_{Y_0|X}^{-1}(\tau|x),u)\Big|_{u=x} \\
&\quad - 2\frac{p_0(F_{Y_0|X}^{-1}(\tau|x),x)}{f_X(x)}\frac{\partial^2}{\partial u^2} p_0(F_{Y_0|X}^{-1}(\tau|x),u)\Big|_{u=x} \\
&+ \frac{p_0(F_{Y_0|X}^{-1}(\tau|x),x)^2}{f_X(x)^2}\frac{\partial^2}{\partial x^2} f_X(x)\bigg]\,dx\,\mu(d\tau) w\,dw + \mathcal{O}\Big(h_y h_x^{1-\frac{d_X}{2}} \Big|\Big|\int \mathbf{K}(w)^2 w\,dw\Big|\Big|\Big) \\
&+ \mathcal{O}\big(h_y h_x^{2-\frac{d_X}{2}}\big) + \mathcal{O}\Big(h_x^{3-\frac{d_X}{2}}\Big) + \mathcal{O}\Big(h_y^2 h_x^{-\frac{d_X}{2}}\Big) + o_p(1) \\
&= b + \mathcal{O}\Big(h_x^{3-\frac{d_X}{2}} + h_y^2 h_x^{-\frac{d_X}{2}} + h_y h_x^{2-\frac{d_X}{2}}\Big) + \mathcal{O}\Big(h_y h_x^{1-\frac{d_X}{2}} \Big|\Big|\int \mathbf{K}(w)^2 w\,dw\Big|\Big|\Big) \\
&= b + o_p(1)
\end{aligned}
$$

by the bandwidth assumptions (2.19) and (2.20). Let $C > 0$ be a sufficiently large constant. Then, the variance of T_1 can be bounded by

$$
\begin{aligned}
&\operatorname{Var}\left(\frac{h_x^{\frac{d_X}{2}}}{n}\sum_{i=1}^n \int\int \kappa(x,\tau)\Big(\mathcal{K}_{h_y}(F_{Y_0|X}^{-1}(\tau|x) - g_{\beta_0}(X_i) - c_0 - \varepsilon_i) - \frac{p_0(F_{Y_0|X}^{-1}(\tau|x),x)}{f_X(x)}\Big)^2\right. \\
&\left.\mathbf{K}_{h_x}(x-X_i)^2\,dx\,\mu(d\tau)\right) \\
&\le \frac{h_x^{d_X}}{n} E\bigg[\Big(\int\int \kappa(x,\tau)\Big(\mathcal{K}_{h_y}(F_{Y_0|X}^{-1}(\tau|x) - g_{\beta_0}(X_1) - c_0 - \varepsilon_1) - \frac{p_0(F_{Y_0|X}^{-1}(\tau|x),x)}{f_X(x)}\Big)^2 \\
&\qquad \mathbf{K}_{h_x}(x-X_1)^2\,dx\,\mu(d\tau)\Big)^2\bigg] \\
&\le \frac{C h_x^{d_X}}{n} E\bigg[\Big(\int \mathbf{K}_{h_x}(x-X_1)^2\,dx\Big)^2\bigg] \\
&= \frac{C}{n h_x^{d_X}}\Big(\int \mathbf{K}(x)^2\,dx\Big)^2 \\
&= o(1),
\end{aligned}
$$

so that $T_1 = b + o_p(1)$.

Asymptotic Behaviour of T_2

Similar to Lemma 1.1.1, one has

$$
\begin{aligned}
&E\bigg[\mathbf{K}_{h_x}(x-X)\bigg(\mathcal{K}_{h_y}((F_{Y_0|X})^{-1}(\tau|x)-Y_0)-\frac{p_0(F^{-1}_{Y_0|X}(\tau|x),x)}{f_X(x)}\bigg)\bigg]\\
&=\int\int \mathbf{K}_{h_x}(x-w)\bigg(\mathcal{K}_{h_y}((F_{Y_0|X})^{-1}(\tau|x)-z)-\frac{p_0(F^{-1}_{Y_0|X}(\tau|x),x)}{f_X(x)}\bigg)f_{Y_0,X}(z,w)\,dz\,dw\\
&=\int \mathbf{K}(w)\bigg(\int \mathcal{K}_{h_y}(F^{-1}_{Y_0|X}(\tau|x)-z)f_{Y_0,X}(z,x-h_xw)\,dz\\
&\quad-\frac{p_0(F^{-1}_{Y_0|X}(\tau|x),x)}{f_X(x)}f_X(x-h_xw)\bigg)\,dw\\
&=\int \mathbf{K}(w)\bigg(\int\int_{-\infty}^{\frac{F^{-1}_{Y_0|X}(\tau|x)-z}{h_y}} K(u)f_{Y_0,X}(z,x-h_xw)\,du\,dz\\
&\quad-\frac{p_0(F^{-1}_{Y_0|X}(\tau|x),x)}{f_X(x)}f_X(x-h_xw)\bigg)\,dw\\
&=\int \mathbf{K}(w)\bigg(\int K(u)\int_{-\infty}^{(F_{Y_0|X})^{-1}(\tau|x)-h_yu} f_{Y_0,X}(z,x-h_xw)\,dz\,du\\
&\quad-\frac{p_0(F^{-1}_{Y_0|X}(\tau|x),x)}{f_X(x)}f_X(x-h_xw)\bigg)\,dw\\
&=\int \mathbf{K}(w)\bigg(\int K(u)p_0(F^{-1}_{Y_0|X}(\tau|x)-h_yu,x-h_xw)\,dz\,du\\
&\quad-p_0(F^{-1}_{Y_0|X}(\tau|x),x)\frac{f_X(x-h_xw)}{f_X(x)}\bigg)\,dw\\
&=\mathcal{O}\big(h_x^q\big)\\
&=o\bigg(\frac{1}{\sqrt{n}}\bigg)
\end{aligned}
\tag{2.48}
$$

uniformly in $x\in\operatorname{supp}(v)$. Therefore, the expectation of T_2 can be written as

$$
\begin{aligned}
E[T_2]=(n-1)h_x^{\frac{d_X}{2}}\int\int \kappa(x,\tau)E\bigg[\bigg(&\mathcal{K}_{h_y}(F^{-1}_{Y_0|X}(\tau|x)-g_{\beta_0}(X_1)-c_0-\varepsilon_1)\\
&-\frac{p_0(F^{-1}_{Y_0|X}(\tau|x),x)}{f_X(x)}\bigg)\mathbf{K}_{h_x}(x-X_1)\bigg]^2\,dx\,\mu(d\tau)\\
=o(1).\qquad&
\end{aligned}
$$

Define

$$
Z_i(x)=\bigg(\mathcal{K}_{h_y}(F^{-1}_{Y_0|X}(\tau|x)-g_{\beta_0}(X_i)-c_0-\varepsilon_i)-\frac{p_0(F^{-1}_{Y_0|X}(\tau|x),x)}{f_X(x)}\bigg)\mathbf{K}_{h_x}(x-X_i),
$$

so that

$$T_2 = \frac{h_x^{\frac{d_X}{2}}}{n} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \int \int \kappa(x,\tau) Z_i(x) Z_j(x) \, dx \, \mu(d\tau).$$

Later, Theorem 2.1 of De Jong (1987) will be used to show asymptotic normality of T_2. By the same reasoning as before, it can be proven that

$$E\big[Z_1(x)Z_2(x)Z_3(u)Z_4(u)\big] = o\bigg(\frac{1}{n^2}\bigg)$$

and

$$E\big[Z_1(x)Z_2(x)Z_2(u)Z_3(u)\big] = o\bigg(\frac{1}{nh_x^{d_X}}\bigg)$$

uniformly in $x, u \in \mathrm{supp}(v)$, which results in

$$\begin{aligned}
&E[T_2^2] \\
&= \frac{2(n-1)h_x^{d_X}}{n} E\bigg[\bigg(\int \int \kappa(x,\tau) Z_1(x) Z_2(x) \, dx \, \mu(d\tau)\bigg)^2\bigg] + o(1) \\
&= \frac{2(n-1)h_x^{d_X}}{n} E\bigg[\bigg(\int \int \kappa(x,\tau)\bigg(\mathcal{K}_{h_y}(F_{Y_0|X}^{-1}(\tau|x) - g_{\beta_0}(X_1) - c_0 - \varepsilon_1) \\
&\quad - \frac{p_0(F_{Y_0|X}^{-1}(\tau|x), x)}{f_X(x)}\bigg)\mathbf{K}_{h_x}(x - X_1)\bigg(\mathcal{K}_{h_y}(F_{Y_0|X}^{-1}(\tau|x) - g_{\beta_0}(X_2) - c_0 - \varepsilon_2) \\
&\quad - \frac{p_0(F_{Y_0|X}^{-1}(\tau|x), x)}{f_X(x)}\bigg)\mathbf{K}_{h_x}(x - X_2) \, dx \, \mu(d\tau)\bigg)^2\bigg] + o(1) \\
&= 2h_x^{-d_X} E\bigg[\bigg(\int \int \kappa(X_1 - h_x x, \tau)\bigg(\mathcal{K}_{h_y}(F_{Y_0|X}^{-1}(\tau|X_1 - h_x x) - g_{\beta_0}(X_1) - c_0 - \varepsilon_1) \\
&\quad - \frac{p_0(F_{Y_0|X}^{-1}(\tau|X_1 - h_x x), X_1 - h_x x)}{f_X(X_1 - h_x x)}\bigg)\mathbf{K}(x) \\
&\quad \bigg(\mathcal{K}_{h_y}(F_{Y_0|X}^{-1}(\tau|X_1 - h_x x) - g_{\beta_0}(X_2) - c_0 - \varepsilon_2) - \frac{p_0(F_{Y_0|X}^{-1}(\tau|X_1 - h_x x), X_1 - h_x x)}{f_X(X_1 - h_x x)}\bigg) \\
&\quad \mathbf{K}\bigg(x + \frac{X_1 - X_2}{h_x}\bigg) \, dx \, \mu(d\tau)\bigg)^2\bigg] + o(1) \\
&= 2h_x^{-d_X} \int \int \int \int \bigg(\int \int \kappa(w_1 - h_x x, \tau)\bigg(\mathcal{K}_{h_y}(F_{Y_0|X}^{-1}(\tau|w_1 - h_x x) - z_1) \\
&\quad - \frac{p_0(F_{Y_0|X}^{-1}(\tau|w_1 - h_x x), w_1 - h_x x)}{f_X(w_1 - h_x x)}\bigg)\mathbf{K}(x)\bigg(\mathcal{K}_{h_y}(F_{Y_0|X}^{-1}(\tau|w_1 - h_x x) - z_2) \\
&\quad - \frac{p_0(F_{Y_0|X}^{-1}(\tau|w_1 - h_x x), w_1 - h_x x)}{f_X(w_1 - h_x x)}\bigg)\mathbf{K}\bigg(x + \frac{w_1 - w_2}{h_x}\bigg) \, dx \, \mu(d\tau)\bigg)^2 f_{Y_0,X}(z_1, w_1) \\
&\quad f_{Y_0,X}(z_2, w_2) \, dw_1 \, dw_2 \, dz_1 \, dz_2 + o(1) \\
&= 2 \int \int \int \int \bigg(\int \int \kappa(w_1 - h_x x, \tau)\bigg(\mathcal{K}_{h_y}(F_{Y_0|X}^{-1}(\tau|w_1 - h_x x) - z_1)
\end{aligned}$$

$$- \frac{p_0(F^{-1}_{Y_0|X}(\tau|w_1 - h_x x), w_1 - h_x x)}{f_X(w_1 - h_x x)}\Bigg)\mathbf{K}(x)\Bigg(\mathcal{K}_{h_y}(F^{-1}_{Y_0|X}(\tau|w_1 - h_x x) - z_2)$$

$$- \frac{p_0(F^{-1}_{Y_0|X}(\tau|w_1 - h_x x), w_1 - h_x x)}{f_X(w_1 - h_x x)}\Bigg)\mathbf{K}(x + w_2)\, dx\, \mu(d\tau)\Bigg)^2 f_{Y_0,X}(z_1, w_1)$$

$$f_{Y_0,X}(z_2, w_1 - h_x w_2)\, dw_1\, dw_2\, dz_1\, dz_2 + o(1).$$

Note that $\mathcal{K}_{h_y}(F^{-1}_{Y_0|X}(\tau|w_1 - h_x x) - z_1) = I_{\{z_1 \le F^{-1}_{Y_0|X}(\tau|w_1)\}} + o(1)$ for Lebesgue all $w_1, x \in \mathbb{R}^{d_X}$, $z \in \mathbb{R}$ and μ-all $\tau \in (0,1)$, so that the dominated convergence theorem yields

$$E[T_2^2] = 2\int\int\int\int\Bigg(\int\int \kappa(w_1,\tau)\Bigg(I_{\{z_1 \le F^{-1}_{Y_0|X}(\tau|w_1)\}} - \frac{p_0(F^{-1}_{Y_0|X}(\tau|w_1), w_1)}{f_X(w_1)}\Bigg)\mathbf{K}(x)$$

$$\Bigg(I_{\{z_2 \le F^{-1}_{Y_0|X}(\tau|w_1)\}} - \frac{p_0(F^{-1}_{Y_0|X}(\tau|w_1), w_1)}{f_X(w_1)}\Bigg)\mathbf{K}(x + w_2)\, dx\, \mu(d\tau)\Bigg)^2$$

$$f_{Y_0,X}(z_1, w_1) f_{Y_0,X}(z_2, w_1)\, dw_1\, dw_2\, dz_1\, dz_2 + o(1)$$

$$= 2\int\int\int\Bigg(\int \kappa(w_1,\tau)\Bigg(I_{\{z_1 \le F^{-1}_{Y_0|X}(\tau|w_1)\}} - \frac{p_0(F^{-1}_{Y_0|X}(\tau|w_1), w_1)}{f_X(w_1)}\Bigg)$$

$$\Bigg(I_{\{z_2 \le F^{-1}_{Y_0|X}(\tau|w_1)\}} - \frac{p_0(F^{-1}_{Y_0|X}(\tau|w_1), w_1)}{f_X(w_1)}\Bigg)\mu(d\tau)\Bigg)^2$$

$$f_{Y_0,X}(z_1, w_1) f_{Y_0,X}(z_2, w_1)\, dw_1\, dz_1\, dz_2 \int\Bigg(\int \mathbf{K}(x)\mathbf{K}(x + w_2)\, dx\Bigg)^2 dw_2 + o(1). \tag{2.49}$$

Later, it will be shown, that the asymptotically non negligible term is equal to V. Define

$$W_{i,j} = 2\frac{h_x^{\frac{d_X}{2}}}{n}\int\int \kappa(x,\tau)\big(Z_i(x) - E[Z_1(x)]\big)\big(Z_j(x) - E[Z_1(x)]\big)\, dx\, \mu(d\tau).$$

Then,

$$W(n) := \sum_{i<j} W_{i,j} = T_2$$

is what De Jong (1987) called clean, that is $E[W_{i,j}|(Y_i, X_i)] = 0$ for all $i \neq j \in \{1, ..., n\}$. In (2.48), it was proven that $E[Z_1(x)] = o(\frac{1}{\sqrt{n}})$ uniformly in $\mathrm{supp}(v)$. Moreover, one can show $W(n) = T_2 + o_p(1)$ as well as $E[W(n)^2] = E[T_2^2] + o(1)$ similarly to before. Therefore,

$$\frac{\max_{i<j} E[W_{i,j}^2]}{E[W(n)^2]} = \frac{4h_x^{d_X} E\Big[\Big(\int\int \kappa(x,\tau) Z_1(x) Z_2(x)\, dx\, \mu(d\tau)\Big)^2\Big]}{nE[W(n)^2]} = \mathcal{O}\Big(\frac{1}{n}\Big) = o(1),$$

so that in order to prove normality of $W(n)$ and thus normality of T_2 it remains to show

$$\frac{E[W(n)^4]}{E[W(n)^2]^2} \to 3$$

(see Theorem 2.1 of De Jong (1987)). It holds that

$$E[W(n)^4] = \sum_{i<j}\sum_{k<l}\sum_{r<s}\sum_{t<u} E[W_{i,j} W_{k,l} W_{r,s} W_{t,u}]$$

$$= \frac{n(n-1)}{2} E[W_{1,2}^4] + \frac{3n(n-1)(n-2)(n-3)}{4} E[W_{1,2}^2]^2$$
$$+ 3n(n-1)(n-2)\frac{E[W_{1,2}^2 W_{2,3}^2]}{4} + 6n(n-1)(n-2)E[W_{1,2}W_{2,3}W_{3,1}^2]$$
$$+ 6n(n-1)(n-2)(n-3)E[W_{1,2}W_{2,3}W_{3,4}W_{4,1}], \tag{2.50}$$

where the prefactors are explained later.
In the following, consider

$$\tilde{W}_{i,j} = 2\frac{h_x^{\frac{d_X}{2}}}{n} \int\int \kappa(x,\tau) Z_i(x) Z_j(x)\, dx\, \mu(d\tau)$$

instead of $W_{i,j}$ as this makes calculations (a little bit) clearer and more convenient and the proof of the asymptotic negligibility of these replacements follows in a similar manner (for example $E[\tilde{W}_{1,2}^4] = E[W_{1,2}^4] + o(1)$).
First, one has for an appropriate constant $C > 0$

$$\begin{aligned}
n^2 E[\tilde{W}_{1,2}^4] &= \frac{16 h_x^{2d_X}}{n^2} E\bigg[\bigg(\int\int \kappa(x,\tau) Z_1(x) Z_2(x)\, dx\, \mu(d\tau)\bigg)^4\bigg] \\
&\leq \frac{C h_x^{2d_X}}{n^2} E\bigg[\bigg(\int\int \kappa(x,\tau) |\mathbf{K}_{h_x}(x - X_1)\mathbf{K}_{h_x}(x - X_2)|\, dx\, \mu(d\tau)\bigg)^4\bigg] \\
&= \frac{C}{n^2 h_x^{2d_X}} E\bigg[\bigg(\int\int \kappa(X_1 - h_x x, \tau)\bigg|\mathbf{K}(x)\mathbf{K}\Big(x + \frac{X_1 - X_2}{h_x}\Big)\bigg|\, dx\, \mu(d\tau)\bigg)^4\bigg] \\
&= o(1).
\end{aligned}$$

In equation (2.49) was shown that

$$\frac{n^4}{16} E[\tilde{W}_{1,2}^2]^2 = V^2 + o(1).$$

For a sufficiently large constant $C > 0$, $E[\tilde{W}_{1,2}^2 \tilde{W}_{2,3}^2]$ can be bounded by

$$\begin{aligned}
&n^3 E[\tilde{W}_{1,2}^2 \tilde{W}_{2,3}^2] \\
&\leq \frac{C h_x^{2d_X}}{n} E\bigg[\bigg(\int\int \kappa(x_1,\tau_1)\big|\mathbf{K}_{h_x}(x_1 - X_1)\mathbf{K}_{h_x}(x_1 - X_2)\big|\, dx_1\, \mu(d\tau_1)\bigg)^2 \\
&\qquad \bigg(\int\int \kappa(x_2,\tau_2)\big|\mathbf{K}_{h_x}(x_2 - X_3)\mathbf{K}_{h_x}(x_2 - X_2)\big|\, dx_2\, \mu(d\tau_2)\bigg)^2\bigg] \\
&= \frac{C}{n h_x^{2d_X}} E\bigg[\bigg(\int\int \kappa(X_1 - h_x x_1, \tau_1)\bigg|\mathbf{K}(x_1)\mathbf{K}\Big(x_1 + \frac{X_1 - X_2}{h_x}\Big)\bigg|\, dx_1\, \mu(d\tau_1)\bigg)^2 \\
&\qquad \bigg(\int\int \kappa(X_3 - h_x x_2, \tau_2)\bigg|\mathbf{K}(x_2)\mathbf{K}\Big(x_2 + \frac{X_3 - X_2}{h_x}\Big)\bigg|\, dx_2\, \mu(d\tau_2)\bigg)^2\bigg] \\
&= \frac{C}{n h_x^{2d_X}} \int\int\int \bigg(\int\int \kappa(w_1 - h_x x_1, \tau_1)\bigg|\mathbf{K}(x_1)\mathbf{K}\Big(x_1 + \frac{w_1 - w_2}{h_x}\Big)\bigg|\, dx_1\, \mu(d\tau_1)\bigg)^2 \\
&\qquad \bigg(\int\int \kappa(w_3 - h_x x_2, \tau_2)\bigg|\mathbf{K}(x_2)\mathbf{K}\Big(x_2 + \frac{w_3 - w_2}{h_x}\Big)\bigg|\, dx_2\, \mu(d\tau_2)\bigg)^2
\end{aligned}$$

$$f_X(w_1)f_X(w_2)f_X(w_3)\,dw_1\,dw_2\,dw_3$$

$$= \frac{C}{n}\int\int\int\Big(\int\int \kappa(w_2 + h_x w_1 - h_x x_1, \tau_1)\big|\mathbf{K}(x_1)\mathbf{K}(x_1 + w_1)\big|\,dx_1\,\mu(d\tau_1)\Big)^2$$

$$\Big(\int\int \kappa(w_2 + h_x w_3 - h_x x_2, \tau_2)\big|\mathbf{K}(x_2)\mathbf{K}(x_2 + w_3)\big|\,dx_2\,\mu(d\tau_2)\Big)^2$$

$$f_X(w_2 + h_x w_1)f_X(w_2)f_X(w_2 + h_x w_3)\,dw_1\,dw_2\,dw_3$$

$$= o(1).$$

$E[\tilde{W}_{1,2}\tilde{W}_{2,3}\tilde{W}_{3,1}^2]$ can be treated similar since

$$n^3E[\tilde{W}_{1,2}\tilde{W}_{2,3}\tilde{W}_{3,1}^2]$$

$$\leq \frac{Ch_x^{2d_X}}{n}E\Big[\int\int \kappa(x_1, \tau_1)\big|\mathbf{K}_{h_x}(x_1 - X_1)\mathbf{K}_{h_x}(x_1 - X_2)\big|\,dx_1\,\mu(d\tau_1)$$

$$\int\int \kappa(x_2, \tau_2)\big|\mathbf{K}_{h_x}(x_2 - X_2)\mathbf{K}_{h_x}(x_2 - X_3)\big|\,dx_2\,\mu(d\tau_2)$$

$$\Big(\int\int \kappa(x_3, \tau_3)\big|\mathbf{K}_{h_x}(x_3 - X_3)\mathbf{K}_{h_x}(x_3 - X_1)\big|\,dx_3\,\mu(d\tau_3)\Big)^2\Big]$$

$$= \frac{C}{nh_x^{2d_X}}E\Big[\int\int \kappa(X_1 - h_x x_1, \tau_1)\Big|\mathbf{K}(x_1)\mathbf{K}\Big(x_1 + \frac{X_1 - X_2}{h_x}\Big)\Big|\,dx_1\,\mu(d\tau_1)$$

$$\int\int \kappa(X_2 - h_x x_2, \tau_2)\Big|\mathbf{K}(x_2)\mathbf{K}\Big(x_2 + \frac{X_2 - X_3}{h_x}\Big)\Big|\,dx_2\,\mu(d\tau_2)$$

$$\Big(\int\int \kappa(X_3 - h_x x_3, \tau_3)\mathbf{K}(x_3)\Big|\mathbf{K}\Big(x_3 + \frac{X_3 - X_1}{h_x}\Big)\Big|\,dx_3\,\mu(d\tau_3)\Big)^2\Big]$$

$$\leq \frac{C^2}{nh_x^{2d_X}}\int\int\int\Big(\int\Big|\mathbf{K}(x_1)\mathbf{K}\Big(x_1 + \frac{w_1 - w_2}{h_x}\Big)\Big|\,dx_1\Big)$$

$$\Big(\int\Big|\mathbf{K}(x_2)\mathbf{K}\Big(x_2 + \frac{w_2 - w_3}{h_x}\Big)\Big|\,dx_2\Big)\Big(\int\Big|\mathbf{K}(x_3)\mathbf{K}\Big(x_3 + \frac{w_3 - w_1}{h_x}\Big)\Big|\,dx_3\Big)^2$$

$$f_X(w_1)f_X(w_2)f_X(w_3)\,dw_1\,dw_2\,dw_3$$

$$\leq \frac{C^3}{n}\int\int\int\Big(\int\big|\mathbf{K}(x_1)\mathbf{K}(x_1 + w_1)\big|\,dx_1\Big)\Big(\int\Big|\mathbf{K}(x_2)\mathbf{K}(x_2 + w_3)\Big|\,dx_2\Big)$$

$$f_X(w_2 + h_x w_1)f_X(w_2)f_X(w_2 - h_x w_3)\,dw_1\,dw_2\,dw_3$$

$$= o(1)$$

for an appropriate constant $C > 0$. It remains to consider $E[\tilde{W}_{1,2}\tilde{W}_{2,3}\tilde{W}_{3,4}\tilde{W}_{4,1}]$. This expectation can be treated by

$$n^4E[\tilde{W}_{1,2}\tilde{W}_{2,3}\tilde{W}_{3,4}\tilde{W}_{4,1}]$$

$$\leq Ch_x^{2d_X}E\Big[\int\int \kappa(x_1, \tau_1)\big|\mathbf{K}_{h_x}(x_1 - X_1)\mathbf{K}_{h_x}(x_1 - X_2)\big|\,dx_1\,\mu(d\tau_1)$$

$$\int\int \kappa(x_2,\tau_2)\big|\mathbf{K}_{h_x}(x_2-X_2)\mathbf{K}_{h_x}(x_2-X_3)\big|\,dx_2\,\mu(d\tau_2)$$
$$\int\int \kappa(x_3,\tau_3)\big|\mathbf{K}_{h_x}(x_3-X_3)\mathbf{K}_{h_x}(x_3-X_4)\big|\,dx_3\,\mu(d\tau_3)$$
$$\int\int \kappa(x_4,\tau_4)\big|\mathbf{K}_{h_x}(x_4-X_4)\mathbf{K}_{h_x}(x_4-X_1)\big|\,dx_4\,\mu(d\tau_4)\Big]$$
$$= \frac{C}{h_x^{2d_X}}E\Bigg[\int\int \kappa(X_1-h_xx_1,\tau_1)\bigg|\mathbf{K}(x_1)\mathbf{K}\Big(x_1+\frac{X_1-X_2}{h_x}\Big)\bigg|\,dx_1\,\mu(d\tau_1)$$
$$\int\int \kappa(X_2-h_xx_2,\tau_2)\bigg|\mathbf{K}(x_2)\mathbf{K}\Big(x_2+\frac{X_2-X_3}{h_x}\Big)\bigg|\,dx_2\,\mu(d\tau_2)$$
$$\int\int \kappa(X_3-h_xx_3,\tau_3)\mathbf{K}(x_3)\bigg|\mathbf{K}\Big(x_3+\frac{X_3-X_4}{h_x}\Big)\bigg|\,dx_3\,\mu(d\tau_3)$$
$$\int\int \kappa(X_4-h_xx_4,\tau_4)\mathbf{K}(x_4)\bigg|\mathbf{K}\Big(x_4+\frac{X_4-X_1}{h_x}\Big)\bigg|\,dx_4\,\mu(d\tau_4)\Bigg]$$
$$\leq \frac{C^2}{h_x^{2d_X}}\int\int\int\int\bigg(\int\bigg|\mathbf{K}(x_1)\mathbf{K}\Big(x_1+\frac{w_1-w_2}{h_x}\Big)\bigg|\,dx_1\bigg)$$
$$\bigg(\int\bigg|\mathbf{K}(x_2)\mathbf{K}\Big(x_2+\frac{w_2-w_3}{h_x}\Big)\bigg|\,dx_2\bigg)\bigg(\int\bigg|\mathbf{K}(x_3)\mathbf{K}\Big(x_3+\frac{w_3-w_4}{h_x}\Big)\bigg|\,dx_3\bigg)$$
$$\bigg(\int\bigg|\mathbf{K}(x_3)\mathbf{K}\Big(x_4+\frac{w_4-w_1}{h_x}\Big)\bigg|\,dx_4\bigg)f_X(w_1)f_X(w_2)f_X(w_3)f_X(w_4)\,dw_1\,dw_2\,dw_3\,dw_4$$
$$\leq C^3\int\int\int\int\bigg(\int\big|\mathbf{K}(x_1)\mathbf{K}(x_1+w_1)\big|\,dx_1\bigg)\bigg(\int\bigg|\mathbf{K}(x_2)\mathbf{K}\Big(x_2+\frac{w_2-w_3}{h_x}\Big)\bigg|\,dx_2\bigg)$$
$$\bigg(\int\big|\mathbf{K}(x_3)\mathbf{K}(x_3+w_4)\big|\,dx_3\bigg)$$
$$f_X(w_2+h_xw_1)f_X(w_2)f_X(w_3)f_X(w_3-h_xw_4)\,dw_1\,dw_2\,dw_3\,dw_4$$
$$= C^3h_x^{d_X}\int\int\int\int\bigg(\int\big|\mathbf{K}(x_1)\mathbf{K}(x_1+w_1)\big|\,dx_1\bigg)\bigg(\int\big|\mathbf{K}(x_2)\mathbf{K}(x_2+w_3)\big|\,dx_2\bigg)$$
$$\bigg(\int\big|\mathbf{K}(x_3)\mathbf{K}(x_3+w_4)\big|\,dx_3\bigg)$$
$$f_X(w_2+h_xw_1)f_X(w_2)f_X(w_2-h_xw_3)f_X(w_2-h_xw_3-h_xw_4)\,dw_1\,dw_2\,dw_3\,dw_4$$
$$= o(1).$$

Finally, this leads to

$$E[W(n)^4] = \frac{3n^4}{4}E[W_{1,2}^2]^2 + o(1) = 3V^2 + o(1) = 3E[T_2^2]^2 + o(1) = 3E[W(n)^2]^2 + o(1)$$

and thus $T_2 \overset{\mathcal{D}}{\to} \mathcal{N}(0,V)$.
Note that the prefactor of $\frac{3n(n-1)(n-2)(n-3)}{4} = \binom{n}{4}\cdot 3\cdot 6$ in (2.50) results from the fact that

- $\binom{n}{4}$ is the number of possibilities to choose a set of four indices out of $\{1,...,n\}$ (without ordering them),

- 3 is the number of possibilities to assign these indices to the corresponding four tuples $(i,j),(k,l),(r,s),(t,u)$ to obtain $E[W_{1,2}^2]^2$ and
- 6 is the number of possible permutations of these tuples.

The other prefactors in (2.50) can be derived similarly, but do not matter for the asymptotic behaviour of $\frac{E[T_2^4]}{E[T_2^2]^2}$. □

Rewriting b and V

The expressions for V and b given in (2.22) and (2.25), respectively, follow from (compare (2.7))

$$p_0(F^{-1}_{Y_0|X}(\tau|x),x) = F_\varepsilon(F^{-1}_{Y_0|X}(\tau|x)-g(x))f_X(x) = \tau f_X(x)$$

and

$$f_{Y,X}(F^{-1}_{Y_0|X}(\tau|x),x) = f_\varepsilon(F^{-1}_{Y_0|X}(\tau|x)-g(x))f_X(x) = f_\varepsilon(F_\varepsilon^{-1}(\tau))f_X(x).$$

To specify this, use the definition of $\kappa(x,\tau)$ in (2.18) and write under the assumptions (2.23) and (2.24)

$$\begin{aligned}
b &= h_x^{-\frac{d_X}{2}} \int \mathbf{K}(w)^2\,dw \int\int \kappa(x,\tau)p_0(F^{-1}_{Y_0|X}(\tau|x),x)\Big(1-\frac{p_0(F^{-1}_{Y_0|X}(\tau|x),x)}{f_X(x)}\Big)\,dx\,\mu(d\tau)\\
&\quad + o(1)\\
&= h_x^{-\frac{d_X}{2}} \int \mathbf{K}(w)^2\,dw \int\int \frac{v(x)}{f_\varepsilon(F_\varepsilon^{-1}(\tau))^2 f_X(x)}\tau(1-\tau)\,dx\,\mu(d\tau)+o(1)\\
&= h_x^{-\frac{d_X}{2}} \int \mathbf{K}(w)^2\,dw \int \frac{v(x)}{f_X(x)}\,dx \int \frac{\tau(1-\tau)}{f_\varepsilon(F_\varepsilon^{-1}(\tau))^2}\,\mu(d\tau)+o(1)
\end{aligned}$$

and (see (2.49))

$$\begin{aligned}
V &= 2\int\Big(\int \mathbf{K}(x)\mathbf{K}(x+s)\,dx\Big)^2 ds \int\int\int\Big(\int \kappa(w,\tau)\\
&\quad \Big(I_{\{z_1\le F^{-1}_{Y_0|X}(\tau|w)\}}-\frac{p_0(F^{-1}_{Y_0|X}(\tau|w),w)}{f_X(w)}\Big)\Big(I_{\{z_2\le F^{-1}_{Y_0|X}(\tau|w)\}}-\frac{p_0(F^{-1}_{Y_0|X}(\tau|w),w)}{f_X(w)}\Big)\\
&\quad \mu(d\tau)\Big)^2 f_{Y_0,X}(z_1,w)f_{Y_0,X}(z_2,w)\,dw\,dz_1\,dz_2\\
&= 2\int\Big(\int \mathbf{K}(x)\mathbf{K}(x+s)\,dx\Big)^2 ds \int\int\int\Big(\int \kappa(w,\tau)\big(I_{\{z_1-g(w)\le F_\varepsilon^{-1}(\tau)\}}-\tau\big)\\
&\quad \big(I_{\{z_2-g(w)\le F_\varepsilon^{-1}(\tau)\}}-\tau\big)\,\mu(d\tau)\Big)^2 f_\varepsilon(z_1-g(w))f_\varepsilon(z_2-g(w))\,dz_1\,dz_2\,f_X(w)^2\,dw\\
&= 2\int\Big(\int \mathbf{K}(x)\mathbf{K}(x+s)\,dx\Big)^2 ds \int \frac{v(w)^2}{f_X(w)^2}\int\int\Big(\int \frac{\big(I_{\{F_\varepsilon(z_1-g(w))\le\tau\}}-\tau\big)}{f_\varepsilon(F_\varepsilon^{-1}(\tau))}\\
&\quad \frac{\big(I_{\{F_\varepsilon(z_2-g(w))\le\tau\}}-\tau\big)}{f_\varepsilon(F_\varepsilon^{-1}(\tau))}\,\mu(d\tau)\Big)^2 f_\varepsilon(z_1-g(w))f_\varepsilon(z_2-g(w))\,dz_1\,dz_2\,dw
\end{aligned}$$

$$= 2\int\Big(\int \mathbf{K}(x)\mathbf{K}(x+s)\,dx\Big)^2 ds \int \frac{v(w)^2}{f_X(w)^2}\,dw$$
$$\int_0^1\int_0^1\Big(\int \frac{\big(I_{\{u_1\leq\tau\}}-\tau\big)\big(I_{\{u_2\leq\tau\}}-\tau\big)}{f_\varepsilon(F_\varepsilon^{-1}(\tau))^2}\,\mu(d\tau)\Big)^2 du_1\,du_2.$$

□

2.8.3 Proof of Theorem 2.3.4

Later, it will be shown, that the test statistic T_n defined in (2.13) is asymptotically equivalent to $\tilde{T}_n+\delta_{2,n}+\delta_{3,n}$, where

$$\tilde{T}_n = nh_x^{\frac{d_X}{2}}\int\int v(x)\big(\hat{F}^{-1}_{Y|X}(\tau|x)-F^{-1}_{Y_0|X}(\tau|x)\big)^2\,dx\,\mu(d\tau),$$

$$\delta_{2,n} = -\Big(\int v(x)D_\beta g_{\beta_0}(x)\Delta_n(x)\,dx\Big)\Omega^{-1}\Big(\int v(x)D_\beta g_{\beta_0}(x)\Delta_n(x)\,dx\Big)^t$$

and

$$\delta_{3,n} = -\mu([0,1])\frac{\big(\int v(x)\Delta_n(x)\,dx\big)^2}{\int v(w)\,dw}.$$

$\tilde{T}_n$ in turn can be split into

$$\begin{aligned}\tilde{T}_n &= nh_x^{\frac{d_X}{2}}\int\int v(x)\big(\hat{F}^{-1}_{Y|X}(\tau|x)-\hat{F}^{-1}_{Y_0|X}(\tau|x)+\hat{F}^{-1}_{Y_0|X}(\tau|x)-F^{-1}_{Y_0|X}(\tau|x)\big)^2\,dx\,\mu(d\tau)\\
&= nh_x^{\frac{d_X}{2}}\int\int v(x)\big(\hat{F}^{-1}_{Y|X}(\tau|x)-\hat{F}^{-1}_{Y_0|X}(\tau|x)\big)^2\,dx\,\mu(d\tau)\\
&\quad + nh_x^{\frac{d_X}{2}}\int\int v(x)\big(\hat{F}^{-1}_{Y_0|X}(\tau|x)-F^{-1}_{Y_0|X}(\tau|x)\big)^2\,dx\,\mu(d\tau)\\
&\quad + 2nh_x^{\frac{d_X}{2}}\int\int v(x)\big(\hat{F}^{-1}_{Y|X}(\tau|x)-\hat{F}^{-1}_{Y_0|X}(\tau|x)\big)\big(\hat{F}^{-1}_{Y_0|X}(\tau|x)-F^{-1}_{Y_0|X}(\tau|x)\big)\,dx\,\mu(d\tau)\\
&= T_1+T_2+T_3.\end{aligned}$$

While Lemma 2.3.2 can be applied for T_2 to obtain

$$nh_x^{\frac{d_X}{2}}\int\int v(x)\big(\hat{F}^{-1}_{Y_0|X}(\tau|x)-F^{-1}_{Y_0|X}(\tau|x)\big)^2\,dx\,\mu(d\tau)-b\xrightarrow{\mathcal{D}} Z$$

with $Z\sim\mathcal{N}(0,V)$ and b as well as V from Lemma 2.3.2, T_1 can be treated as follows. Remember

$$\kappa(x,\tau)=\frac{v(x)}{f_{Y_0|X}(F^{-1}_{Y_0|X}(\tau|x)|x)^2 f_X(x)^2}$$

as well as (2.46) and write

$$\begin{aligned}&nh_x^{\frac{d_X}{2}}\int\int v(x)\big(\hat{F}^{-1}_{Y|X}(\tau|x)-\hat{F}^{-1}_{Y_0|X}(\tau|x)\big)^2\,dx\,\mu(d\tau)\\
&= nh_x^{\frac{d_X}{2}}\int\int \kappa(x,\tau)\Big(\frac{c_n}{n}\sum_{i=1}^n K_{h_y}(F^{-1}_{Y_0|X}(\tau|x)-g_{\beta_0}(X_i)-c_0-\varepsilon_i)\mathbf{K}_{h_x}(x-X_i)\Delta_n(X_i)\end{aligned}$$

$$+ o_p\left(\frac{1}{\sqrt{n}}\right)\Big)^2 dx\,\mu(d\tau)$$

$$\leq \frac{c_n^2 h_x^{\frac{d_X}{2}}}{n}\sum_{i=1}^n\sum_{j=1}^n\int\int\kappa(x,\tau)K_{h_y}(F^{-1}_{Y_0|X}(\tau|x)-g_{\beta_0}(X_i)-c_0-\varepsilon_i)\mathbf{K}_{h_x}(x-X_i)$$

$$K_{h_y}(F^{-1}_{Y_0|X}(\tau|x)-g_{\beta_0}(X_j)-c_0-\varepsilon_j)\mathbf{K}_{h_x}(x-X_j)\Delta_n(X_i)\Delta_n(X_j)\,dx\,\mu(d\tau)$$

$$+\frac{o_p\big(\sqrt{n}c_n h_x^{\frac{d_X}{2}}\big)}{n}\sum_{i=1}^n\int\int\kappa(x,\tau)\Big|K_{h_y}(F^{-1}_{Y_0|X}(\tau|x)-g_{\beta_0}(X_i)-c_0-\varepsilon_i)$$

$$\mathbf{K}_{h_x}(x-X_i)\Delta_n(X_i)\Big|\,dx\,\mu(d\tau)+o_p(1)$$

$$=\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\int\int\kappa(x,\tau)K_{h_y}(F^{-1}_{Y_0|X}(\tau|x)-g_{\beta_0}(X_i)-c_0-\varepsilon_i)\mathbf{K}_{h_x}(x-X_i)\mathbf{K}_{h_x}(x-X_j)$$

$$K_{h_y}(F^{-1}_{Y_0|X}(\tau|x)-g_{\beta_0}(X_j)-c_0-\varepsilon_j)\Delta_n(X_i)\Delta_n(X_j)\,dx\,\mu(d\tau)$$

$$+o_p(1)$$

by the definition of c_n in (**A2**). Then, (2.42) leads to $nh_x^{d_X}h_y\to\infty$ and thus

$$\frac{1}{n}E\Big[\int\int\kappa(x,\tau)K_{h_y}(F^{-1}_{Y_0|X}(\tau|x)-g_{\beta_0}(X_1)-c_0-\varepsilon_1)^2\mathbf{K}_{h_x}(x-X_1)^2\Delta_n(X_1)^2\,dx\,\mu(d\tau)\Big]$$

$$=\frac{1}{n}\int\int\int\int\kappa(x,\tau)K_{h_y}(F^{-1}_{Y_0|X}(\tau|x)-g_{\beta_0}(w)-c_0-e)^2\mathbf{K}_{h_x}(x-w)^2\Delta_n(w)^2$$

$$f_X(w)f_\varepsilon(e)\,dw\,de\,dx\,\mu(d\tau)$$

$$=\frac{1}{nh_x^{d_X}}\int\int\int\int\kappa(x,\tau)K_{h_y}(F^{-1}_{Y_0|X}(\tau|x)-g_{\beta_0}(x-h_xw)-c_0-e)^2\mathbf{K}(w)^2\Delta_n(x-h_xw)^2$$

$$f_X(x-h_xw)f_\varepsilon(e)\,dw\,de\,dx\,\mu(d\tau)$$

$$=\frac{1}{nh_x^{d_X}h_y}\int\int\int\int\kappa(x,\tau)K(e)^2\mathbf{K}(w)^2\Delta_n(x-h_xw)^2$$

$$f_X(x-h_xw)f_\varepsilon(F^{-1}_{Y_0|X}(\tau|x)-g_{\beta_0}(x-h_xw)-c_0-h_ye)\,dw\,de\,dx\,\mu(d\tau)$$

$$=o(1),$$

that is

$$T_1=\frac{1}{n^2}\sum_{i=1}^n\sum_{\substack{j=1\\ j\neq i}}^n\int\int\kappa(x,\tau)K_{h_y}(F^{-1}_{Y_0|X}(\tau|x)-g_{\beta_0}(X_i)-c_0-\varepsilon_i)\mathbf{K}_{h_x}(x-X_i)$$

$$K_{h_y}(F^{-1}_{Y_0|X}(\tau|x)-g_{\beta_0}(X_j)-c_0-\varepsilon_j)\mathbf{K}_{h_x}(x-X_j)\Delta_n(X_i)\Delta_n(X_j)\,dx\,\mu(d\tau)+o_p(1).$$

Due to

$$E\big[K_{h_y}(F^{-1}_{Y_0|X}(\tau|x)-g_{\beta_0}(X_1)-c_0-\varepsilon_1)\mathbf{K}_{h_x}(x-X_1)\Delta_n(X_1)\big]$$

$$
\begin{aligned}
&= \int\int K_{h_y}(F_{Y_0|X}^{-1}(\tau|x) - g_{\beta_0}(w) - c_0 - e)\mathbf{K}_{h_x}(x-w)\Delta_n(w)f_\varepsilon(e)f_X(w)\,de\,dw \\
&= \int\int K(e)\mathbf{K}(w)\Delta_n(x-h_xw)f_\varepsilon(F_{Y_0|X}^{-1}(\tau|x) - g_{\beta_0}(x-h_xw) - c_0 - h_ye) \\
&\qquad f_X(x-h_xw)\,de\,dw \\
&= \Delta_n(x)f_\varepsilon(F_{Y_0|X}^{-1}(\tau|x) - g_{\beta_0}(x) - c_0)f_X(x) + o(1)
\end{aligned}
$$

uniformly in $x \in \operatorname{supp}(v)$ and $\tau \in \operatorname{supp}(\mu)$, the expectation of T_1 can be written as

$$
\begin{aligned}
E[T_1] &= \int\int \kappa(x,\tau)E\big[K_{h_y}(F_{Y_0|X}^{-1}(\tau|x) - g_{\beta_0}(X_1) - c_0 - \varepsilon_1) \\
&\qquad \mathbf{K}_{h_x}(x-X_1)\Delta_n(X_1)\big]^2\,dx\,\mu(d\tau) + o(1) \\
&= \delta_{1,n} + o(1)
\end{aligned}
$$

with

$$
\begin{aligned}
\delta_{1,n} &= \int\int \kappa(x,\tau)\Delta_n(x)^2 f_\varepsilon\big(F_{Y_0|X}^{-1}(\tau|x) - g_{\beta_0}(x) - c_0\big)^2 f_X^2(x)\,dx\,\mu(d\tau) \\
&= \int\int v(x)\Delta_n(x)^2\,dx\,\mu(d\tau) \\
&= \mu([0,1])\int v(x)\Delta_n(x)^2\,dx.
\end{aligned}
$$

Here, the definition of $\kappa(x,\tau)$ and the fact were used that (compare (2.7))

$$
f_{Y_0|X}(F_{Y_0|X}^{-1}(\tau|x)|x) = f_\varepsilon(F_{Y_0|X}^{-1}(\tau|x) - g_{\beta_0}(x) - c_0) = f_\varepsilon(F_\varepsilon^{-1}(\tau)).
$$

In the following, it is shown that the variance of the asymptotically nonnegligible terms converges to zero. For reasons of clarity and comprehensibility, define

$$
\begin{aligned}
Z_{i,j} = \int\int &\kappa(x,\tau)\mathbf{K}_{h_x}(x-X_i)\mathbf{K}_{h_x}(x-X_j)K_{h_y}(F_{Y_0|X}^{-1}(\tau|x) - g_{\beta_0}(X_i) - c_0 - \varepsilon_i) \\
&K_{h_y}(F_{Y_0|X}^{-1}(\tau|x) - g_{\beta_0}(X_j) - c_0 - \varepsilon_j)\Delta_n(X_i)\Delta_n(X_j)\,dx\,\mu(d\tau),
\end{aligned}
$$

so that $T_1 = \frac{2}{n^2}\sum_{i=1}^n\sum_{j=i+1}^n Z_{i,j} + o_p(1)$. To show that the variance of T_1 converges to zero, write

$$
\begin{aligned}
&\operatorname{Var}\left(\frac{2}{n^2}\sum_{i=1}^n\sum_{j=i+1}^n Z_{i,j}\right) \\
&= \frac{4}{n^4}\sum_{i=1}^n\sum_{j=i+1}^n\sum_{k=1}^n\sum_{l=k+1}^n \operatorname{Cov}(Z_{i,j}, Z_{k,l}) \\
&= \frac{4}{n^4}\sum_{i=1}^n\sum_{j=i+1}^n \operatorname{Var}(Z_{i,j}) + \frac{4}{n^4}\sum_{i=1}^n\sum_{j=i+1}^n\sum_{\substack{l=i+1\\ l\neq j}}^n \operatorname{Cov}(Z_{i,j}, Z_{i,l}) \\
&\quad + \frac{4}{n^4}\sum_{k=1}^n\sum_{i=k+1}^n\sum_{j=i+1}^n \operatorname{Cov}(Z_{i,j}, Z_{k,i}) + \frac{4}{n^4}\sum_{i=1}^n\sum_{j=i+1}^n\sum_{l=j+1}^n \operatorname{Cov}(Z_{i,j}, Z_{j,l})
\end{aligned}
$$

$$+\frac{4}{n^4}\sum_{j=1}^{n}\sum_{i=j-1}^{n}\sum_{\substack{k=j-1\\k\neq i}}^{n}\mathrm{Cov}(Z_{i,j},Z_{k,j}),$$

so that it suffices to prove that

$$E[Z_{1,2}^2], E[|Z_{1,2}Z_{1,3}|], E[|Z_{1,2}Z_{2,3}|] \quad \text{and} \quad E[|Z_{1,3}Z_{2,3}|] \tag{2.51}$$

converge to zero. For an appropriate constant $C>0$ it holds that

$$\begin{aligned}
&\frac{E[Z_{1,2}^2]}{n^2}\\
&\leq \frac{C}{n^2}E\Bigg[\bigg(\int\int \kappa(x,\tau)\Big|K_{h_y}(F^{-1}_{Y_0|X}(\tau|x)-g_{\beta_0}(X_1)-c_0-\varepsilon_1)\\
&\quad K_{h_y}(F^{-1}_{Y_0|X}(\tau|x)-g_{\beta_0}(X_2)-c_0-\varepsilon_2)\mathbf{K}_{h_x}(x-X_1)\mathbf{K}_{h_x}(x-X_2)\Big|\,dx\,\mu(d\tau)\bigg)^2\Bigg]\\
&=\frac{C}{n^2h_x^{2d_X}}E\Bigg[\bigg(\int\int \kappa(X_1-h_xx,\tau)\Big|K_{h_y}(F^{-1}_{Y_0|X}(\tau|X_1-h_xx)-g_{\beta_0}(X_1)-c_0-\varepsilon_1)\\
&\quad K_{h_y}(F^{-1}_{Y_0|X}(\tau|X_1-h_xx)-g_{\beta_0}(X_2)-c_0-\varepsilon_2)\mathbf{K}(x)\mathbf{K}\Big(x+\frac{X_1-X_2}{h_x}\Big)\Big|\,dx\,\mu(d\tau)\bigg)^2\Bigg]\\
&=\frac{C}{n^2h_x^{2d_X}}\int\int\int\int\bigg(\int\int \kappa(w_1-h_xx,\tau)\Big|K_{h_y}(F^{-1}_{Y_0|X}(\tau|w_1-h_xx)-g_{\beta_0}(w_1)-c_0-e_1)\\
&\quad K_{h_y}(F^{-1}_{Y_0|X}(\tau|w_1-h_xx)-g_{\beta_0}(w_2)-c_0-e_2)\mathbf{K}(x)\mathbf{K}\Big(x+\frac{w_1-w_2}{h_x}\Big)\Big|\,dx\,\mu(d\tau)\bigg)^2\\
&\quad f_X(w_1)f_X(w_2)f_\varepsilon(e_1)f_\varepsilon(e_2)\,dw_1\,dw_2\,de_1\,de_2\\
&=\frac{C}{n^2h_x^{d_X}}\int\int\int\int\bigg(\int\int \kappa(w_1-h_xx,\tau)\Big|K_{h_y}(F^{-1}_{Y_0|X}(\tau|w_1-h_xx)-g_{\beta_0}(w_1)-c_0-e_1)\\
&\quad K_{h_y}(F^{-1}_{Y_0|X}(\tau|w_1-h_xx)-g_{\beta_0}(w_1-h_xw_2)-c_0-e_2)\mathbf{K}(x)\mathbf{K}(x+w_2)\Big|\,dx\,\mu(d\tau)\bigg)^2\\
&\quad f_X(w_1)f_X(w_1-h_xw_2)f_\varepsilon(e_1)f_\varepsilon(e_2)\,dw_1\,dw_2\,de_1\,de_2\\
&=\frac{C}{n^2h_x^{d_X}}\int\int\int\int\int\int\int\int\\
&\quad\bigg(\kappa(w_1-h_xx_1,\tau_1)\Big|K_{h_y}(F^{-1}_{Y_0|X}(\tau_1|w_1-h_xx)-g_{\beta_0}(w_1)-c_0-e_1)\\
&\quad K_{h_y}(F^{-1}_{Y_0|X}(\tau_1|w_1-h_xx_1)-g_{\beta_0}(w_1-h_xw_2)-c_0-e_2)\mathbf{K}(x_1)\mathbf{K}(x_1+w_2)\Big|\bigg)\\
&\quad\bigg(\kappa(w_1-h_xx_2,\tau_2)\Big|K_{h_y}(F^{-1}_{Y_0|X}(\tau_2|w_1-h_xx_2)-g_{\beta_0}(w_1)-c_0-e_1)\\
&\quad K_{h_y}(F^{-1}_{Y_0|X}(\tau_2|w_1-h_xx_2)-g_{\beta_0}(w_1-h_xw_2)-c_0-e_2)\mathbf{K}(x_2)\mathbf{K}(x_2+w_2)\Big|\bigg)\\
&\quad f_X(w_1)f_X(w_1-h_xw_2)f_\varepsilon(e_1)f_\varepsilon(e_2)\,dw_1\,dw_2\,de_1\,de_2\,dx_1\,dx_2\,\mu(d\tau_1)\,\mu(d\tau_2)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{C^2}{n^2 h_x^{d_X} h_y^2} \int\int\int\int\int\int \Big(\kappa(w_1 - h_x x_1, \tau_1) \Big| K(e_1)K(e_2)\mathbf{K}(x_1)\mathbf{K}(x_1 + w_2) \Big| \Big) \\
&\qquad f_X(w_1) f_X(w_1 - h_x w_2) f_\varepsilon(F^{-1}_{Y_0|X}(\tau_1|w_1 - h_x x) - g_{\beta_0}(w_1) - h_y e_1) \\
&\qquad f_\varepsilon(F^{-1}_{Y_0|X}(\tau_1|w_1 - h_x x_1) - g_{\beta_0}(w_1 - h_x w_2) - h_y e_2)\, dw_1\, dw_2\, de_1\, de_2\, dx_1\, \mu(d\tau_1) \\
&\leq \frac{C^2}{n^2 h_x^{d_X} h_y^2} \int\int\int\int\int\int \Big| K(e_1)K(e_2)\mathbf{K}(x_1)\mathbf{K}(x_1 + w_2) \Big| f_X(w_1)\, dx_1\, dw_1\, dw_2\, de_1\, de_2 \\
&= o(1).
\end{aligned}
$$

Again for an appropriate constant $C > 0$, the second expectation in (2.51) can be written as

$$
\begin{aligned}
&\frac{E[|Z_{1,2}Z_{1,3}|]}{n} \\
&\leq \frac{C}{n h_x^{2d_X}} E\Bigg[\Bigg(\int\int \kappa(x,\tau) \Big| K_{h_y}(F^{-1}_{Y_0|X}(\tau|X_1 - h_x x) - g_{\beta_0}(X_1) - c_0 - \varepsilon_1) \\
&\qquad K_{h_y}(F^{-1}_{Y_0|X}(\tau|X_1 - h_x x) - g_{\beta_0}(X_2) - c_0 - \varepsilon_2)\mathbf{K}(x)\mathbf{K}\Big(x + \frac{X_1 - X_2}{h_x} \Big) \Big|\, dx\, \mu(d\tau) \Bigg) \\
&\qquad \Bigg(\int\int \kappa(x,\tau) \Big| K_{h_y}(F^{-1}_{Y_0|X}(\tau|X_1 - h_x x) - g_{\beta_0}(X_1) - c_0 - \varepsilon_1) \\
&\qquad K_{h_y}(F^{-1}_{Y_0|X}(\tau|X_1 - h_x x) - g_{\beta_0}(X_1) - c_0 - \varepsilon_3)\mathbf{K}(x)\mathbf{K}\Big(x + \frac{X_1 - X_3}{h_x} \Big) \Big|\, dx\, \mu(d\tau) \Bigg) \Bigg] \\
&= \frac{C}{n h_x^{2d_X}} \int\int\int\int\int\int\int\int\int\int \kappa(x_1,\tau_1)\kappa(x_2,\tau_2) \\
&\qquad \Big| K_{h_y}(F^{-1}_{Y_0|X}(\tau_1|w_1 - h_x x_1) - g_{\beta_0}(w_1) - c_0 - e_1) \\
&\qquad K_{h_y}(F^{-1}_{Y_0|X}(\tau_1|w_1 - h_x x_1) - g_{\beta_0}(w_2) - c_0 - e_2)\mathbf{K}(x_1)\mathbf{K}\Big(x_1 + \frac{w_1 - w_2}{h_x} \Big) \\
&\qquad K_{h_y}(F^{-1}_{Y_0|X}(\tau_2|w_1 - h_x x_2) - g_{\beta_0}(w_1) - c_0 - e_1) \\
&\qquad K_{h_y}(F^{-1}_{Y_0|X}(\tau_2|w_1 - h_x x_2) - g_{\beta_0}(w_3) - c_0 - e_3)\mathbf{K}(x_2)\mathbf{K}\Big(x_2 + \frac{w_1 - w_3}{h_x} \Big) \Big| f_X(w_1) \\
&\qquad f_X(w_2) f_X(w_3) f_\varepsilon(e_1) f_\varepsilon(e_2) f_\varepsilon(e_3)\, dw_1\, dw_2\, dw_3\, de_1\, de_2\, de_3\, dx_1\, dx_2\, \mu(d\tau_1)\, \mu(d\tau_2) \\
&\leq \frac{C^2}{n h_y h_x^{2d_X}} \int\int\int\int\int\int\int\int \Big| K(e_1)K(e_2)K(e_3)\mathbf{K}(x_1)\mathbf{K}\Big(x_1 + \frac{w_1 - w_2}{h_x} \Big) \\
&\qquad \mathbf{K}(x_2)\mathbf{K}\Big(x_2 + \frac{w_1 - w_3}{h_x} \Big) \Big| f_X(w_1) f_X(w_2) f_X(w_3)\, dw_1\, dw_2\, dw_3\, de_1\, de_2\, de_3\, dx_1\, dx_2 \\
&\leq \frac{C^3}{n h_y} \int\int\int \Big(\int |\mathbf{K}(x_1)\mathbf{K}(x_1 + w_2)|\, dx_1 \Big) \Big(\int |\mathbf{K}(x_2)\mathbf{K}(x_2 + w_3)|\, dx_2 \Big) \\
&\qquad f_X(w_1) f_X(w_1 - h_x w_2) f_X(w_1 - h_x w_3)\, dw_1\, dw_2\, dw_3
\end{aligned}
$$

$$= o(1).$$

Similarly,

$$\frac{E[|Z_{1,2}Z_{2,3}|]}{n}$$
$$\leq \frac{C}{nh_x^{2d_X}} E\Bigg[\bigg(\int\int \kappa(x,\tau)\bigg| K_{h_y}(F^{-1}_{Y_0|X}(\tau|X_1 - h_x x) - g_{\beta_0}(X_1) - c_0 - \varepsilon_1)$$
$$K_{h_y}(F^{-1}_{Y_0|X}(\tau|X_1 - h_x x) - g_{\beta_0}(X_2) - c_0 - \varepsilon_2)\mathbf{K}(x)\mathbf{K}\bigg(x + \frac{X_1 - X_2}{h_x}\bigg)\bigg| \, dx\, \mu(d\tau)\bigg)$$
$$\bigg(\int\int \kappa(x,\tau)\bigg| K_{h_y}(F^{-1}_{Y_0|X}(\tau|X_2 - h_x x) - g_{\beta_0}(X_2) - c_0 - \varepsilon_2)$$
$$K_{h_y}(F^{-1}_{Y_0|X}(\tau|X_2 - h_x x) - g_{\beta_0}(X_3) - c_0 - \varepsilon_3)\mathbf{K}(x)\mathbf{K}\bigg(x + \frac{X_2 - X_3}{h_x}\bigg)\bigg| \, dx\, \mu(d\tau)\bigg)\Bigg]$$
$$= \frac{C}{nh_x^{2d_X}} \int\int\int\int\int\int \bigg(\int\int \kappa(x,\tau)\bigg| K_{h_y}(F^{-1}_{Y_0|X}(\tau|w_1 - h_x x) - g_{\beta_0}(w_1) - c_0 - e_1)$$
$$K_{h_y}(F^{-1}_{Y_0|X}(\tau|w_1 - h_x x) - g_{\beta_0}(w_2) - c_0 - e_2)\mathbf{K}(x)\mathbf{K}\bigg(x + \frac{w_1 - w_2}{h_x}\bigg)\bigg| \, dx\, \mu(d\tau)\bigg)$$
$$\bigg(\int\int \kappa(x,\tau)\bigg| K_{h_y}(F^{-1}_{Y_0|X}(\tau|w_2 - h_x x) - g_{\beta_0}(w_2) - c_0 - e_2)$$
$$K_{h_y}(F^{-1}_{Y_0|X}(\tau|w_2 - h_x x) - g_{\beta_0}(w_3) - c_0 - e_3)\mathbf{K}(x)\mathbf{K}\bigg(x + \frac{w_2 - w_3}{h_x}\bigg)\bigg| \, dx\, \mu(d\tau)\bigg)$$
$$f_X(w_1)f_X(w_2)f_X(w_3)f_\varepsilon(e_1)f_\varepsilon(e_2)f_\varepsilon(e_3)\, dw_1\, dw_2\, dw_3\, de_1\, de_2\, de_3$$
$$\leq \frac{C^2}{nh_y}\int\int\int \bigg(\int |\mathbf{K}(x)\mathbf{K}(x + w_1)|\, dx\bigg)\bigg(\int |\mathbf{K}(x)\mathbf{K}(x + w_3)|\, dx\bigg)$$
$$f_X(w_2 + h_x w_1)f_X(w_2)f_X(w_2 - h_x w_3)\, dw_1\, dw_2\, dw_3$$
$$= o(1)$$

and

$$\frac{E[|Z_{1,3}Z_{2,3}|]}{n} \leq \frac{C^3}{nh_y}\int\int\int \bigg(\int |\mathbf{K}(x)\mathbf{K}(x + w_2)|\, dx\bigg)\bigg(\int |\mathbf{K}(x)\mathbf{K}(x + w_3)|\, dx\bigg)$$
$$f_X(w_3 + h_x w_1)f_X(w_2 + h_x w_3)f_X(w_3)\, dw_1\, dw_2\, dw_3$$
$$= o(1).$$

In total,

$$T_1 = \delta_{1,n} + o_p(1)$$

has been proven, so that only T_3 is left to be examined. Inserting equations (2.45) and (2.46) yields

$$T_3 = 2nh_x^{\frac{d_X}{2}} \int\int v(x)\big(\hat{F}^{-1}_{Y|X}(\tau|x) - \hat{F}^{-1}_{Y_0|X}(\tau|x)\big)\big(\hat{F}^{-1}_{Y_0|X}(\tau|x) - F^{-1}_{Y_0|X}(\tau|x)\big)\, dx\, \mu(d\tau)$$

$$
\begin{aligned}
&= 2nh_x^{\frac{d_X}{2}} \int\int \kappa(x,\tau)\Big(\frac{c_n}{n}\sum_{i=1}^n K_{h_y}(F^{-1}_{Y_0|X}(\tau|x)-g_{\beta_0}(X_i)-c_0-\varepsilon_i)\mathbf{K}_{h_x}(x-X_i)\Delta_n(X_i)\\
&\quad + o_p\Big(\frac{1}{\sqrt{n}}\Big)\Big)\Big(\frac{1}{n}\sum_{i=1}^n\Big(\mathcal{K}_{h_y}(F^{-1}_{Y_0|X}(\tau|x)-g_{\beta_0}(X_i)-c_0-\varepsilon_i)\\
&\quad - \frac{p_0(F^{-1}_{Y_0|X}(\tau|x),x)}{f_X(x)}\Big)\mathbf{K}_{h_x}(x-X_i)+o_p\Big(\frac{1}{\sqrt{n}}\Big)\Big)\,dx\,\mu(d\tau)\\
&= 2nh_x^{\frac{d_X}{2}} \int\int \kappa(x,\tau)\Big(\frac{c_n}{n}\sum_{i=1}^n K_{h_y}(F^{-1}_{Y_0|X}(\tau|x)-g_{\beta_0}(X_i)-c_0-\varepsilon_i)\\
&\quad \mathbf{K}_{h_x}(x-X_i)\Delta_n(X_i)\Big)\Big(\frac{1}{n}\sum_{i=1}^n\Big(\mathcal{K}_{h_y}(F^{-1}_{Y_0|X}(\tau|x)-g_{\beta_0}(X_i)-c_0-\varepsilon_i)\\
&\quad - \frac{p_0(F^{-1}_{Y_0|X}(\tau|x),x)}{f_X(x)}\Big)\mathbf{K}_{h_x}(x-X_i)\Big)\,dx\,\mu(d\tau)+o_p(1)\\
&= \frac{2c_nh_x^{\frac{d_X}{2}}}{n}\sum_{i=1}^n\sum_{j=1}^n\int\int \kappa(x,\tau)K_{h_y}(F^{-1}_{Y_0|X}(\tau|x)-g_{\beta_0}(X_i)-c_0-\varepsilon_i)\\
&\quad \mathbf{K}_{h_x}(x-X_i)\Delta_n(X_i)\Big(\mathcal{K}_{h_y}(F^{-1}_{Y_0|X}(\tau|x)-g_{\beta_0}(X_j)-c_0-\varepsilon_j)-\frac{p_0(F^{-1}_{Y_0|X}(\tau|x),x)}{f_X(x)}\Big)\\
&\quad \mathbf{K}_{h_x}(x-X_j)\,dx\,\mu(d\tau)+o_p(1),
\end{aligned}
$$

where the second to last equality follows similarly to the proof of Lemma 2.3.2 and the treatment of T_1. For a sufficiently large constant $C>0$ one has (see (2.44))

$$
\begin{aligned}
&c_nh_x^{\frac{d_X}{2}}E\Big[\Big|\int\int \kappa(x,\tau)K_{h_y}(F^{-1}_{Y_0|X}(\tau|x)-g_{\beta_0}(X_1)-c_0-\varepsilon_1)\mathbf{K}_{h_x}(x-X_1)^2\Delta_n(X_1)\\
&\Big(\mathcal{K}_{h_y}(F^{-1}_{Y_0|X}(\tau|x)-g_{\beta_0}(X_1)-c_0-\varepsilon_1)-\frac{p_0(F^{-1}_{Y_0|X}(\tau|x),x)}{f_X(x)}\Big)\,dx\,\mu(d\tau)\Big|\Big]\\
&\le Cc_nh_x^{\frac{d_X}{2}}E\Big[\int\int \kappa(x,\tau)|K_{h_y}(F^{-1}_{Y_0|X}(\tau|x)-g_{\beta_0}(X_1)-c_0-\varepsilon_i)|\mathbf{K}_{h_x}(x-X_1)^2\,dx\,\mu(d\tau)\Big]\\
&= Cc_nh_x^{\frac{d_X}{2}}\int\int\int\int \kappa(x,\tau)|K_{h_y}(F^{-1}_{Y_0|X}(\tau|x)-g_{\beta_0}(w)-c_0-e)|\mathbf{K}_{h_x}(x-w)^2\\
&\qquad f_X(w)f_\varepsilon(e)\,dw\,de\,dx\,\mu(d\tau)\\
&= \frac{C}{n^{\frac{1}{2}}h_x^{\frac{3d_X}{4}}}\int\int\int\int \kappa(x,\tau)|K(e)|\mathbf{K}(w)^2\\
&\qquad f_X(x-h_xw)f_\varepsilon(F^{-1}_{Y_0|X}(\tau|x)-g_{\beta_0}(w)-c_0-h_ye)\,dw\,de\,dx\,\mu(d\tau)\\
&= o(1),
\end{aligned}
$$

so that

$$T_3 = \frac{2c_n h_x^{\frac{d_X}{2}}}{n} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} \tilde{Z}_{i,j} + o_p(1)$$

with

$$\tilde{Z}_{i,j} = \int\int \kappa(x,\tau)\mathbf{K}_{h_x}(x-X_i)K_{h_y}(F^{-1}_{Y_0|X}(\tau|x) - g_{\beta_0}(X_i) - c_0 - \varepsilon_i)\Delta_n(X_i)\mathbf{K}_{h_x}(x-X_j)$$
$$\left(\mathcal{K}_{h_y}(F^{-1}_{Y_0|X}(\tau|x) - g_{\beta_0}(X_j) - c_0 - \varepsilon_j) - \frac{p_0(F^{-1}_{Y_0|X}(\tau|x),x)}{f_X(x)}\right) dx\, \mu(d\tau).$$

To prove asymptotic negligibility of T_3, it suffices to show $\frac{c_n^2 h_x^{d_X}}{n^2} E\big[\big(\sum_{i=1}^n \sum_{\substack{j=1 \\ j\neq i}}^n \tilde{Z}_{i,j}\big)^2\big] = o(1)$. This leads to the proof of

$$c_n^2 h_x^{d_X} E[\tilde{Z}_{1,2}^2] = o(1), \tag{2.52}$$
$$c_n^2 h_x^{d_X} E[\tilde{Z}_{1,2}\tilde{Z}_{2,1}] = o(1),$$
$$c_n^2 n h_x^{d_X} E[\tilde{Z}_{1,2}\tilde{Z}_{1,3}] = o(1), \tag{2.53}$$
$$c_n^2 n h_x^{d_X} E[\tilde{Z}_{1,2}\tilde{Z}_{3,1}] = o(1),$$
$$c_n^2 n h_x^{d_X} E[\tilde{Z}_{1,2}\tilde{Z}_{2,3}] = o(1),$$
$$c_n^2 n h_x^{d_X} E[\tilde{Z}_{1,2}\tilde{Z}_{3,2}] = o(1),$$
$$c_n^2 n^2 h_x^{d_X} E[\tilde{Z}_{1,2}\tilde{Z}_{3,4}] = o(1). \tag{2.54}$$

For the sake of brevity, only equations (2.52),(2.53) and (2.54) are proven. The other assertions follow similarly. Let $C > 0$ be a sufficiently large constant. Equation (2.52) results from

$$\begin{aligned}
&c_n^2 h_x^{d_X} E[\tilde{Z}_{1,2}^2] \\
&\leq C c_n^2 h_x^{d_X} E\bigg[\bigg(\int\int \kappa(x,\tau)|K_{h_y}(F^{-1}_{Y_0|X}(\tau|x) - g_{\beta_0}(X_1) - c_0 - \varepsilon_1) \\
&\qquad \mathbf{K}_{h_x}(x-X_1)\mathbf{K}_{h_x}(x-X_2)|\, dx\, \mu(d\tau)\bigg)^2\bigg] \\
&\leq \frac{Cc_n^2}{h_x^{d_X}} E\bigg[\bigg(\int\int \kappa(X_1 - h_x x,\tau)\bigg|K_{h_y}(F^{-1}_{Y_0|X}(\tau|X_1 - h_x x) - g_{\beta_0}(X_1) - c_0 - \varepsilon_1) \\
&\qquad \mathbf{K}(x)\mathbf{K}\bigg(x + \frac{X_1 - X_2}{h_x}\bigg)\bigg|\, dx\, \mu(d\tau)\bigg)^2\bigg] \\
&= \frac{Cc_n^2}{h_x^{d_X}} \int\int\int \bigg(\int\int \kappa(w_1 - h_x x,\tau)\bigg|K_{h_y}(F^{-1}_{Y_0|X}(\tau|w_1 - h_x x) - g_{\beta_0}(w_1) - c_0 - e) \\
&\qquad \mathbf{K}(x)\mathbf{K}\bigg(x + \frac{w_1 - w_2}{h_x}\bigg)\bigg|\, dx\, \mu(d\tau)\bigg)^2 f_X(w_1) f_X(w_2) f_\varepsilon(e)\, dw_1\, dw_2\, de \\
&= Cc_n^2 \int\int\int \bigg(\int\int \kappa(w_1 - h_x x,\tau)\bigg|K_{h_y}(F^{-1}_{Y_0|X}(\tau|w_1 - h_x x) - g_{\beta_0}(w_1) - c_0 - e)
\end{aligned}$$

$$\mathbf{K}(x)\mathbf{K}(x+w_2)\Big| \, dx \, \mu(d\tau)\Big)^2 f_X(w_1) f_X(w_1 - h_x w_2) f_\varepsilon(e) \, dw_1 \, dw_2 \, de$$

$$= C c_n^2 \int\int\int\int\int\int\int \kappa(w_1 - h_x x_1, \tau_1)\kappa(w_2 - h_x x_2, \tau_2)\Big|\mathbf{K}(x_1 + w_2)\mathbf{K}(x_2)\mathbf{K}(x_2 + w_2)$$

$$\mathbf{K}(x_1) K_{h_y}(F^{-1}_{Y_0|X}(\tau_1|w_1 - h_x x_1) - g_{\beta_0}(w_1) - c_0 - e)$$

$$K_{h_y}(F^{-1}_{Y_0|X}(\tau_2|w_1 - h_x x_2) - g_{\beta_0}(w_1) - c_0 - e)$$

$$\Big| f_X(w_1) f_X(w_1 - h_x w_2) f_\varepsilon(e) \, dw_1 \, dw_2 \, de \, dx_1 \, dx_2 \, \mu(d\tau_1) \, \mu(d\tau_2)$$

$$= \frac{C^2 c_n^2}{h_y} \int\int\int\int\int \kappa(w_1 - h_x x_1, \tau_1)\Big|K(e)\mathbf{K}(x_1)\mathbf{K}(x_1 + w_2)\Big|$$

$$f_X(w_1) f_X(w_1 - h_x w_2) f_\varepsilon(F^{-1}_{Y_0|X}(\tau_1|w_1 - h_x x_1) - g_{\beta_0}(w_1) - h_y e) dw_1 \, dw_2 \, de \, dx_1 \, \mu(d\tau_1)$$

$$= \mathcal{O}\Big(n^{-1} h_x^{-\frac{d_X}{2}} h_y^{-1}\Big)$$

$$= o(1).$$

In (2.48), it was shown that

$$E\Big[\mathbf{K}_{h_x}(x - X_1)\Big(\mathcal{K}_{h_y}((F_{Y_0|X})^{-1}(\tau|x) - g_{\beta_0}(X_1) - c_0 - \varepsilon_1) - \frac{p_0(F^{-1}_{Y_0|X}(\tau|x), x)}{f_X(x)}\Big)\Big] = o\Big(\frac{1}{\sqrt{n}}\Big)$$

uniformly in $x \in \operatorname{supp}(v)$ and $\tau \in \operatorname{supp}(\mu)$, so that

$$c_n^2 n h_x^{d_X} E[\tilde{Z}_{1,2}\tilde{Z}_{1,3}]$$

$$= c_n^2 n h_x^{d_X} \int\int\int\int \kappa(x_1, \tau_1)\kappa(x_2, \tau_2) E\big[\Delta_n(X_1)^2 \mathbf{K}_{h_x}(x_1 - X_1)\mathbf{K}_{h_x}(x_2 - X_1)$$

$$K_{h_y}(F^{-1}_{Y_0|X}(\tau_1|x_1) - g_{\beta_0}(X_1) - c_0 - \varepsilon_1) K_{h_y}(F^{-1}_{Y_0|X}(\tau_2|x_2) - g_{\beta_0}(X_1) - c_0 - \varepsilon_1)\big]$$

$$E\Big[\mathbf{K}_{h_x}(x_1 - X_2)\mathcal{K}_{h_y}(F^{-1}_{Y_0|X}(\tau_1|x_1) - g_{\beta_0}(X_2) - c_0 - \varepsilon_2) - \frac{p_0(F^{-1}_{Y_0|X}(\tau_1|x_1), x_1)}{f_X(x_1)}\Big]$$

$$E\Big[\mathbf{K}_{h_x}(x_2 - X_3)\mathcal{K}_{h_y}(F^{-1}_{Y_0|X}(\tau_2|x_2) - g_{\beta_0}(X_3) - c_0 - \varepsilon_3) - \frac{p_0(F^{-1}_{Y_0|X}(\tau_2|x_2), x_2)}{f_X(x_2)}\Big]$$

$$dx_1 \, dx_2 \, \mu(d\tau_1) \, \mu(d\tau_2)$$

$$= o\Bigg(\frac{h_x^{\frac{d_X}{2}}}{n}\Bigg) \int\int\int\int \kappa(x_1, \tau_1)\kappa(x_2, \tau_2) E\big[|\mathbf{K}_{h_x}(x_1 - X_1)\mathbf{K}_{h_x}(x_2 - X_1)$$

$$K_{h_y}(F^{-1}_{Y_0|X}(\tau_1|x_1) - g_{\beta_0}(X_1) - c_0 - \varepsilon_1) K_{h_y}(F^{-1}_{Y_0|X}(\tau_2|x_2) - g_{\beta_0}(X_1) - c_0 - \varepsilon_1)|\big]$$

$$dx_1 \, dx_2 \, \mu(d\tau_1) \, \mu(d\tau_2)$$

$$= o\Bigg(\frac{1}{n h_x^{\frac{d_X}{2}}}\Bigg) E\Big[\int\int\int\int \kappa(X_1 - h_x x_1, \tau_1)\kappa(x_2, \tau_2)|\mathbf{K}(x_1)$$

$$K_{h_y}(F^{-1}_{Y_0|X}(\tau_1|X_1 - h_x x_1) - g_{\beta_0}(X_1) - c_0 - \varepsilon_1)K_{h_y}(F^{-1}_{Y_0|X}(\tau_2|x_2) - g_{\beta_0}(X_1) - c_0 - \varepsilon_1)|$$

$$dx_1\, dx_2\, \mu(d\tau_1)\, \mu(d\tau_2)\Big]$$

$$= o\left(\frac{1}{nh_x^{\frac{d_X}{2}}}\right) \int\int\int\int\int\int \kappa(x_2,\tau_2)|\mathbf{K}(x_1)K_{h_y}(F^{-1}_{Y_0|X}(\tau_1|w - h_x x_1) - g_{\beta_0}(w) - c_0 - e)$$

$$K_{h_y}(F^{-1}_{Y_0|X}(\tau_2|x_2) - g_{\beta_0}(w) - c_0 - e)|f_X(w)f_\varepsilon(e)\, dw\, de\, dx_1\, dx_2\, \mu(d\tau_1)\, \mu(d\tau_2)$$

$$= o\left(\frac{1}{nh_x^{\frac{d_X}{2}} h_y}\right) \int\int\int\int\int |\mathbf{K}(x_1)K(e)|$$

$$f_X(w)f_\varepsilon(F^{-1}_{Y_0|X}(\tau_1|w - h_x x_1) - g_{\beta_0}(w) - c_0 - h_y e)\, dw\, de\, dx_1\, \mu(d\tau_1)$$

$$= o(1).$$

Moreover, equation (2.54) follows from (2.48) by

$$c_n^2 n^2 h_x^{d_X} E[\tilde{Z}_{1,2}\tilde{Z}_{3,4}]$$

$$= c_n^2 n^2 h_x^{d_X} \int\int\int\int E\big[\Delta_n(X_1)\mathbf{K}_{h_x}(x_1 - X_1)K_{h_y}(F^{-1}_{Y_0|X}(\tau_1|x_1) - g_{\beta_0}(X_1) - c_0 - \varepsilon_1)\big]$$

$$E\big[\Delta_n(X_3)\mathbf{K}_{h_x}(x_2 - X_3)K_{h_y}(F^{-1}_{Y_0|X}(\tau_2|x_2) - g_{\beta_0}(X_3) - c_0 - \varepsilon_3)\big]$$

$$E\left[\mathbf{K}_{h_x}(x_1 - X_2)\left(\mathcal{K}_{h_y}(F^{-1}_{Y_0|X}(\tau_1|x_1) - g_{\beta_0}(X_2) - c_0 - \varepsilon_2) - \frac{p_0(F^{-1}_{Y_0|X}(\tau_1|x_1), x_1)}{f_X(x_1)}\right)\right]$$

$$E\left[\mathbf{K}_{h_x}(x_2 - X_4)\left(\mathcal{K}_{h_y}(F^{-1}_{Y_0|X}(\tau_2|x_2) - g_{\beta_0}(X_4) - c_0 - \varepsilon_4) - \frac{p_0(F^{-1}_{Y_0|X}(\tau_2|x_2), x_2)}{f_X(x_2)}\right)\right]$$

$$\kappa(x_1,\tau_1)\kappa(x_2,\tau_2)dx_1\, dx_2\, \mu(d\tau_1)\, \mu(d\tau_2)$$

$$= o\big(h_x^{\frac{d_X}{2}}\big) \int\int\int\int \kappa(x_1,\tau_1)\kappa(x_2,\tau_2)\, dx_1\, dx_2\, \mu(d\tau_1)\, \mu(d\tau_2)$$

$$= o(1).$$

All in all, it was proven that $T_3 = o_p(1)$ and thus

$$\tilde{T}_n - b = nh_x^{\frac{d_X}{2}} \int\int v(x)\big(\hat{F}^{-1}_{Y|X}(\tau|x) - F^{-1}_{Y_0|X}(\tau|x)\big)^2\, dx\, \mu(d\tau) - b$$

$$\xrightarrow{\mathcal{D}} Z + \delta_{1,n}$$

with $Z \sim \mathcal{N}(0,V)$ and $\delta_{1,n} = \mu([0,1])\int v(x)\Delta_n(x)^2\, dx$.

Asymptotic Equivalence of T_n and $\tilde{T}_n + \delta_{2,n} + \delta_{3,n}$

Recall Remark 2.3.1 and the definition of $\hat{c}_{\beta,\tau}$ in (2.14). Due to

$$\delta_n = \delta_{1,n} + \delta_{2,n} + \delta_{3,n}$$

(see Remark 2.3.5) it remains to show asymptotic equivalence of $\tilde{T}_n + \delta_{2,n} + \delta_{3,n}$ and

$$T_n = nh_x^{\frac{d_X}{2}} \min_{\beta \in B} \int \int v(x)\big(\hat{F}^{-1}_{Y|X}(\tau|x) - g_\beta(x) - \hat{c}_{\beta,\tau}\big)^2 \, dx \, \mu(d\tau).$$

For that purpose define

$$\begin{aligned}
&G(\beta)\\
&= -2nh_x^{\frac{d_X}{2}} \int \int v(x)\big(\hat{F}^{-1}_{Y|X}(\tau|x) - g_{\beta_0}(x) - c_{\beta_0,\tau}\big) \, dx (\hat{c}_{\beta_0,\tau} - c_{\beta_0,\tau}) \, \mu(d\tau)\\
&\quad - 2nh_x^{\frac{d_X}{2}} \int \int v(x)\big(\hat{F}^{-1}_{Y|X}(\tau|x) - g_{\beta_0}(x) - c_{\beta_0,\tau}\big) D_\beta(g_{\beta_0}(x) + c_{\beta_0,\tau}) \, dx \, \mu(d\tau)(\beta - \beta_0)\\
&\quad + nh_x^{\frac{d_X}{2}} \int v(x) \, dx \int (\hat{c}_{\beta_0,\tau} - c_{\beta_0,\tau})^2 \, \mu(d\tau) + nh_x^{\frac{d_X}{2}} \mu([0,1])(\beta - \beta_0)^t \Omega (\beta - \beta_0)
\end{aligned}$$

as well as $\bar{\beta} = \arg\min_{\beta \in B} G(\beta)$ and

$$\hat{\beta} = nh_x^{\frac{d_X}{2}} \arg\min_{\beta \in B} \int \int v(x)\big(\hat{F}^{-1}_{Y|X}(\tau|x) - g_\beta(x) - \hat{c}_{\beta,\tau}\big)^2 \, dx \, \mu(d\tau).$$

First, it will be shown that

$$||\hat{\beta} - \beta_0|| = \mathcal{O}_p\Big(n^{-\frac{1}{2}} h_x^{-\frac{d_X}{4}}\Big), \tag{2.55}$$

$$||\bar{\beta} - \beta_0|| = \mathcal{O}_p\Big(n^{-\frac{1}{2}} h_x^{-\frac{d_X}{4}}\Big). \tag{2.56}$$

Due to

$$\hat{F}^{-1}_{Y|X}(\tau|x) - g_{\beta_0}(x) - c_{\beta_0,\tau} = \hat{F}^{-1}_{Y|X}(\tau|x) - F^{-1}_{Y_0|X}(\tau|x) = o_p(1)$$

uniformly in $x \in \operatorname{supp}(v), \tau \in \mu$, assumption (**A7**) implies $\hat{\beta} - \beta_0 = o_p(1)$ and $\bar{\beta} - \beta_0 = o_p(1)$. Further, for all sequences $\beta_n \in B, n \in \mathbb{N}$ with $||\beta_n - \beta_0|| \to 0$, Lemma 2.8.1, assumption (**A7**) and equation (2.14) yield

$$\begin{aligned}
&\hat{c}_{\beta_n,\tau} - c_{\beta_0,\tau}\\
&= \frac{\int v(x)\big(\hat{F}^{-1}_{Y|X}(\tau|x) - F^{-1}_{Y_0|X}(\tau|x) + g_{\beta_0}(x) - g_{\beta_n}(x)\big) \, dx}{\int v(x) \, dx}\\
&= \frac{\int v(x)\big(\hat{F}^{-1}_{Y|X}(\tau|x) - \hat{F}^{-1}_{Y_0|X}(\tau|x) + \hat{F}^{-1}_{Y_0|X}(\tau|x) - F^{-1}_{Y_0|X}(\tau|x)\big) \, dx}{\int v(x) \, dx} + \mathcal{O}_p(||\beta_n - \beta_0||)\\
&= \frac{c_n}{n \int v(x) \, dx} \sum_{i=1}^n \int \frac{v(x)}{f_{Y_0|X}(F^{-1}_{Y_0|X}(\tau|x)|x) f_X(x)} K_{h_y}(F^{-1}_{Y_0|X}(\tau|x) - g_{\beta_0}(X_i) - c_0 - \varepsilon_i)\\
&\qquad \mathbf{K}_{h_x}(x - X_i) \Delta_n(X_i) \, dx + \mathcal{O}_p\bigg(\frac{1}{\sqrt{n}} + ||\beta_n - \beta_0||\bigg)\\
&= \frac{c_n}{\int v(x) \, dx} E\bigg[\int \frac{v(x)}{f_{Y_0|X}(F^{-1}_{Y_0|X}(\tau|x)|x) f_X(x)} K_{h_y}(F^{-1}_{Y_0|X}(\tau|x) - g_{\beta_0}(X_i) - c_0 - \varepsilon_i)
\end{aligned}$$

$$\mathbf{K}_{h_x}(x - X_i)\Delta_n(X_i)\,dx\Big] + o_p(c_n) + \mathcal{O}_p\Big(\frac{1}{\sqrt{n}} + ||\beta_n - \beta_0||\Big)$$

$$= c_n \frac{\int v(x)\Delta_n(x)\,dx}{\int v(x)\,dx} + o_p(c_n) + \mathcal{O}_p\Big(\frac{1}{\sqrt{n}} + ||\beta_n - \beta_0||\Big) \tag{2.57}$$

uniformly in $\tau \in \mathrm{supp}(\mu)$, where the second to last equation can be shown analogously to the reasoning in the proof of Lemma 4.2.12 later. Moreover, note that

$$D_\beta \hat{c}_{\beta,\tau} = -\frac{\int v(x) D_\beta g_\beta(x)\,dx}{\int v(x)\,dx} = D_\beta c_{\beta,\tau} \tag{2.58}$$

and

$$\int\int v(x)(c_{\beta_0,\tau} - \hat{c}_{\beta_0,\tau}) D_\beta(g_{\beta_0}(x) + c_{\beta_0,\tau})\,dx\,\mu(d\tau)$$

$$\overset{(2.57)}{=} c_n \frac{\int v(x)\Delta_n(x)\,dx}{\int v(x)\,dx} \int\int v(x) D_\beta(g_{\beta_0}(x) + c_{\beta_0,\tau})\,dx\,\mu(d\tau) + o_p(c_n)$$

$$= c_n \frac{\int v(x)\Delta_n(x)\,dx}{\int v(x)\,dx} \int\int v(x)\Big(D_\beta g_{\beta_0}(x) - \frac{\int v(w) D_\beta g_{\beta_0}(w)\,dw}{\int v(w)\,dw}\Big)\,dx\,\mu(d\tau) + o_p(c_n)$$

$$= o_p(c_n).$$

Therefore, a Taylor expansion of $\beta \mapsto \big(\hat{F}^{-1}_{Y|X}(\tau|x) - g_\beta(x) - c_{\beta,\tau}\big)^2$ and the binomial formula yield for some β^* between $\hat{\beta}$ and β_0

$$0 \le nh_x^{\frac{d_X}{2}} \int\int v(x)\big(\hat{F}^{-1}_{Y|X}(\tau|x) - g_{\beta_0}(x) - c_{\beta_0,\tau}\big)^2\,dx\,\mu(d\tau) - T_n$$

$$= nh_x^{\frac{d_X}{2}} \int\int v(x)\big(\hat{F}^{-1}_{Y|X}(\tau|x) - g_{\beta_0}(x) - c_{\beta_0,\tau}\big)^2\,dx\,\mu(d\tau)$$

$$- nh_x^{\frac{d_X}{2}} \int\int v(x)\big(\hat{F}^{-1}_{Y|X}(\tau|x) - g_{\beta_0}(x) - \hat{c}_{\beta_0,\tau}\big)^2\,dx\,\mu(d\tau)$$

$$+ 2nh_x^{\frac{d_X}{2}} \int\int v(x)\big(\hat{F}^{-1}_{Y|X}(\tau|x) - g_{\beta_0}(x) - \hat{c}_{\beta_0,\tau}\big) D_\beta(g_{\beta_0}(x) + \hat{c}_{\beta_0,\tau})\,dx\,\mu(d\tau)(\hat{\beta} - \beta_0)$$

$$+ nh_x^{\frac{d_X}{2}} (\hat{\beta} - \beta_0)^t \int\int \Big(v(x)\big(\hat{F}^{-1}_{Y|X}(\tau|x) - g_{\beta^*}(x) - \hat{c}_{\beta^*,\tau}\big)\,\mathrm{Hess}(g_{\beta^*}(x) + \hat{c}_{\beta^*,\tau})$$

$$- \big(D_\beta(g_{\beta^*}(x) + \hat{c}_{\beta^*,\tau})\big)^t D_\beta(g_{\beta^*}(x) + \hat{c}_{\beta^*,\tau})\Big)\,dx\,\mu(d\tau)(\hat{\beta} - \beta_0)$$

$$= 2nh_x^{\frac{d_X}{2}} \int\int v(x)\big(\hat{F}^{-1}_{Y|X}(\tau|x) - g_{\beta_0}(x) - c_{\beta_0,\tau}\big)\,dx(\hat{c}_{\beta_0,\tau} - c_{\beta_0,\tau})\,\mu(d\tau)$$

$$- nh_x^{\frac{d_X}{2}} \int v(x)\,dx \int (\hat{c}_{\beta_0,\tau} - c_{\beta_0,\tau})^2\,\mu(d\tau)$$

$$+ 2nh_x^{\frac{d_X}{2}} \int\int v(x)\big(\hat{F}^{-1}_{Y|X}(\tau|x) - g_{\beta_0}(x) - \hat{c}_{\beta_0,\tau}\big) D_\beta(g_{\beta_0}(x) + c_{\beta_0,\tau})\,dx\,\mu(d\tau)(\hat{\beta} - \beta_0)$$

$$- nh_x^{\frac{d_X}{2}} \mu([0,1])(\hat{\beta} - \beta_0)^t \Omega(\hat{\beta} - \beta_0) + o_p\Big(nh_x^{\frac{d_X}{2}} ||\hat{\beta} - \beta_0||^2\Big) \tag{2.59}$$

$$= -G(\hat{\beta}) + o_p\Big(\sqrt{n} h_x^{\frac{d_X}{4}} ||\hat{\beta} - \beta_0||\Big) + o_p\Big(nh_x^{\frac{d_X}{2}} ||\hat{\beta} - \beta_0||^2\Big).$$

Here, it was used that
$$\hat{F}^{-1}_{Y|X}(\tau|x) - g_{\beta^*}(x) - \hat{c}_{\beta^*,\tau} = o_p(1)$$
uniformly in $x \in \operatorname{supp}(v), \tau \in \operatorname{supp}(\mu)$ and
$$\int\int v(x)\big(D_\beta(g_{\beta^*}(x) + \hat{c}_{\beta^*,\tau})\big)^t D_\beta(g_{\beta^*}(x) + \hat{c}_{\beta^*,\tau})\,dx\,\mu(d\tau) = \Omega + o_p(1)$$
componentwise due to $\beta^* - \beta_0 = o_p(1)$.
Later, it will be shown that (again componentwise)
$$\begin{aligned}
&\sqrt{n}h_x^{\frac{d_X}{4}}\int\left|\int\big(\hat{F}^{-1}_{Y|X}(\tau|x) - g_{\beta_0}(x) - c_{\beta_0,\tau}\big)v(x)D_\beta(g_{\beta_0}(x) + c_{\beta_0,\tau})\,dx\right|\mu(d\tau)\\
&\leq \sqrt{n}h_x^{\frac{d_X}{4}}\int\left|\int\big(\hat{F}^{-1}_{Y|X}(\tau|x) - \hat{F}^{-1}_{Y_0|X}(\tau|x)\big)v(x)D_\beta(g_{\beta_0}(x) + c_{\beta_0,\tau})\,dx\right|\mu(d\tau)\\
&\quad + \sqrt{n}h_x^{\frac{d_X}{4}}\int\left|\int\big(\hat{F}^{-1}_{Y_0|X}(\tau|x) - F^{-1}_{Y_0|X}(\tau|x)\big)v(x)D_\beta(g_{\beta_0}(x) + c_{\beta_0,\tau})\,dx\right|\mu(d\tau)\\
&= \mathcal{O}_p(1) \qquad (2.60)
\end{aligned}$$
as well as
$$\sqrt{n}h_x^{\frac{d_X}{4}}\int\left|\int\big(\hat{F}^{-1}_{Y|X}(\tau|x) - g_{\beta_0}(x) - c_{\beta_0,\tau}\big)v(x)\,dx\right|\mu(d\tau) = \mathcal{O}_p(1).$$
Since (2.57) implies
$$\hat{c}_{\beta_0,\tau} - c_{\beta_0,\tau} = \mathcal{O}_p\Big(c_n + \frac{1}{\sqrt{n}}\Big),$$
equation (2.59) then leads to
$$\begin{aligned}
0 &\leq -nh_x^{\frac{d_X}{2}}\mu([0,1])(\hat{\beta} - \beta_0)^t\Omega(\hat{\beta} - \beta_0)\\
&\quad + \mathcal{O}_p\Big(\sqrt{n}h_x^{\frac{d_X}{4}}||\hat{\beta} - \beta_0||\Big) + o_p\Big(nh_x^{\frac{d_X}{2}}||\hat{\beta} - \beta_0||^2\Big) + \mathcal{O}_p(1),
\end{aligned}$$
that is $\hat{\beta} - \beta_0 = \mathcal{O}_p\big(n^{-\frac{1}{2}}h_x^{-\frac{d_X}{4}}\big)$.
To prove the equations from above, define
$$\tilde{\kappa}(x,\tau) = \frac{v(x)D_\beta(g_{\beta_0}(x) + c_{\beta_0,\tau})}{f_{Y_0|X}(F^{-1}_{Y_0|X}(\tau|x)|x)f_X(x)} \qquad (2.61)$$
and write with Lemma 2.8.1
$$\begin{aligned}
&\sqrt{n}h_x^{\frac{d_X}{4}}\int\left|\int\big(\hat{F}^{-1}_{Y|X}(\tau|x) - \hat{F}^{-1}_{Y_0|X}(\tau|x)\big)v(x)D_\beta(g_{\beta_0}(x) + c_{\beta_0,\tau})\,dx\right|\mu(d\tau)\\
&= \frac{1}{n}\sum_{i=1}^n\int\left|\int\tilde{\kappa}(x,\tau)K_{h_y}(F^{-1}_{Y_0|X}(\tau|x) - g_{\beta_0}(X_i) - c_0 - \varepsilon_i)\mathbf{K}_{h_x}(x - X_i)\Delta_n(X_i)\,dx\right|\mu(d\tau)\\
&\quad + o_p(1)\\
&= E\Big[\int\left|\int\tilde{\kappa}(x,\tau)K_{h_y}(F^{-1}_{Y_0|X}(\tau|x) - g_{\beta_0}(X_1) - c_0 - \varepsilon_1)\mathbf{K}_{h_x}(x - X_1)\Delta_n(X_1)\,dx\right|\mu(d\tau)\Big]
\end{aligned}$$

$$
\begin{aligned}
&+ o_p(1)\\
&= \mathcal{O}_p(1)
\end{aligned}
$$

as well as (with Lemma 2.8.1)

$$
\begin{aligned}
&\sqrt{n} h_x^{\frac{d_X}{4}} \int \left| \int \left(\hat{F}^{-1}_{Y_0|X}(\tau|x) - F^{-1}_{Y_0|X}(\tau|x)\right) v(x) D_\beta (g_{\beta_0}(x) + c_{\beta_0,\tau})\, dx \right| \mu(d\tau)\\
&= \int \left| \frac{h_x^{\frac{d_X}{4}}}{\sqrt{n}} \sum_{i=1}^n \int \tilde{\kappa}(x,\tau) \Bigg(\mathcal{K}_{h_y}(F^{-1}_{Y_0|X}(\tau|x) - g_{\beta_0}(X_i) - c_0 - \varepsilon_i) - \frac{p_0(F^{-1}_{Y_0|X}(\tau|x), x)}{f_X(x)} \Bigg) \right.\\
&\qquad \left. \mathbf{K}_{h_x}(x - X_i)\, dx \right| \mu(d\tau) + o_p(1)\\
&= \int \left| \frac{h_x^{\frac{d_X}{4}}}{\sqrt{n}} \sum_{i=1}^n \int \tilde{\kappa}(X_i - h_x x, \tau) \Bigg(\mathcal{K}_{h_y}(F^{-1}_{Y_0|X}(\tau|X_i - h_x x) - g_{\beta_0}(X_i) - c_0 - \varepsilon_i) \right.\\
&\qquad \left. - \frac{p_0(F^{-1}_{Y_0|X}(\tau|X_i - h_x x), X_i - h_x x)}{f_X(X_i - h_x x)} \Bigg) \mathbf{K}(x)\, dx \right| \mu(d\tau) + o_p(1).
\end{aligned}
$$

Let $C > 0$ be a sufficiently large constant. Then, for each of the components $\tilde{\kappa}_k, k = 1, ..., d_B$, one has

$$
\begin{aligned}
&E\Bigg[\Bigg(\int \left| \frac{h_x^{\frac{d_X}{4}}}{\sqrt{n}} \sum_{i=1}^n \int \tilde{\kappa}_k(X_i - h_x x, \tau) \Bigg(\mathcal{K}_{h_y}(F^{-1}_{Y_0|X}(\tau|X_i - h_x x) - g_{\beta_0}(X_i) - c_0 - \varepsilon_i) \right.\\
&\left. - \frac{p_0(F^{-1}_{Y_0|X}(\tau|X_i - h_x x), X_i - h_x x)}{f_X(X_i - h_x x)} \Bigg) \mathbf{K}(x)\, dx \right| \mu(d\tau) \Bigg)^2 \Bigg]\\
&\leq \mu([0,1]) E\Bigg[\int \Bigg(\frac{h_x^{\frac{d_X}{4}}}{\sqrt{n}} \sum_{i=1}^n \int \tilde{\kappa}_k(X_i - h_x x, \tau)\\
&\quad \Bigg(\mathcal{K}_{h_y}(F^{-1}_{Y_0|X}(\tau|X_i - h_x x) - g_{\beta_0}(X_i) - c_0 - \varepsilon_i) - \frac{p_0(F^{-1}_{Y_0|X}(\tau|X_i - h_x x), X_i - h_x x)}{f_X(X_i - h_x x)} \Bigg)\\
&\quad \mathbf{K}(x)\, dx \Bigg)^2 \mu(d\tau) \Bigg]\\
&\leq \mu([0,1]) h_X^{\frac{d_X}{2}} E\Bigg[\int \Bigg(\int \tilde{\kappa}_k(X_1 - h_x x, \tau) \Bigg(\mathcal{K}_{h_y}(F^{-1}_{Y_0|X}(\tau|X_1 - h_x x) - g_{\beta_0}(X_1) - c_0 - \varepsilon_1)\\
&\quad - \frac{p_0(F^{-1}_{Y_0|X}(\tau|X_1 - h_x x), X_1 - h_x x)}{f_X(X_1 - h_x x)} \Bigg) \mathbf{K}(x)\, dx \Bigg)^2 \mu(d\tau) \Bigg]\\
&\quad + \mu([0,1]) n h_X^{\frac{d_X}{2}} \int \int E\Bigg[\tilde{\kappa}_k(X_1 - h_x x, \tau) \Bigg(\mathcal{K}_{h_y}(F^{-1}_{Y_0|X}(\tau|X_1 - h_x x) - g_{\beta_0}(X_1) - c_0 - \varepsilon_1)\\
&\quad - \frac{p_0(F^{-1}_{Y_0|X}(\tau|X_1 - h_x x), X_1 - h_x x)}{f_X(X_1 - h_x x)} \Bigg) \Bigg]^2 \mathbf{K}(x)\, dx\, \mu(d\tau)
\end{aligned}
$$

$$\leq Ch_X^{\frac{d_X}{2}}\Big(\int |\mathbf{K}(x)|\,dx\Big)^2 + o\Big(h_X^{\frac{d_X}{2}}\Big)$$

$$= o(1),$$

where the last inequality can be shown similarly to (2.48), so that

$$\sqrt{n}h_x^{\frac{d_X}{4}}\int\Big|\int\big(\hat{F}^{-1}_{Y_0|X}(\tau|x)-F^{-1}_{Y_0|X}(\tau|x)\big)v(x)D_\beta(g_{\beta_0}(x)+c_{\beta_0,\tau})\,dx\Big|\,\mu(d\tau)=o_p(1) \quad (2.62)$$

and

$$\sqrt{n}h_x^{\frac{d_X}{4}}\int\Big|\int\big(\hat{F}^{-1}_{Y|X}(\tau|x)-g_{\beta_0}(x)-c_{\beta_0,\tau}\big)v(x)D_\beta(g_{\beta_0}(x)+c_{\beta_0,\tau})\,dx\Big|\,\mu(d\tau)=\mathcal{O}_p(1).$$

Completely analogously, it can be shown that

$$\sqrt{n}h_x^{\frac{d_X}{4}}\int\Big|\int\big(\hat{F}^{-1}_{Y|X}(\tau|x)-g_{\beta_0}(x)-c_{\beta_0,\tau}\big)v(x)\,dx\Big|\,\mu(d\tau)=\mathcal{O}_p(1).$$

Therefore, it holds that $\hat{\beta}-\beta_0=\mathcal{O}_p\big(n^{-\frac{1}{2}}h_x^{-\frac{d_X}{4}}\big)$. Especially, (2.59) implies

$$T_n = nh_x^{\frac{d_X}{2}}\int\int v(x)\big(\hat{F}^{-1}_{Y|X}(\tau|x)-g_{\beta_0}(x)-c_{\beta_0,\tau}\big)^2\,dx\,\mu(d\tau)+\mathcal{O}_p(1).$$

$\bar{\beta}$ is defined as the due to (**A8**) unique minimizer of G. Hence,

$$\begin{aligned}
0 &= D_\beta G(\bar{\beta})\\
&= -2nh_x^{\frac{d_X}{2}}\int\int v(x)\big(\hat{F}^{-1}_{Y|X}(\tau|x)-g_{\beta_0}(x)-c_{\beta_0,\tau}\big)D_\beta(g_{\beta_0}(x)+c_{\beta_0,\tau})\,dx\,\mu(d\tau)\\
&\quad + 2nh_x^{\frac{d_X}{2}}\mu([0,1])(\beta-\beta_0)^t\Omega,
\end{aligned}$$

that is, (2.60) leads to

$$\begin{aligned}
\bar{\beta} &= \beta_0+\frac{1}{\mu([0,1])}\Omega^{-1}\int\int v(x)\big(\hat{F}^{-1}_{Y|X}(\tau|x)-g_{\beta_0}(x)-c_{\beta_0,\tau}\big)\\
&\qquad \big(D_\beta(g_{\beta_0}(x)+c_{\beta_0,\tau})\big)^t\,dx\,\mu(d\tau)\\
&= \beta_0+\mathcal{O}_p\Big(n^{-\frac{1}{2}}h_x^{-\frac{d_X}{4}}\Big).
\end{aligned} \quad (2.63)$$

Note that for all $\beta\in B$ with $||\beta-\beta_0||=\mathcal{O}_p\Big(n^{-\frac{1}{2}}h_x^{-\frac{d_X}{4}}\Big)$, one has (see (2.59))

$$\begin{aligned}
&\int\int v(x)\big(\hat{F}^{-1}_{Y|X}(\tau|x)-g_\beta(x)-\hat{c}_{\beta,\tau}\big)^2\,dx\,\mu(d\tau)\\
&=\int\int v(x)\big(\hat{F}^{-1}_{Y|X}(\tau|x)-g_{\beta_0}(x)-c_{\beta_0,\tau}\big)^2\,dx\,\mu(d\tau)+G(\beta)+o_p(1),
\end{aligned}$$

so that

$$T_n = nh_x^{\frac{d_X}{2}}\int\int v(x)\big(\hat{F}^{-1}_{Y|X}(\tau|x)-g_{\hat{\beta}}(x)-\hat{c}_{\hat{\beta},\tau}\big)^2\,dx\,\mu(d\tau)$$

$$
\begin{aligned}
&= nh_x^{\frac{d_X}{2}} \int\int v(x)\big(\hat{F}_{Y|X}^{-1}(\tau|x) - g_{\beta_0}(x) - c_{\beta_0,\tau}\big)^2\, dx\, \mu(d\tau) + G(\hat{\beta}) + o_p(1)\\
&\geq nh_x^{\frac{d_X}{2}} \int\int v(x)\big(\hat{F}_{Y|X}^{-1}(\tau|x) - g_{\beta_0}(x) - c_{\beta_0,\tau}\big)^2\, dx\, \mu(d\tau) + G(\bar{\beta}) + o_p(1)\\
&= nh_x^{\frac{d_X}{2}} \int\int v(x)\big(\hat{F}_{Y|X}^{-1}(\tau|x) - g_{\bar{\beta}}(x) - \hat{c}_{\bar{\beta},\tau}\big)^2\, dx\, \mu(d\tau) + o_p(1)\\
&\geq T_n + o_p(1).
\end{aligned}
$$

Consequently, to obtain the asymptotic distribution of T_n it suffices to calculate that of

$$
nh_x^{\frac{d_X}{2}} \int\int v(x)\big(\hat{F}_{Y|X}^{-1}(\tau|x) - g_{\beta_0}(x) - c_{\beta_0,\tau}\big)^2\, dx\, \mu(d\tau) + G(\bar{\beta}) = \tilde{T}_n + G(\bar{\beta}).
$$

Inserting $\bar{\beta}$ from (2.63) into $G(\bar{\beta})$ yields

$$
\begin{aligned}
&\tilde{T}_n + G(\bar{\beta})\\
&= \tilde{T}_n - 2nh_x^{\frac{d_X}{2}} \int\int v(x)\big(\hat{F}_{Y|X}^{-1}(\tau|x) - g_{\beta_0}(x) - c_{\beta_0,\tau}\big)\, dx (\hat{c}_{\beta_0,\tau} - c_{\beta_0,\tau})\, \mu(d\tau)\\
&\quad + nh_x^{\frac{d_X}{2}} \int v(x)\, dx \int (\hat{c}_{\beta_0,\tau} - c_{\beta_0,\tau})^2 \mu(d\tau)\\
&\quad - \frac{nh_x^{\frac{d_X}{2}}}{\mu([0,1])}\Big(\int\int v(x)\big(\hat{F}_{Y|X}^{-1}(\tau|x) - g_{\beta_0}(x) - c_{\beta_0,\tau}\big) D_\beta(g_{\beta_0}(x) + c_{\beta_0,\tau})\, dx\, \mu(d\tau)\Big)\Omega^{-1}\\
&\quad \Big(\int\int v(x)\big(\hat{F}_{Y|X}^{-1}(\tau|x) - g_{\beta_0}(x) - c_{\beta_0,\tau}\big) D_\beta(g_{\beta_0}(x) + c_{\beta_0,\tau})\, dx\, \mu(d\tau)\Big)^t + o_p(1).
\end{aligned}
$$

Since $\hat{c}_{\beta_0,\tau}$ was defined as the minimizer of $c \mapsto \int v(x)\big(\hat{F}_{Y|X}^{-1}(\tau|x) - g_{\beta_0}(x) - c\big)^2\, dx$, it holds that

$$
\int v(x)\big(\hat{F}_{Y|X}^{-1}(\tau|x) - g_{\beta_0}(x) - \hat{c}_{\beta_0,\tau}\big)\, dx = 0
$$

and thus

$$
\begin{aligned}
&\int v(x)\big(\hat{F}_{Y|X}^{-1}(\tau|x) - g_{\beta_0}(x) - c_{\beta_0,\tau}\big)\, dx\\
&= \int v(x)\big(\hat{F}_{Y|X}^{-1}(\tau|x) - g_{\beta_0}(x) - \hat{c}_{\beta_0,\tau} + \hat{c}_{\beta_0,\tau} - c_{\beta_0,\tau}\big)\, dx\\
&= \int v(x)(\hat{c}_{\beta_0,\tau} - c_{\beta_0,\tau})\, dx
\end{aligned}
$$

for all $\tau \in \operatorname{supp}(\mu)$. Together with $F_{Y_0|X}^{-1}(\tau|x) = g_{\beta_0}(x) + c_{\beta_0,\tau}$, this results in

$$
\begin{aligned}
&\tilde{T}_n + G(\bar{\beta})\\
&= \tilde{T}_n - \delta_{1,n} + \mu([0,1]) \int v(x)\Delta_n(x)^2\, dx - nh_x^{\frac{d_X}{2}} \int v(x)\, dx \int (\hat{c}_{\beta_0,\tau} - c_{\beta_0,\tau})^2 \mu(d\tau)\\
&\quad - \frac{nh_x^{\frac{d_X}{2}}}{\mu([0,1])}\Big(\int\int v(x)\big(\hat{F}_{Y|X}^{-1}(\tau|x) - F_{Y_0|X}^{-1}(\tau|x)\big) D_\beta(g_{\beta_0}(x) + c_{\beta_0,\tau})\, dx\, \mu(d\tau)\Big)\Omega^{-1}
\end{aligned}
$$

$$\Big(\int\int v(x)\big(\hat{F}^{-1}_{Y|X}(\tau|x) - F^{-1}_{Y_0|X}(\tau|x)\big)D_\beta(g_{\beta_0}(x) + c_{\beta_0,\tau})\,dx\,\mu(d\tau)\Big)^t + o_p(1)$$

$$\overset{(2.62)}{=} \tilde{T}_n - \delta_{1,n} + \mu([0,1])\int v(x)\Delta_n(x)^2\,dx - nh_x^{\frac{d_X}{2}}\int v(x)\,dx\int(\hat{c}_{\beta_0,\tau} - c_{\beta_0,\tau})^2\mu(d\tau)$$

$$- \frac{nh_x^{\frac{d_X}{2}}}{\mu([0,1])}\Big(\int\int v(x)\big(\hat{F}^{-1}_{Y|X}(\tau|x) - \hat{F}^{-1}_{Y_0|X}(\tau|x)\big)D_\beta(g_{\beta_0}(x) + c_{\beta_0,\tau})\,dx\,\mu(d\tau)\Big)\Omega^{-1}$$

$$\Big(\int\int v(x)\big(\hat{F}^{-1}_{Y|X}(\tau|x) - \hat{F}^{-1}_{Y_0|X}(\tau|x)\big)D_\beta(g_{\beta_0}(x) + c_{\beta_0,\tau})\,dx\,\mu(d\tau)\Big)^t + o_p(1)$$

$$= \tilde{T}_n - \delta_{1,n} + \mu([0,1])\int v(x)\Delta_n(x)^2\,dx - \mu([0,1])\frac{\big(\int v(x)\Delta_n(x)\,dx\big)^2}{\int v(x)\,dx}$$

$$- \frac{1}{\mu([0,1])}\Big(\frac{1}{n}\sum_{i=1}^n\int\int\frac{v(x)}{f_{Y_0|X}(F^{-1}_{Y_0|X}(\tau|x)|x)f_X(x)}D_\beta(g_{\beta_0}(x) + c_{\beta_0,\tau})$$

$$K_{h_y}(F^{-1}_{Y_0|X}(\tau|x) - g_{\beta_0}(X_i) - c_0 - \varepsilon_i)\mathbf{K}_{h_x}(x - X_i)\Delta_n(X_i)\,dx\,\mu(d\tau)\Big)$$

$$\Omega^{-1}\Big(\frac{1}{n}\sum_{i=1}^n\int\int\frac{v(x)}{f_{Y_0|X}(F^{-1}_{Y_0|X}(\tau|x)|x)f_X(x)}D_\beta(g_{\beta_0}(x) + c_{\beta_0,\tau})$$

$$K_{h_y}(F^{-1}_{Y_0|X}(\tau|x) - g_{\beta_0}(X_i) - c_0 - \varepsilon_i)\mathbf{K}_{h_x}(x - X_i)\Delta_n(X_i)\,dx\,\mu(d\tau)\Big)^t + o_p(1),$$

where (2.46) and (2.57) were applied to obtain the last equation. Let $\tilde{\kappa}$ be as in (2.61). Then, one has

$$\frac{1}{n}\sum_{i=1}^n\int\int\frac{v(x)}{f_{Y_0|X}(F^{-1}_{Y_0|X}(\tau|x)|x)f_X(x)}D_\beta(g_{\beta_0}(x) + c_{\beta_0,\tau})$$

$$K_{h_y}(F^{-1}_{Y_0|X}(\tau|x) - g_{\beta_0}(X_i) - c_0 - \varepsilon_i)\mathbf{K}_{h_x}(x - X_i)\Delta_n(X_i)\,dx\,\mu(d\tau)$$

$$= \frac{1}{n}\sum_{i=1}^n\int\int\tilde{\kappa}(x,\tau)K_{h_y}(F^{-1}_{Y_0|X}(\tau|x) - g_{\beta_0}(X_i) - c_0 - \varepsilon_i)\mathbf{K}_{h_x}(x - X_i)\Delta_n(X_i)\,dx\,\mu(d\tau)$$

$$= \int\int\tilde{\kappa}(x,\tau)E\big[K_{h_y}(F^{-1}_{Y_0|X}(\tau|x) - g_{\beta_0}(X_1) - c_0 - \varepsilon_1)\mathbf{K}_{h_x}(x - X_1)\Delta_n(X_1)\big]\,dx\,\mu(d\tau)$$

$$+ o_p(1)$$

$$= \int\int\int\int\tilde{\kappa}(x,\tau)K_{h_y}(F^{-1}_{Y_0|X}(\tau|x) - z)\mathbf{K}_{h_x}(x - w)\Delta_n(w)f_{Y_0,X}(z,w)\,dz\,dw\,dx\,\mu(d\tau)$$

$$+ o_p(1)$$

$$= \int\int\int\int\tilde{\kappa}(x,\tau)K(z)\mathbf{K}(w)\Delta_n(x - h_xw)$$

$$f_{Y_0,X}(F^{-1}_{Y_0|X}(\tau|x - h_xw) - h_yz, x - h_xw)\,dz\,dw\,dx\,\mu(d\tau) + o_p(1)$$

$$= \int\int\tilde{\kappa}(x,\tau)\Delta_n(x)f_{Y_0,X}(F^{-1}_{Y_0|X}(\tau|x), x)\,dx\,\mu(d\tau) + o_p(1)$$

$$= \int \int v(w) D_\beta (g_{\beta_0}(w) + c_{\beta_0,\tau}) \Delta_n(w)\, dw\, \mu(d\tau) + o_p(1),$$

so that with (2.58)

$$\begin{aligned}
&\tilde{T}_n + G(\bar{\beta}) \\
&= \tilde{T}_n - \delta_{1,n} + \mu([0,1]) \int v(x)\Delta_n(x)^2\, dx - \mu([0,1]) \frac{\left(\int v(x)\Delta_n(x)\, dx\right)^2}{\int v(x)\, dx} \\
&\quad - \frac{1}{\mu([0,1])} \left(\int \int v(x) D_\beta (g_{\beta_0}(x) + c_{\beta_0,\tau}) \Delta_n(x)\, dx\, \mu(d\tau) \right) \\
&\quad \Omega^{-1} \left(\int \int v(x) D_\beta (g_{\beta_0}(x) + c_{\beta_0,\tau}) \Delta_n(x)\, dx\, \mu(d\tau) \right)^t + o_p(1) \\
&= \tilde{T}_n - \delta_{1,n} + \mu([0,1]) \int v(x)\Delta_n(x)^2\, dx - \mu([0,1]) \frac{\left(\int v(x)\Delta_n(x)\, dx\right)^2}{\int v(x)\, dx} \\
&\quad - \mu([0,1]) \left(\int v(x)\Delta_n(x) \left(D_\beta g_{\beta_0}(x) - \frac{\int D_\beta g_{\beta_0}(w)\, dw}{\int v(w)\, dw} \right) dx \right) \\
&\quad \Omega^{-1} \left(\int v(x)\Delta_n(x) \left(D_\beta g_{\beta_0}(x) - \frac{\int D_\beta g_{\beta_0}(w)\, dw}{\int v(w)\, dw} \right) dx \right)^t + o_p(1) \\
&= \tilde{T}_n - \delta_{1,n} + \mu([0,1]) \int v(x) \Bigg(\Delta_n(x) - \frac{\int v(w_1)\Delta_n(w_1)\, dw_1}{\int v(w_2)\, dw_2} \\
&\quad - \left(D_\beta g_{\beta_0}(x) - \frac{\int D_\beta g_{\beta_0}(w_3)\, dw_3}{\int v(w_4)\, dw_4} \right) \\
&\quad \Omega^{-1} \left(\int v(w_5)\Delta_n(w_5) \left(D_\beta g_{\beta_0}(w_5) - \frac{\int D_\beta g_{\beta_0}(w_6)\, dw_6}{\int v(w_7)\, dw_7} \right) dw_5 \right)^t \Bigg)^2 dx + o_p(1) \\
&= \tilde{T}_n - \delta_{1,n} + \delta_n + o_p(1) \\
&= \tilde{T}_n + \delta_{2,n} + \delta_{3,n} + o_p(1),
\end{aligned}$$

where the third from last equality was obtained by standard calculations. Finally, Lemma 2.3.2 leads to

$$T_n - b - \delta_n \xrightarrow{\mathcal{D}} Z.$$

□

2.8.4 Proof of Remark 2.3.5

δ_n was defined as

$$\begin{aligned}
\delta_n = \mu([0,1]) \int v(x) \Bigg(&\Delta_n(x) - \frac{\int v(w_1)\Delta_n(w_1)\, dw_1}{\int v(w_2)\, dw_2} - \left(D_\beta g_{\beta_0}(x) - \frac{\int D_\beta g_{\beta_0}(w_3)\, dw_3}{\int v(w_4)\, dw_4} \right) \\
&\Omega^{-1} \left(\int v(w_5)\Delta_n(w_5) \left(D_\beta g_{\beta_0}(w_5) - \frac{\int D_\beta g_{\beta_0}(w_6)\, dw_6}{\int v(w_7)\, dw_7} \right) dw_5 \right)^t \Bigg)^2 dx.
\end{aligned}$$

The alternative expression for δ_n can be obtained by simply expanding that from above. While doing so, the fact is used that

$$\int v(w_1)\frac{\int v(w_2)D_\beta g_{\beta_0}(w_2)\,dw_2}{\int v(w_3)\,dw_3}\left(D_\beta g_{\beta_0}(w_1)-\frac{\int v(w_4)D_\beta g_{\beta_0}(w_4)\,dw_4}{\int v(w_5)\,dw_5}\right)^t dw_1=0. \quad (2.64)$$

To prove the second assertion, rewrite Δ_n as

$$\Delta_n(x)=\frac{g_{\beta_n}(x)-g_{\beta_0}(x)}{c_n}=D_\beta g_{\beta_0}(x)\frac{\beta_n-\beta_0}{c_n}+o(1)$$

uniformly in $x\in\operatorname{supp}(v)$ and $n\in\mathbb{N}$. Hence, the distributive law yields

$$\begin{aligned}
\delta_n=\mu([0,1])\int v(x)\Bigg(& D_\beta g_{\beta_0}(x)\frac{\beta_n-\beta_0}{c_n}-\left(\frac{\int v(w_1)D_\beta g_{\beta_0}(w_1)\,dw_1}{\int v(w_2)\,dw_2}\right)\frac{\beta_n-\beta_0}{c_n}\\
&-\left(D_\beta g_{\beta_0}(x)-\frac{\int D_\beta g_{\beta_0}(w_3)\,dw_3}{\int v(w_4)\,dw_4}\right)\\
&\Omega^{-1}\left(\int v(w_5)\Delta_n(w_5)\left(D_\beta g_{\beta_0}(w_5)-\frac{\int D_\beta g_{\beta_0}(w_6)\,dw_6}{\int v(w_7)\,dw_7}\right)dw_5\right)^t\Bigg)^2 dx+o(1)\\
=\mu([0,1])\int v(x)\Bigg(&\left(D_\beta g_{\beta_0}(x)-\frac{\int v(w_1)D_\beta g_{\beta_0}(w_1)\,dw_1}{\int v(w_2)\,dw_2}\right)\\
&\left(\frac{\beta_n-\beta_0}{c_n}-\Omega^{-1}\left(\int v(w_5)\Delta_n(w_5)\left(D_\beta g_{\beta_0}(w_5)-\frac{\int D_\beta g_{\beta_0}(w_6)\,dw_6}{\int v(w_7)\,dw_7}\right)dw_5\right)^t\right)\Bigg)^2 dx\\
&+o(1).
\end{aligned}$$

Equation (2.64) leads to

$$\begin{aligned}
&\frac{\beta_n-\beta_0}{c_n}-\Omega^{-1}\left(\int v(w_5)\Delta_n(w_5)\left(D_\beta g_{\beta_0}(w_5)-\frac{\int D_\beta g_{\beta_0}(w_6)\,dw_6}{\int v(w_7)\,dw_7}\right)dw_5\right)^t\\
&=\frac{\beta_n-\beta_0}{c_n}-\Omega^{-1}\left(\int v(w_5)D_\beta g_{\beta_0}(w_5)\frac{\beta_n-\beta_0}{c_n}\left(D_\beta g_{\beta_0}(w_5)-\frac{\int D_\beta g_{\beta_0}(w_6)\,dw_6}{\int v(w_7)\,dw_7}\right)dw_5\right)^t\\
&\quad+o(1)\\
&=\frac{\beta_n-\beta_0}{c_n}-\Omega^{-1}\left(\frac{(\beta_n-\beta_0)^t}{c_n}\int v(w_5)\left(D_\beta g_{\beta_0}(w_5)-\frac{\int D_\beta g_{\beta_0}(w_3)\,dw_3}{\int v(w_4)\,dw_4}\right)^t\right.\\
&\qquad\left.\left(D_\beta g_{\beta_0}(w_5)-\frac{\int D_\beta g_{\beta_0}(w_6)\,dw_6}{\int v(w_7)\,dw_7}\right)dw_5\right)^t+o(1)\\
&=\frac{\beta_n-\beta_0}{c_n}-\Omega^{-1}\Omega\frac{\beta_n-\beta_0}{c_n}+o(1)\\
&=o(1),
\end{aligned}$$

that is, $\delta_n=o(1)$. $\square$

2.8.5 Proof of Theorem 2.3.6

The proof of the second part directly follows from Theorem 2.3.4 and Slutsky's theorem, so that it remains to prove the first assertion. Since H_0 is violated in this case, one has

$$\min_{\beta\in B,\, c\in\mathbb{R}} \int\int v(x)\big(F_{Y|X}^{-1}(\tau|x) - g_\beta(x) - c\big)^2\, dx\, \mu(d\tau) > 0.$$

Recall

$$\hat{F}_{Y|X}(y|x) - F_{Y|X}(y|x) = \frac{\hat{p}(y,x)}{\hat{f}_X(x)} - \frac{p(y,x)}{f_X(x)}$$

Again, the results of Hansen (2008) yield $\hat{p}(y,x) - p(y,x) = o_p(1)$ as well as $\hat{f}_X(x) - f_X(x) = o_p(1)$ uniformly on compact sets and thus

$$\hat{F}_{Y|X}(y|x) - F_{Y|X}(y|x) = o_p(1)$$

uniformly on $x \in \operatorname{supp}(v)$ and y belonging to some compact set $\mathcal{K} \subseteq \mathbb{R}$. When choosing $\mathcal{K} = [y_1, y_2]$ with

$$y_1 = \inf_{x\in\operatorname{supp}(v),\tau\in\operatorname{supp}(\mu)} F_{Y|X}^{-1}(\tau,x) \quad \text{and} \quad y_2 = \sup_{x\in\operatorname{supp}(v),\tau\in\operatorname{supp}(\mu)} F_{Y|X}^{-1}(\tau,x),$$

assumption (2.28) ensures that the functions $y \mapsto F_{Y|X}(y|x)$ are strictly increasing for all $x \in \operatorname{supp}(v)$, so that

$$\hat{F}_{Y|X}^{-1}(\tau|x) - F_{Y|X}^{-1}(\tau|x) = o_p(1)$$

uniformly on $x \in \operatorname{supp}(v)$ and $\tau \in \operatorname{supp}(\mu)$. Especially, it holds that

$$\sup_{x\in\operatorname{supp}(v),\tau\in\operatorname{supp}(\mu)} |\hat{F}_{Y|X}^{-1}(\tau|x)| \le \sup_{x\in\operatorname{supp}(v),\tau\in\operatorname{supp}(\mu)} |F_{Y|X}^{-1}(\tau|x)| + o_p(1)$$

and the minimization in (2.13) with respect to c can be replaced by that over some appropriate compact set $[c_1, c_2] \subseteq \mathbb{R}$ to obtain

$$\begin{aligned}
\frac{T_n}{nh_x^{\frac{d_X}{2}}} &= \min_{\beta\in B, c\in[c_1,c_2]} \int\int v(x)\big(\hat{F}_{Y|X}^{-1}(\tau|x) - g_\beta(x) - c\big)^2\, dx\, \mu(d\tau) + o_p(1) \\
&\ge \int\int v(x)\big(\hat{F}_{Y|X}^{-1}(\tau|x) - F_{Y|X}^{-1}(\tau|x)\big)^2\, dx\, \mu(d\tau) \\
&\quad - 2 \sup_{x\in\operatorname{supp}(v),\tau\in\operatorname{supp}(\mu)} \big|\hat{F}_{Y|X}^{-1}(\tau|x) - F_{Y|X}^{-1}(\tau|x)\big| \\
&\quad \max_{\beta\in B, c\in[c_1,c_2]} \int\int v(x)\big|F_{Y|X}^{-1}(\tau|x) - g_\beta(x) - c\big|\, dx\, \mu(d\tau) \\
&\quad + \min_{\beta\in B, c\in[c_1,c_2]} \int\int v(x)\big(F_{Y|X}^{-1}(\tau|x) - g_\beta(x) - c\big)^2\, dx\, \mu(d\tau) + o_p(1) \\
&= \min_{\beta\in B, c\in[c_1,c_2]} \int\int v(x)\big(F_{Y|X}^{-1}(\tau|x) - g_\beta(x) - c\big)^2\, dx\, \mu(d\tau) + o_p(1).
\end{aligned}$$

Hence

$$P(\Phi(Y_1, X_1, ..., Y_n, X_n) = 1)$$

$$= P\Big(T_n > \hat{b} + \sqrt{\hat{V}} u_{1-\alpha}\Big)$$

$$= P\left(\frac{T_n}{nh_x^{\frac{d_X}{2}}} > \frac{\hat{b} + \sqrt{\hat{V}} u_{1-\alpha}}{nh_x^{\frac{d_X}{2}}}\right)$$

$$= P\left(\min_{\beta \in B, c \in [c_1, c_2]} \int\int v(x)\big(F_{Y|X}^{-1}(\tau|x) - g_\beta(x) - c\big)^2 \, dx \, \mu(d\tau) > o_p(1)\right)$$

$$= 1 + o(1).$$

□

2.8.6 Proof of Theorem 2.4.1

Note that there exists a compact interval $\mathcal{C} = [c_1, c_2] \subseteq \mathbb{R}$ such that

$$(F_{Y|X}^h)^{-1}(\tau|x) \in (h(c_1), h(c_2)) \quad \text{for all } x \in \operatorname{supp}(v), \tau \in \operatorname{supp}(\mu). \tag{2.65}$$

Similar to the case without transformations, one has

$$\hat{F}_{Y|X}^{\hat{h}}(y|x) - F_{Y|X}^h(y|x)$$

$$= \frac{\hat{p}^{\hat{h}}(y,x)}{\hat{f}_X(x)} - \frac{p^h(y,x)}{f_X(x)}$$

$$= \frac{1}{f_X(x)}(p^{\hat{h}}(y,x) - p^h(y,x)) - \frac{p^h(y,x)}{f_X(x)^2}(\hat{f}_X(x) - f_X(x))$$

$$- \frac{\hat{f}_X(x) - f_X(x)}{\hat{f}_X(x) f_X(x)}\left(\hat{p}^{\hat{h}}(y,x) - p^h(y,x) - \frac{p^h(y,x)(\hat{f}_X(x) - f_X(x))}{f_X(x)}\right),$$

$$\hat{p}^h(y,x) - p^h(y,x) = o_p\big(n^{-\frac{1}{4}}\big) \tag{2.66}$$

and

$$\hat{f}_X(x) - f_X(x) = o_p\big(n^{-\frac{1}{4}}\big)$$

uniformly in $x \in \operatorname{supp}(v)$ and $y \in h(\mathcal{C})$. First, the asymptotic behaviour of $\hat{p}^{\hat{h}}(y,x) - \hat{p}^h(y,x)$ is examined. To this end, let $\delta > 0$. As the support of K is compact and $\hat{h}(y) \leq \hat{h}(c_1 - \delta) = h(c_1 - \delta) + o_p(1)$ uniformly in $y \in (-\infty, c_1 - \delta)$ and analogously $\hat{h}(y) \geq \hat{h}(c_2 + \delta) = h(c_2 + \delta) + o_p(1)$ uniformly in $y \in (c_2 + \delta, \infty)$ one has

$$P\Big(\forall z \in h(\mathcal{C}), y \notin [c_1 - \delta, c_2 + \delta] : \mathcal{K}_{h_y}\big(z - \hat{h}(y)\big) = \mathcal{K}_{h_y}\big(z - h(y)\big) \in \{0,1\}\Big) \to 1. \tag{2.67}$$

(2.41) yields

$$\left|\frac{1}{nh_y^j}\sum_{i=1}^n I_{\{Y_i \in [c_1-\delta, c_2+\delta]\}} K^{(j-1)}\left(\frac{y - h(Y_i)}{h_y}\right)(h(Y_i) - \hat{h}(Y_i))^j \mathbf{K}_{h_x}(x - X_i)\right|$$

$$\leq \frac{\sup_{y \in [c_1-\delta, c_2+\delta]} (h(y) - \hat{h}(y))^j}{h_y^{j-1}} \frac{1}{nh_y}\sum_{i=1}^n \left|K^{(j-1)}\left(\frac{y - h(Y_i)}{h_y}\right)\mathbf{K}_{h_x}(x - X_i)\right|$$

$$= \mathcal{O}_p\left(\frac{1}{\sqrt{n}^j h_y^{j-1}}\right)$$

$$= o_p\left(\frac{1}{\sqrt{n}}\right)$$

for all $j = 2, ..., r$ as well as

$$\left|\frac{1}{nh_y^{r+1}}\sum_{i=1}^n I_{\{Y_i\in[c_1-\delta,c_2+\delta]\}}K^{(r)}(y_i^*)(h(Y_i)-\hat{h}(Y_i))^{r+1}\mathbf{K}_{h_x}(x-X_i)\right|$$

$$\leq \frac{\sup\limits_{y\in[c_1-\delta,c_2+\delta]}(h(y)-\hat{h}(y))^{r+1}}{h_y^{r+1}}\sup_{y\in\mathbb{R}}K^{(r)}(y)\frac{1}{nh_y}\sum_{i=1}^n|\mathbf{K}_{h_x}(x-X_i)|$$

$$= \mathcal{O}_p\left(\frac{1}{\sqrt{n}^{r+1}h_y^{r+1}}\right)$$

$$= o_p\left(\frac{1}{\sqrt{n}}\right).$$

Hence, one has for appropriate $y_i^* \in \mathbb{R}, i = 1, ..., n$,

$$\begin{aligned}
&\hat{p}^{\hat{h}}(y,x) - \hat{p}^h(y,x)\\
&= \frac{1}{n}\sum_{i=1}^n \mathcal{K}_{h_y}\big(y-\hat{h}(Y_i)\big)\mathbf{K}_{h_x}(x-X_i) - \frac{1}{n}\sum_{i=1}^n \mathcal{K}_{h_y}\big(y-h(Y_i)\big)\mathbf{K}_{h_x}(x-X_i)\\
&= \frac{1}{n}\sum_{i=1}^n I_{\{Y_i\in[c_1-\delta,c_2+\delta]\}}\mathcal{K}_{h_y}\big(y-\hat{h}(Y_i)\big)\mathbf{K}_{h_x}(x-X_i)\\
&\quad - \frac{1}{n}\sum_{i=1}^n I_{\{Y_i\in[c_1-\delta,c_2+\delta]\}}\mathcal{K}_{h_y}\big(y-h(Y_i)\big)\mathbf{K}_{h_x}(x-X_i) + o_p\left(\frac{1}{\sqrt{n}}\right)\\
&= \frac{1}{n}\sum_{i=1}^n I_{\{Y_i\in[c_1-\delta,c_2+\delta]\}}\sum_{j=1}^r \frac{1}{h_y^j j!}K^{(j-1)}\left(\frac{y-h(Y_i)}{h_y}\right)(h(Y_i)-\hat{h}(Y_i))^j\mathbf{K}_{h_x}(x-X_i)\\
&\quad + \frac{1}{n}\sum_{i=1}^n I_{\{Y_i\in[c_1-\delta,c_2+\delta]\}}\frac{1}{h_y^{r+1}(r+1)!}K^{(r)}(y)\Big|_{y=y_i^*}(h(Y_i)-\hat{h}(Y_i))^{r+1}\mathbf{K}_{h_x}(x-X_i)\\
&\quad + o_p\left(\frac{1}{\sqrt{n}}\right)\\
&= \frac{1}{n^2}\sum_{i=1}^n\sum_{k=1}^n \psi(Y_k,X_k,Y_i)I_{\{Y_i\in[c_1-\delta,c_2+\delta]\}}K_{h_y}\big(y-h(Y_i)\big)\mathbf{K}_{h_x}(x-X_i) + o_p\left(\frac{1}{\sqrt{n}}\right) \qquad (2.68)
\end{aligned}$$

and

$$\begin{aligned}
&|\hat{p}^{\hat{h}}(y,x) - \hat{p}^h(y,x)|\\
&\quad = \left|\frac{1}{n}\sum_{i=1}^n I_{\{Y_i\in[c_1-\delta,c_2+\delta]\}}K_{h_y}(y-h(Y_i))(h(Y_i)-\hat{h}(Y_i))\mathbf{K}_{h_x}(x-X_i)\right| + o_p\left(\frac{1}{\sqrt{n}}\right)\\
&\quad \leq \sup_{z\in[c_1-\delta,c_2+\delta]}|h(z)-\hat{h}(z)|\frac{1}{n}\sum_{i=1}^n|K_{h_y}(y-h(Y_i))\mathbf{K}_{h_x}(x-X_i)| + o_p\left(\frac{1}{\sqrt{n}}\right)
\end{aligned}$$

$$\overset{(2.31)}{=} \mathcal{O}_p\Big(\frac{1}{\sqrt{n}}\Big) \tag{2.69}$$

uniformly in $y \in h(\mathcal{C}), x \in \operatorname{supp}(v)$. Due to (2.69), equation (2.66) can be extended to $\hat{p}^{\hat{h}}(y,x) - p^h(y,x) = o_p\big(n^{-\frac{1}{4}}\big)$, so that

$$\begin{aligned}
\hat{F}^{\hat{h}}_{Y|X}(y|x) - F^h_{Y|X}(y|x) &= \frac{1}{f_X(x)}(\hat{p}^{\hat{h}}(y,x) - p^h(y,x)) - \frac{p^h(y,x)}{f_X(x)^2}(\hat{f}_X(x) - f_X(x)) \\
&\quad + o_p\Big(\frac{1}{\sqrt{n}}\Big) \\
&= \frac{1}{f_X(x)}\hat{p}^{\hat{h}}(y,x) - \frac{p^h(y,x)}{f_X(x)^2}\hat{f}_X(x) + o_p\Big(\frac{1}{\sqrt{n}}\Big) \\
&= o_p\big(n^{-\frac{1}{4}}\big)
\end{aligned}$$

uniformly on $x \in \operatorname{supp}(v)$ and $y \in h(\mathcal{C})$. A similar reasoning leads to

$$\hat{f}^{\hat{h}}_{Y|X}(y|x) - f^{\hat{h}}_{Y|X}(y|x) = o_p\big(n^{-\frac{1}{4}}\big) \quad \text{and} \quad \frac{\partial}{\partial y}\hat{f}^{\hat{h}}_{Y|X}(y|x) = \mathcal{O}_p(1)$$

uniformly on $x \in \operatorname{supp}(v)$ and $y \in h(\mathcal{C})$, so that for an appropriate y^* one has

$$\begin{aligned}
0 &= \hat{F}^{\hat{h}}_{Y|X}((\hat{F}^{\hat{h}}_{Y|X})^{-1}(\tau|x)|x) - F^h_{Y|X}((F^h_{Y|X})^{-1}(\tau|x)|x) \\
&= \hat{F}^{\hat{h}}_{Y|X}((F^h_{Y|X})^{-1}(\tau|x)|x) + \hat{f}^{\hat{h}}_{Y|X}((F^h_{Y|X})^{-1}(\tau|x)|x)\big((\hat{F}^{\hat{h}}_{Y|X})^{-1}(\tau|x) - (F^h_{Y|X})^{-1}(\tau|x)\big) \\
&\quad + \frac{\partial}{\partial y}\hat{f}^{\hat{h}}_{Y|X}(y^*|x)\big((\hat{F}^{\hat{h}}_{Y|X})^{-1}(\tau|x) - (F^h_{Y|X})^{-1}(\tau|x)\big)^2 - F^h_{Y|X}((F^h_{Y|X})^{-1}(\tau|x)|x) \\
&= \hat{F}^{\hat{h}}_{Y|X}((F^h_{Y|X})^{-1}(\tau|x)|x) - F^h_{Y|X}((F^h_{Y|X})^{-1}(\tau|x)|x) \\
&\quad + f^h_{Y|X}((F^h_{Y|X})^{-1}(\tau|x)|x)\big((\hat{F}^{\hat{h}}_{Y|X})^{-1}(\tau|x) - (F^h_{Y|X})^{-1}(\tau|x)\big) + o_p\Big(\frac{1}{\sqrt{n}}\Big)
\end{aligned}$$

uniformly in $x \in \operatorname{supp}(v)$ and $\tau \in \operatorname{supp}(\mu)$. This in turn results in

$$\begin{aligned}
&(F^h_{Y|X})^{-1}(\tau|x) - (\hat{F}^{\hat{h}}_{Y|X})^{-1}(\tau|x) \\
&= \frac{\hat{F}^{\hat{h}}_{Y|X}((F^h_{Y|X})^{-1}(\tau|x)|x) - F^h_{Y|X}((F^h_{Y|X})^{-1}(\tau|x)|x)}{f^h_{Y|X}((F^h_{Y|X})^{-1}(\tau|x)|x)} + o_p\Big(\frac{1}{\sqrt{n}}\Big) \\
&= \frac{1}{f^h_{Y|X}((F^h_{Y|X})^{-1}(\tau|x)|x)}\bigg(\frac{1}{f_X(x)}\hat{p}^{\hat{h}}((F^h_{Y|X})^{-1}(\tau|x),x) - \frac{p^h((F^h_{Y|X})^{-1}(\tau|x),x)}{f_X(x)^2}\hat{f}_X(x)\bigg) \\
&\quad + o_p\Big(\frac{1}{\sqrt{n}}\Big) \\
&= \frac{1}{f^h_{Y|X}((F^h_{Y|X})^{-1}(\tau|x)|x)}\frac{1}{n}\sum_{i=1}^n \mathbf{K}_{h_x}(x - X_i)\bigg(\frac{1}{f_X(x)}\mathcal{K}_{h_y}((F^h_{Y|X})^{-1}(\tau|x) - \hat{h}(Y_i)) \\
&\quad - \frac{p^h((F^h_{Y|X})^{-1}(\tau|x),x)}{f_X(x)^2}\bigg) + o_p\Big(\frac{1}{\sqrt{n}}\Big)
\end{aligned} \tag{2.70}$$

uniformly in $x \in \operatorname{supp}(v)$ and $\tau \in \operatorname{supp}(\mu)$. Note that validity of H_0 is assumed, that is, $h(Y)$ here corresponds to $Y_0 = Y$ in Section 2.3.2. Therefore, equation (2.45) leads to

$$
\begin{aligned}
&(F^h_{Y|X})^{-1}(\tau|x) - (\hat{F}^h_{Y|X})^{-1}(\tau|x) \\
&= \frac{1}{f^h_{Y|X}((F^h_{Y|X})^{-1}(\tau|x)|x)}\bigg(\frac{1}{f_X(x)}\hat{p}^h((F^h_{Y|X})^{-1}(\tau|x),x) - \frac{p^h((F^h_{Y|X})^{-1}(\tau|x),x)}{f_X(x)^2}\hat{f}_X(x)\bigg) \\
&= \frac{1}{f^h_{Y|X}((F^h_{Y|X})^{-1}(\tau|x)|x)}\frac{1}{n}\sum_{i=1}^n \mathbf{K}_{h_x}(x-X_i)\bigg(\frac{1}{f_X(x)}\mathcal{K}_{h_y}((F^h_{Y|X})^{-1}(\tau|x) - h(Y_i)) \\
&\quad - \frac{p^h((F^h_{Y|X})^{-1}(\tau|x),x)}{f_X(x)^2}\bigg) + o_p\bigg(\frac{1}{\sqrt{n}}\bigg)
\end{aligned}
$$

uniformly in $x \in \operatorname{supp}(v)$ and $\tau \in \operatorname{supp}(\mu)$. Hence, (2.68), (2.69) and (2.70) yield

$$
\begin{aligned}
&(\hat{F}^h_{Y|X})^{-1}(\tau|x) - (\hat{F}^{\hat{h}}_{Y|X})^{-1}(\tau|x) \\
&= (F^h_{Y|X})^{-1}(\tau|x) - (\hat{F}^{\hat{h}}_{Y|X})^{-1}(\tau|x) - \big((F^h_{Y|X})^{-1}(\tau|x) - (\hat{F}^h_{Y|X})^{-1}(\tau|x)\big) \\
&\overset{(2.70)}{=} \frac{1}{f^h_{Y|X}((F^h_{Y|X})^{-1}(\tau|x)|x)f_X(x)}\big(\hat{p}^{\hat{h}}((F^h_{Y|X})^{-1}(\tau|x),x) - \hat{p}^h((F^h_{Y|X})^{-1}(\tau|x),x)\big) \\
&\quad + o_p\bigg(\frac{1}{\sqrt{n}}\bigg) \\
&\overset{(2.68)}{=} \frac{1}{f^h_{Y|X}((F^h_{Y|X})^{-1}(\tau|x)|x)f_X(x)n^2}\sum_{i=1}^n\sum_{k=1}^n \psi(Y_k,X_k,Y_i)I_{\{Y_i\in[c_1-\delta,c_2+\delta]\}} \\
&\quad K_{h_y}\big((F^h_{Y|X})^{-1}(\tau|x) - h(Y_i)\big)\mathbf{K}_{h_x}(x-X_i) + o_p\bigg(\frac{1}{\sqrt{n}}\bigg) \\
&\overset{(2.69)}{=} \mathcal{O}_p\bigg(\frac{1}{\sqrt{n}}\bigg)
\end{aligned}
\tag{2.71}
$$

uniformly in $x \in \operatorname{supp}(v)$ and $\tau \in \operatorname{supp}(\mu)$. Recall that $(F^h_{Y|X})^{-1}(\tau|\cdot) = g_{\beta_0}(\cdot)+c_0+F_\varepsilon^{-1}(\tau)$. Extend definitions (2.14) and (2.16) to

$$
\begin{aligned}
c^h_{\beta,\tau} &= \frac{\int v(x)((F^h_{Y|X})^{-1}(\tau|x) - g_\beta(x))\,dx}{\int v(x)\,dx} \\
\hat{c}^h_{\beta,\tau} &= \frac{\int v(x)((\hat{F}^h_{Y|X})^{-1}(\tau|x) - g_\beta(x))\,dx}{\int v(x)\,dx} \\
\hat{c}^{\hat{h}}_{\beta,\tau} &= \frac{\int v(x)((\hat{F}^{\hat{h}}_{Y|X})^{-1}(\tau|x) - g_\beta(x))\,dx}{\int v(x)\,dx}.
\end{aligned}
$$

Recall Theorem 2.3.4 and the definitions of δ_n there. Due to validity of H_0 it holds that $\delta_n = 0$. In the proof of Theorem 2.3.4 it was shown that

$$
\begin{aligned}
T^h_n &= \min_{\beta\in B} nh_x^{\frac{d_X}{2}}\int\int v(x)\big((\hat{F}^h_{Y|X})^{-1}(\tau|x) - g_\beta(x) - \hat{c}^h_{\beta,\tau}\big)^2\,dx\,\mu(d\tau) \\
&= nh_x^{\frac{d_X}{2}}\int\int v(x)\big((\hat{F}^h_{Y|X})^{-1}(\tau|x) - (F^h_{Y|X})^{-1}(\tau|x)\big)^2\,dx\,\mu(d\tau) + \mathcal{O}_p(1)
\end{aligned}
$$

$$= b + \mathcal{O}_p(1). \tag{2.72}$$

Similar to the proof of (2.14) one can show that

$$\hat{c}^{\hat{h}}_{\beta,\tau} - \hat{c}^{h}_{\beta,\tau} = \frac{\int v(x)\big((\hat{F}^{\hat{h}}_{Y|X})^{-1}(\tau|x) - (\hat{F}^{h}_{Y|X})^{-1}(\tau|x)\big)}{\int v(x)\,dx} \overset{(2.71)}{=} \mathcal{O}_p\Big(\frac{1}{\sqrt{n}}\Big) \tag{2.73}$$

and the same calculations as in (2.57) lead to (note that H_0 was assumed, that is $c_n = 0$)

$$\hat{c}^{\hat{h}}_{\beta,\tau} - c^{h}_{\beta,\tau} = \hat{c}^{\hat{h}}_{\beta,\tau} - \hat{c}^{h}_{\beta,\tau} + \frac{\int v(x)\big((\hat{F}^{h}_{Y|X})^{-1}(\tau|x) - (F^{h}_{Y|X})^{-1}(\tau|x)\big)}{\int v(x)\,dx}$$

$$= \mathcal{O}\Big(\frac{1}{\sqrt{n}}\Big) \tag{2.74}$$

uniformly in $\beta \in B$ and $\tau \in \mathrm{supp}(\mu)$.
Let $\hat{\beta}^h$ and $\hat{\beta}^{\hat{h}}$ be the minimizing values in T_n^h and $T_n^{\hat{h}}$, respectively.

Lemma 2.8.2 *Let $\bar{\beta}^{\hat{h}}$ and $\bar{\beta}^h$ denote the minimizers of $G^{\hat{h}} : B \times \mathbb{R} \to \mathbb{R}$,*

$$G^{\hat{h}}(\beta) = -2nh_x^{\frac{d_X}{2}} \int\int v(x)\big((\hat{F}^{\hat{h}}_{Y|X})^{-1}(\tau|x) - g_{\beta_0}(x) - c^{h}_{\beta_0,\tau}\big)$$

$$D_\beta(g_{\beta_0}(x) + c^{h}_{\beta_0,\tau})\,dx\,\mu(d\tau)(\beta - \beta_0) + nh_x^{\frac{d_X}{2}}(\beta - \beta_0)^t\Omega(\beta - \beta_0),$$

and $G^h : B \times \mathbb{R} \to \mathbb{R}$,

$$G^{h}(\beta) = -2nh_x^{\frac{d_X}{2}} \int\int v(x)\big((\hat{F}^{h}_{Y|X})^{-1}(\tau|x) - g_{\beta_0}(x) - c^{h}_{\beta_0,\tau}\big)$$

$$D_\beta(g_{\beta_0}(x) + c^{h}_{\beta_0,\tau})\,dx\,\mu(d\tau)(\beta - \beta_0)$$

$$+ nh_x^{\frac{d_X}{2}}(\beta - \beta_0)^t\Omega(\beta - \beta_0).$$

Define

$$\tilde{T}^h_n := nh_x^{\frac{d_X}{2}} \int\int v(x)\big((\hat{F}^{h}_{Y|X})^{-1}(\tau|x) - g_{\beta_0}(x) - c^{h}_{\beta_0,\tau}\big)^2\,dx\,\mu(d\tau)$$

$$= nh_x^{\frac{d_X}{2}} \int\int v(x)\big((\hat{F}^{h}_{Y|X})^{-1}(\tau|x) - (F^{h}_{Y|X})^{-1}(\tau|x)\big)^2\,dx\,\mu(d\tau).$$

Then, one has

$$||\hat{\beta}^{\hat{h}} - \beta_0|| = \mathcal{O}_p\big(n^{-\frac{1}{2}}h_x^{-\frac{d_X}{4}}\big), \tag{2.75}$$

$$||\hat{\beta}^{h} - \beta_0|| = \mathcal{O}_p\big(n^{-\frac{1}{2}}h_x^{-\frac{d_X}{4}}\big),$$

$$||\bar{\beta}^{\hat{h}} - \beta_0|| = \mathcal{O}_p\big(n^{-\frac{1}{2}}\big), \tag{2.76}$$

$$||\bar{\beta}^{h} - \beta_0|| = \mathcal{O}_p\big(n^{-\frac{1}{2}}\big)$$

and

$$nh_x^{\frac{d_X}{2}} \int\int v(x)\big((\hat{F}^{\hat{h}}_{Y|X})^{-1}(\tau|x) - g_{\beta_0}(x) - c^{h}_{\beta_0,\tau}\big)^2\,dx\,\mu(d\tau) = \tilde{T}^h_n + o_p(1).$$

Proof: It is started with proving the last assertion. Write

$$nh_x^{\frac{d_X}{2}}\int\int v(x)\big((\hat{F}^{\hat{h}}_{Y|X})^{-1}(\tau|x)-g_{\beta_0}(x)-c^h_{\beta_0,\tau}\big)^2\,dx\,\mu(d\tau)$$

$$=nh_x^{\frac{d_X}{2}}\int\int v(x)\big((\hat{F}^{\hat{h}}_{Y|X})^{-1}(\tau|x)-(\hat{F}^{h}_{Y|X})^{-1}(\tau|x)$$

$$+(\hat{F}^{h}_{Y|X})^{-1}(\tau|x)-g_{\beta_0}(x)-c^h_{\beta_0,\tau}\big)^2\,dx\,\mu(d\tau)$$

$$=nh_x^{\frac{d_X}{2}}\int\int v(x)\big((\hat{F}^{\hat{h}}_{Y|X})^{-1}(\tau|x)-(\hat{F}^{h}_{Y|X})^{-1}(\tau|x)\big)^2\,dx\,\mu(d\tau)+\tilde{T}^h_n$$

$$+2nh_x^{\frac{d_X}{2}}\int\int v(x)\big((\hat{F}^{\hat{h}}_{Y|X})^{-1}(\tau|x)-(\hat{F}^{h}_{Y|X})^{-1}(\tau|x)\big)$$

$$\big((\hat{F}^{h}_{Y|X})^{-1}(\tau|x)-g_{\beta_0}(x)-c^h_{\beta_0,\tau}\big)\,dx\,\mu(d\tau)$$

Thanks to (2.71) the first term is asymptotically negligible. In Lemma 2.3.2 it was shown that $\tilde{T}^h_n=\mathcal{O}_p\big(h_x^{-\frac{d_X}{2}}\big)$. Moreover, the third term can be expressed alternatively via (2.71) and Lemma 2.8.1 as

$$nh_x^{\frac{d_X}{2}}\int\int v(x)\big((\hat{F}^{\hat{h}}_{Y|X})^{-1}(\tau|x)-(\hat{F}^{h}_{Y|X})^{-1}(\tau|x)\big)$$

$$\big((\hat{F}^{h}_{Y|X})^{-1}(\tau|x)-g_{\beta_0}(x)-c^h_{\beta_0,\tau}\big)\,dx\,\mu(d\tau)$$

$$=nh_x^{\frac{d_X}{2}}\int\int v(x)\bigg(\frac{1}{f^h_{Y|X}((F^h_{Y|X})^{-1}(\tau|x)|x)f_X(x)n^2}\sum_{i=1}^n\sum_{k=1}^n\psi(Y_k,X_k,Y_i)$$

$$I_{\{Y_i\in[c_1-\delta,c_2+\delta]\}}K_{h_y}\big((F^h_{Y|X})^{-1}(\tau|x)-h(Y_i)\big)\mathbf{K}_{h_x}(x-X_i)+o_p\Big(\frac{1}{\sqrt{n}}\Big)\bigg)$$

$$\bigg(\frac{1}{f^h_{Y|X}((F^h_{Y|X})^{-1}(\tau|x)|x)}\frac{1}{n}\sum_{i=1}^n\mathbf{K}_{h_x}(x-X_i)\bigg(\frac{1}{f_X(x)}\mathcal{K}_{h_y}((F^h_{Y|X})^{-1}(\tau|x)-h(Y_i))$$

$$-\frac{p^h((F^h_{Y|X})^{-1}(\tau|x),x)}{f_X(x)^2}\bigg)+\mathcal{O}_p\Big(\frac{1}{\sqrt{n}}\Big)\bigg)\,dx\,\mu(d\tau)$$

$$=\frac{h_x^{\frac{d_X}{2}}}{n^2}\sum_{i=1}^n\sum_{k=1}^n\sum_{l=1}^n\int\int\kappa(x,\tau)\psi(Y_k,X_k,Y_i)I_{\{Y_i\in[c_1-\delta,c_2+\delta]\}}K_{h_y}\big((F^h_{Y|X})^{-1}(\tau|x)-h(Y_i)\big)$$

$$\mathbf{K}_{h_x}(x-X_i)\mathbf{K}_{h_x}(x-X_l)$$

$$\bigg(\mathcal{K}_{h_y}((F^h_{Y|X})^{-1}(\tau|x)-h(Y_l))-\frac{p^h((F^h_{Y|X})^{-1}(\tau|x),x)}{f_X(x)}\bigg)\,dx\,\mu(d\tau)+o_p(1),$$

where

$$\kappa(x,\tau)=\frac{v(x)}{f^h_{Y|X}((F^h_{Y|X})^{-1}(\tau|x)|x)^2f_X(x)^2}$$

has a compact support. For all compact sets $\mathcal{C}\subseteq\mathbb{R}$, the function $(y_1,x_1,y)\mapsto\psi(y_1,x_1,y)$ is uniformly bounded in $(y_1,x_1,y)\in\mathbb{R}^{d_X+1}\times\mathcal{C}$ due to assumption (**A9**). The sum can be split into

$$nh_x^{\frac{d_X}{2}}\int\int v(x)\big((\hat{F}^{\hat{h}}_{Y|X})^{-1}(\tau|x)-(\hat{F}^{h}_{Y|X})^{-1}(\tau|x)\big)$$

$$
\begin{aligned}
&\Big((\hat{F}^h_{Y|X})^{-1}(\tau|x) - g_{\beta_0}(x) - c^h_{\beta_0,\tau}\Big)\, dx\, \mu(d\tau)\\
&= \frac{h_x^{\frac{d_X}{2}}}{n^2} \sum_{i=1}^n \int\int \kappa(x,\tau)\psi(Y_i,X_i,Y_i) I_{\{Y_i\in[c_1-\delta,c_2+\delta]\}} K_{h_y}\big((F^h_{Y|X})^{-1}(\tau|x) - h(Y_i)\big)\\
&\qquad \mathbf{K}_{h_x}(x-X_i)^2 \bigg(\mathcal{K}_{h_y}((F^h_{Y|X})^{-1}(\tau|x) - h(Y_i)) - \frac{p^h((F^h_{Y|X})^{-1}(\tau|x),x)}{f_X(x)}\bigg)\, dx\, \mu(d\tau)\\
&\quad + \frac{h_x^{\frac{d_X}{2}}}{n^2} \sum_{i=1}^n \sum_{\substack{k=1\\k\neq i}}^n \int\int \kappa(x,\tau)\psi(Y_k,X_k,Y_i) I_{\{Y_i\in[c_1-\delta,c_2+\delta]\}} K_{h_y}\big((F^h_{Y|X})^{-1}(\tau|x) - h(Y_i)\big)\\
&\qquad \mathbf{K}_{h_x}(x-X_i)\mathbf{K}_{h_x}(x-X_k) \bigg(\mathcal{K}_{h_y}((F^h_{Y|X})^{-1}(\tau|x) - h(Y_k))\\
&\qquad - \frac{p^h((F^h_{Y|X})^{-1}(\tau|x),x)}{f_X(x)}\bigg)\, dx\, \mu(d\tau)\\
&\quad + \frac{h_x^{\frac{d_X}{2}}}{n^2} \sum_{i=1}^n \sum_{\substack{l=1\\l\neq i}}^n \int\int \kappa(x,\tau)\psi(Y_i,X_i,Y_i) I_{\{Y_i\in[c_1-\delta,c_2+\delta]\}} K_{h_y}\big((F^h_{Y|X})^{-1}(\tau|x) - h(Y_i)\big)\\
&\qquad \mathbf{K}_{h_x}(x-X_i)\mathbf{K}_{h_x}(x-X_l) \bigg(\mathcal{K}_{h_y}((F^h_{Y|X})^{-1}(\tau|x) - h(Y_l))\\
&\qquad - \frac{p^h((F^h_{Y|X})^{-1}(\tau|x),x)}{f_X(x)}\bigg)\, dx\, \mu(d\tau)\\
&\quad + \frac{h_x^{\frac{d_X}{2}}}{n^2} \sum_{i=1}^n \sum_{\substack{k=1\\k\neq i}}^n \int\int \kappa(x,\tau)\psi(Y_k,X_k,Y_i) I_{\{Y_i\in[c_1-\delta,c_2+\delta]\}} K_{h_y}\big((F^h_{Y|X})^{-1}(\tau|x) - h(Y_i)\big)\\
&\qquad \mathbf{K}_{h_x}(x-X_i)^2 \bigg(\mathcal{K}_{h_y}((F^h_{Y|X})^{-1}(\tau|x) - h(Y_i)) - \frac{p^h((F^h_{Y|X})^{-1}(\tau|x),x)}{f_X(x)}\bigg)\, dx\, \mu(d\tau)\\
&\quad + \frac{h_x^{\frac{d_X}{2}}}{n^2} \sum_{i=1}^n \sum_{\substack{k=1\\k\neq i}}^n \sum_{\substack{l=1\\l\neq i,k}}^n \int\int \kappa(x,\tau)\psi(Y_k,X_k,Y_i) I_{\{Y_i\in[c_1-\delta,c_2+\delta]\}}\\
&\qquad K_{h_y}\big((F^h_{Y|X})^{-1}(\tau|x) - h(Y_i)\big)\mathbf{K}_{h_x}(x-X_i)\mathbf{K}_{h_x}(x-X_l)\bigg(\mathcal{K}_{h_y}((F^h_{Y|X})^{-1}(\tau|x) - h(Y_l))\\
&\qquad - \frac{p^h((F^h_{Y|X})^{-1}(\tau|x),x)}{f_X(x)}\bigg)\, dx\, \mu(d\tau) + o_p(1)\\
&= I + II + III + IV + V + o_p(1).
\end{aligned}
$$

For an appropriate constant $C > 0$, term I can be bounded by

$$
I \leq \frac{C h_x^{\frac{d_X}{2}}}{n^2} \sum_{i=1}^n \int\int \kappa(x,\tau)|K_{h_y}((F^h_{Y|X})^{-1}(\tau|x) - h(Y_i))|\mathbf{K}_{h_x}(x-X_i)^2\, dx\, \mu(d\tau)
$$

$$= \frac{C}{n^2 h_x^{\frac{d_X}{2}}} \sum_{i=1}^{n} \int\int \kappa(X_i + h_x x, \tau) |K_{h_y}((F^h_{Y|X})^{-1}(\tau|X_i + h_x x) - h(Y_i))| \mathbf{K}(x)^2 \, dx \, \mu(d\tau),$$

which in turn is asymptotically negligible due to

$$\begin{aligned}
&E[\kappa(X_i + h_x x, \tau)|K_{h_y}((F^h_{Y|X})^{-1}(\tau|X_i + h_x x) - h(Y_i))|] \\
&= \int \kappa(w + h_x x, \tau) K_{h_y}((F^h_{Y|X})^{-1}(\tau|w + h_x x) - g_{\beta_0}(w) - c_0 - e) f_X(w) f_\varepsilon(e) \, dw \, de \\
&= \int \kappa(w + h_x x, \tau) K(e) f_X(w) f_\varepsilon((F^h_{Y|X})^{-1}(\tau|w + h_x x) - g_{\beta_0}(w) - c_0 - h_y e) \, dw \, de \\
&= \mathcal{O}(1)
\end{aligned}$$

uniformly in $x \in \operatorname{supp}(K)$ and $\tau \in \operatorname{supp}(\mu)$, that is $I = o_p(1)$. For the second term define

$$\begin{aligned}
Z^{II}_{i,k} = \int\int &\kappa(x, \tau) \psi(Y_k, X_k, Y_i) I_{\{Y_i \in [c_1 - \delta, c_2 + \delta]\}} K_{h_y}\big((F^h_{Y|X})^{-1}(\tau|x) - h(Y_i)\big) \mathbf{K}_{h_x}(x - X_i) \\
&\mathbf{K}_{h_x}(x - X_k) \bigg(\mathcal{K}_{h_y}((F^h_{Y|X})^{-1}(\tau|x) - h(Y_k)) - \frac{p^h((F^h_{Y|X})^{-1}(\tau|x), x)}{f_X(x)} \bigg) \, dx \, \mu(d\tau),
\end{aligned}$$

so that $II = \frac{h_x^{\frac{d_X}{2}}}{n^2} \sum_{i=1}^n \sum_{\substack{k=1 \\ k \neq i}}^n Z^{II}_{i,k}$. For an appropriate constant $C > 0$, the expectation of $|Z^{II}_{1,2}|$ can be bounded by

$$\begin{aligned}
E[|Z^{II}_{1,2}|] \leq C \int\int &\kappa(x, \tau) E\big[\big|K_{h_y}\big((F^h_{Y|X})^{-1}(\tau|x) - h(Y_1)\big) \mathbf{K}_{h_x}(x - X_1)\big|\big] \\
&E[|\mathbf{K}_{h_x}(x - X_1)|] \, dx \, \mu(d\tau) \\
= \mathcal{O}(1)&,
\end{aligned}$$

that is,

$$E\bigg[\bigg| \frac{h_x^{\frac{d_X}{2}}}{n^2} \sum_{i=1}^n \sum_{\substack{k=1 \\ k \neq i}}^n Z^{II}_{i,k} \bigg|\bigg] \leq \frac{h_x^{\frac{d_X}{2}}}{n^2} \sum_{i=1}^n \sum_{\substack{k=1 \\ k \neq i}}^n E\big[|Z^{II}_{i,k}|\big] = o(1).$$

Term III can be treated similarly to obtain $III = o_p(1)$.
For the fourth term define

$$\begin{aligned}
Z^{IV}_{i,k} = \int\int &\kappa(x, \tau) \psi(Y_k, X_k, Y_i) I_{\{Y_i \in [c_1 - \delta, c_2 + \delta]\}} K_{h_y}\big((F^h_{Y|X})^{-1}(\tau|x) - h(Y_i)\big) \mathbf{K}_{h_x}(x - X_i)^2 \\
&\bigg(\mathcal{K}_{h_y}((F^h_{Y|X})^{-1}(\tau|x) - h(Y_i)) - \frac{p^h((F^h_{Y|X})^{-1}(\tau|x), x)}{f_X(x)} \bigg) \, dx \, \mu(d\tau),
\end{aligned}$$

that is, $IV = \frac{h_x^{\frac{d_X}{2}}}{n^2} \sum_{i=1}^n \sum_{\substack{k=1 \\ k \neq i}}^n Z^{IV}_{i,k}$. As before, it can be shown that $\frac{h_x^{\frac{d_X}{2}}}{n^2} \sum_{i=1}^n Z^{IV}_{i,i} = o_p(1)$, so that for an appropriate constant $C > 0$

$$|IV| = \frac{h_x^{\frac{d_X}{2}}}{n^2} \bigg| \sum_{i=1}^n \sum_{k=1}^n Z^{IV}_{i,k} \bigg| + o_p(1)$$

$$\leq \frac{Ch_x^{\frac{d_X}{2}}}{n}\sum_{i=1}^n \int\int \kappa(x,\tau)\big|K_{h_y}\big((F_{Y|X}^h)^{-1}(\tau|x)-h(Y_i)\big)\big|\mathbf{K}_{h_x}(x-X_i)^2\,dx\,\mu(d\tau)$$
$$\underbrace{I_{\{Y_i\in[c_1-\delta,c_2+\delta]\}}\Big|\frac{1}{n}\sum_{k=1}^n \psi(Y_k,X_k,Y_i)\Big|}_{=\mathcal{O}_p\left(\frac{1}{\sqrt{n}}\right)}$$
$$= \mathcal{O}_p\Big(\frac{1}{\sqrt{n}}\Big)\frac{Ch_x^{\frac{d_X}{2}}}{n}\sum_{i=1}^n \int\int \kappa(x,\tau)\big|K_{h_y}\big((F_{Y|X}^h)^{-1}(\tau|x)-h(Y_i)\big)\big|$$
$$\mathbf{K}_{h_x}(x-X_i)^2\,dx\,\mu(d\tau)$$
$$= \mathcal{O}_p\Big(\frac{1}{\sqrt{nh_x^{d_X}}}\Big),$$

where the last equality follows similar to proving asymptotic negligibility of I, II and III. Hence, $IV = o_p(1)$.
It remains to examine term V. Define

$$Z_{i,k,l}^V = \int\int \kappa(x,\tau)\psi(Y_k,X_k,Y_i)I_{\{Y_i\in[c_1-\delta,c_2+\delta]\}}K_{h_y}\big((F_{Y|X}^h)^{-1}(\tau|x)-h(Y_i)\big)\mathbf{K}_{h_x}(x-X_i)$$
$$\mathbf{K}_{h_x}(x-X_l)\bigg(\mathcal{K}_{h_y}((F_{Y|X}^h)^{-1}(\tau|x)-h(Y_l))-\frac{p^h((F_{Y|X}^h)^{-1}(\tau|x),x)}{f_X(x)}\bigg)\,dx\,\mu(d\tau),$$

so that $V = \frac{h_x^{\frac{d_X}{2}}}{n^2}\sum_{i=1}^n\sum_{\substack{k=1\\k\neq i}}^n\sum_{\substack{l=1\\l\neq i,k}}^n Z_{i,k,l}^V$. One has

$$E[V^2] = \frac{h_x^{d_X}}{n^4}\sum_{i=1}^n\sum_{\substack{k=1\\k\neq i}}^n\sum_{\substack{l=1\\l\neq i,k}}^n\sum_{s=1}^n\sum_{\substack{t=1\\t\neq s}}^n\sum_{\substack{u=1\\u\neq s,t}}^n E\big[Z_{i,k,l}^V Z_{s,t,u}^V\big].$$

Due to $E[\psi(Y_2,X_2,Y_1)|Y_1]=0$ the expectation vanishes whenever k or t are occurring only once in (i,k,l,s,t,u). Only asymptotic negligibility of the summand corresponding to the case, in which $k=t$ and $\#\{i,k,l,s,t,u\}=5$, will be shown, since asymptotic negligibility of the remaining summands can be deduced from this case and the calculations for terms I, II, III, IV. It holds that

$$E\big[Z_{1,2,3}^V Z_{4,2,5}^V\big]$$
$$= E\bigg[\int\int \kappa(x,\tau)\psi(Y_2,X_2,Y_1)\psi(Y_2,X_2,Y_4)I_{\{Y_1\in[c_1-\delta,c_2+\delta]\}}I_{\{Y_4\in[c_1-\delta,c_2+\delta]\}}$$
$$K_{h_y}\big((F_{Y|X}^h)^{-1}(\tau|x)-h(Y_1)\big)\mathbf{K}_{h_x}(x-X_1)K_{h_y}\big((F_{Y|X}^h)^{-1}(\tau|x)-h(Y_4)\big)\mathbf{K}_{h_x}(x-X_4)$$
$$\mathbf{K}_{h_x}(x-X_3)\mathbf{K}_{h_x}(x-X_5)\bigg(\mathcal{K}_{h_y}((F_{Y|X}^h)^{-1}(\tau|x)-h(Y_3))-\frac{p^h((F_{Y|X}^h)^{-1}(\tau|x),x)}{f_X(x)}\bigg)$$
$$\bigg(\mathcal{K}_{h_y}((F_{Y|X}^h)^{-1}(\tau|x)-h(Y_5))-\frac{p^h((F_{Y|X}^h)^{-1}(\tau|x),x)}{f_X(x)}\bigg)\,dx\,\mu(d\tau)\bigg]$$

$$= E\Big[\int\int \kappa(x,\tau)\psi(Y_2,X_2,Y_1)\psi(Y_2,X_2,Y_4)I_{\{Y_1\in[c_1-\delta,c_2+\delta]\}}I_{\{Y_4\in[c_1-\delta,c_2+\delta]\}}$$

$$K_{h_y}\big((F^h_{Y|X})^{-1}(\tau|x)-h(Y_1)\big)\mathbf{K}_{h_x}(x-X_1)K_{h_y}\big((F^h_{Y|X})^{-1}(\tau|x)-h(Y_4)\big)\mathbf{K}_{h_x}(x-X_4)$$

$$E\Big[\mathbf{K}_{h_x}(x-X_3)\Big(\mathcal{K}_{h_y}((F^h_{Y|X})^{-1}(\tau|x)-h(Y_3))-\frac{p^h((F^h_{Y|X})^{-1}(\tau|x),x)}{f_X(x)}\Big)\Big]^2\,dx\,\mu(d\tau)\Big]$$

As in (2.48) the inner expectation can be bounded via

$$E\Big[\mathbf{K}_{h_x}(x-X_3)\Big(\mathcal{K}_{h_y}((F^h_{Y|X})^{-1}(\tau|x)-h(Y_3))-\frac{p^h((F^h_{Y|X})^{-1}(\tau|x),x)}{f_X(x)}\Big)\Big]$$

$$=\int \mathbf{K}(w)\Big(\int K(u)p^h((F^h_{Y|X})^{-1}(\tau|x)-h_yu,x-h_xw)\,dz\,du$$

$$-p^h((F^h_{Y|X})^{-1}(\tau|x),x)\frac{f_X(x-h_xw)}{f_X(x)}\Big)\,dw$$

$$=o_p\Big(\frac{1}{\sqrt{n}}\Big) \tag{2.77}$$

uniformly in $x\in\operatorname{supp}(v)$ and $\tau\in\operatorname{supp}(\mu)$. By the same reasoning as before this results in $E\big[Z^V_{1,2,3}Z^V_{4,2,5}\big]=o\big(\frac{1}{n}\big)$ and thus

$$E[V^2]=\frac{h_x^{d_X}}{n^4}\sum_{i=1}^n\sum_{\substack{k=1\\k\neq i}}^n\sum_{\substack{l=1\\l\neq i,k}}^n\sum_{\substack{s=1\\s\neq i,k,l}}^n\sum_{\substack{u=1\\u\neq i,k,l,s}}^n E\big[Z^V_{i,k,l}Z^V_{s,k,u}\big]=o(1).$$

Finally, this leads to $V=o_p(1)$, that is

$$nh_x^{\frac{d_X}{2}}\int\int v(x)\big((\hat F^{\hat h}_{Y|X})^{-1}(\tau|x)-g_{\beta_0}(x)-c^h_{\beta_0,\tau}\big)^2\,dx\,\mu(d\tau)=\tilde T^h_n+o_p(1).$$

Treatment of $\hat\beta^h$, $\hat\beta^{\hat h}$, $\bar\beta^h$ and $\bar\beta^{\hat h}$

For treating $\hat\beta^h$ and $\hat\beta^{\hat h}$ note that

$$\int\int v(x)\big((\hat F^{\hat h}_{Y|X})^{-1}(\tau|x)-g_\beta(x)-\hat c^{\hat h}_{\beta,\tau}\big)^2\,dx\,\mu(d\tau)$$

$$=\int\int v(x)\big((\hat F^{h}_{Y|X})^{-1}(\tau|x)-g_\beta(x)-\hat c^{h}_{\beta,\tau}\big)^2\,dx\,\mu(d\tau)+o_p(1)$$

$$=\int\int v(x)\big((F^{h}_{Y|X})^{-1}(\tau|x)-g_\beta(x)-c^{h}_{\beta,\tau}\big)^2\,dx\,\mu(d\tau)+o_p(1)$$

uniformly in $\beta\in B$ with

$$\sup_{\beta\in B,||\beta-\beta_0||>\delta}\int\int v(x)\big((F^{h}_{Y|X})^{-1}(\tau|x)-g_\beta(x)-c^{h}_{\beta,\tau}\big)^2\,dx\,\mu(d\tau)>0$$

for all $\delta>0$, which leads because of (**A7**) to $||\hat\beta^h-\beta_0||=o_p(1)$ and $||\hat\beta^{\hat h}-\beta_0||=o_p(1)$. Due to (2.72), one has $T^h_n-b=\mathcal{O}_p(1)$ and a Taylor expansion of $\beta\mapsto\big((\hat F^{\hat h}_{Y|X})^{-1}-g_\beta(x)-c^h_{\beta,\tau}\big)^2$ (compare (2.59)) yields

$$T^h_n-b$$

$$
\begin{aligned}
&\overset{(2.72)}{=} nh_x^{\frac{d_X}{2}} \int\int v(x)\big((\hat{F}_{Y|X}^h)^{-1}(\tau|x) - g_{\beta_0}(x) - c_{\beta_0,\tau}^h\big)^2 \, dx\, \mu(d\tau) - b + \mathcal{O}_p(1) \\
&= nh_x^{\frac{d_X}{2}} \int\int v(x)\big((\hat{F}_{Y|X}^h)^{-1}(\tau|x) - g_{\hat{\beta}^{\hat{h}}}(x) - c_{\hat{\beta}^{\hat{h}},\tau}^h\big)^2 \, dx\, \mu(d\tau) - b \\
&\quad - 2nh_x^{\frac{d_X}{2}} \int\int v(x)\big((\hat{F}_{Y|X}^h)^{-1}(\tau|x) - g_{\hat{\beta}^{\hat{h}}}(x) - c_{\hat{\beta}^{\hat{h}},\tau}^h\big) D_\beta\big(g_{\hat{\beta}^{\hat{h}}}(x) + c_{\hat{\beta}^{\hat{h}},\tau}^h\big) \\
&\qquad dx\, \mu(d\tau)(\beta_0 - \hat{\beta}^{\hat{h}}) + (\beta_0 - \hat{\beta}^{\hat{h}})^t nh_x^{\frac{d_X}{2}} \Omega(\beta_0 - \hat{\beta}^{\hat{h}}) + o_p\big(nh_x^{\frac{d_X}{2}} ||\beta_0 - \hat{\beta}^{\hat{h}}||^2\big) + \mathcal{O}_p(1) \\
&\geq T_n^h - b + nh_x^{\frac{d_X}{2}} (\beta_0 - \hat{\beta}^{\hat{h}})^t \Omega(\beta_0 - \hat{\beta}^{\hat{h}}) - 2nh_x^{\frac{d_X}{2}} \int\int v(x) \\
&\qquad \big((\hat{F}_{Y|X}^h)^{-1}(\tau|x) - g_{\hat{\beta}^{\hat{h}}}(x) - \hat{c}_{\hat{\beta}^{\hat{h}},\tau}^{\hat{h}}\big) D_\beta\big(g_{\hat{\beta}^{\hat{h}}}(x) + \hat{c}_{\hat{\beta}^{\hat{h}},\tau}^{\hat{h}}\big) \, dx\, \mu(d\tau)(\beta_0 - \hat{\beta}^{\hat{h}}) \\
&\quad + \mathcal{O}_p\big(\sqrt{nh_x^{d_X}} ||\beta_0 - \hat{\beta}^{\hat{h}}||\big) + o_p\big(nh_x^{\frac{d_X}{2}} ||\beta_0 - \hat{\beta}^{\hat{h}}||^2\big) + \mathcal{O}_p(1) \\
&= (\beta_0 - \hat{\beta}^{\hat{h}})^t nh_x^{\frac{d_X}{2}} \Omega(\beta_0 - \hat{\beta}^{\hat{h}}) + \mathcal{O}_p\big(\sqrt{nh_x^{d_X}} ||\beta_0 - \hat{\beta}^{\hat{h}}||\big) + o_p\big(nh_x^{\frac{d_X}{2}} ||\beta_0 - \hat{\beta}^{\hat{h}}||^2\big) + \mathcal{O}_p(1).
\end{aligned} \tag{2.78}
$$

Here, the second to last inequality, where $\hat{c}_{\hat{\beta}^{\hat{h}},\tau}^h$ was replaced with $c_{\hat{\beta}^{\hat{h}},\tau}^{\hat{h}}$, follows from (compare (2.58) and (2.74))

$$
\begin{aligned}
&nh_x^{\frac{d_X}{2}} \int\int v(x)\big((\hat{F}_{Y|X}^h)^{-1}(\tau|x) - g_{\hat{\beta}^{\hat{h}}}(x) - c_{\hat{\beta}^{\hat{h}},\tau}^h\big) D_\beta\big(g_{\hat{\beta}^{\hat{h}}}(x) + c_{\hat{\beta}^{\hat{h}},\tau}^h\big) \, dx\, \mu(d\tau) \\
&= nh_x^{\frac{d_X}{2}} \int\int v(x)\big((\hat{F}_{Y|X}^h)^{-1}(\tau|x) - g_{\hat{\beta}^{\hat{h}}}(x) - \hat{c}_{\hat{\beta}^{\hat{h}},\tau}^{\hat{h}}\big) D_\beta\big(g_{\hat{\beta}^{\hat{h}}}(x) + \hat{c}_{\hat{\beta}^{\hat{h}},\tau}^{\hat{h}}\big) \, dx\, \mu(d\tau) \\
&\quad + \mathcal{O}_p\big(\sqrt{nh_x^{d_X}}\big) \\
&= nh_x^{\frac{d_X}{2}} \int\int v(x)\big((\hat{F}_{Y|X}^h)^{-1}(\tau|x) - (\hat{F}_{Y|X}^{\hat{h}})^{-1}(\tau|x) + (\hat{F}_{Y|X}^{\hat{h}})^{-1}(\tau|x) - g_{\hat{\beta}^{\hat{h}}}(x) - \hat{c}_{\hat{\beta}^{\hat{h}},\tau}^{\hat{h}}\big) \\
&\qquad D_\beta\big(g_{\hat{\beta}^{\hat{h}}}(x) + \hat{c}_{\hat{\beta}^{\hat{h}},\tau}^{\hat{h}}\big) \, dx\, \mu(d\tau) + \mathcal{O}_p\big(\sqrt{nh_x^{d_X}}\big) \\
&\overset{(2.71)}{=} nh_x^{\frac{d_X}{2}} \int\int v(x)\big((\hat{F}_{Y|X}^{\hat{h}})^{-1}(\tau|x) - g_{\hat{\beta}^{\hat{h}}}(x) - \hat{c}_{\hat{\beta}^{\hat{h}},\tau}^{\hat{h}}\big)\big(D_\beta g_{\hat{\beta}^{\hat{h}}}(x) + \hat{c}_{\hat{\beta}^{\hat{h}},\tau}^{\hat{h}}\big) \, dx\, \mu(d\tau) \\
&\quad + \mathcal{O}_p\big(\sqrt{nh_x^{d_X}}\big).
\end{aligned}
$$

The last equality in (2.78) follows from the definition of $\hat{\beta}^{\hat{h}}$ as the minimizer of T_n^h, which implies

$$
nh_x^{\frac{d_X}{2}} \int\int v(x)\big((\hat{F}_{Y|X}^{\hat{h}})^{-1}(\tau|x) - g_{\hat{\beta}^{\hat{h}}}(x) - \hat{c}_{\hat{\beta}^{\hat{h}},\tau}^{\hat{h}}\big) D_\beta\big(g_{\hat{\beta}^{\hat{h}}}(x) + \hat{c}_{\hat{\beta}^{\hat{h}},\tau}^{\hat{h}}\big) \, dx\, \mu(d\tau) = 0.
$$

Since Ω is positive definite, equation (2.78) leads to $||\hat{\beta}^{\hat{h}} - \beta_0|| = \mathcal{O}_p\big(n^{-\frac{1}{2}} h_x^{-\frac{d_X}{4}}\big)$. The same assertion for $\hat{\beta}^h$ was already shown in the proof of Theorem 2.3.4.
As the minimizer of $G^{\hat{h}}(\beta)$, $\bar{\beta}^{\hat{h}}$ it is determined by

$$
0 = D_\beta G^{\hat{h}}(\beta)
$$

$$= -2nh_x^{\frac{d_X}{2}} \int\int v(x)\big((\hat{F}^{\hat{h}}_{Y|X})^{-1}(\tau|x) - g_{\beta_0}(x) - c^h_{\beta_0,\tau}\big) D_\beta(g_{\beta_0}(x) + c^h_{\beta_0,\tau})\,dx\,\mu(d\tau)$$

$$+ 2nh_x^{\frac{d_X}{2}}(\beta-\beta_0)^t nh_x^{\frac{d_X}{2}}\Omega$$

and consequently can be expressed as

$$\bar{\beta}^{\hat{h}} = \beta_0 + \Omega^{-1}\int\int v(x)\big((\hat{F}^{\hat{h}}_{Y|X})^{-1}(\tau|x) - (F^h_{Y|X})^{-1}(\tau|x)\big) D_\beta(g_{\beta_0}(x) + c^h_{\beta_0,\tau})^t\,dx\,\mu(d\tau)$$

$$= \beta_0 + \Omega^{-1}\int\int v(x)\big((\hat{F}^{h}_{Y|X})^{-1}(\tau|x) - (F^h_{Y|X})^{-1}(\tau|x)\big) D_\beta(g_{\beta_0}(x) + c^h_{\beta_0,\tau})^t\,dx\,\mu(d\tau)$$

$$+ \mathcal{O}_p\Big(\frac{1}{\sqrt{n}}\Big) \tag{2.79}$$

$$= \beta_0 - \Omega^{-1}\frac{1}{n}\sum_{i=1}^n \int\int \frac{v(x)}{f^h_{Y|X}((F^h_{Y|X})^{-1}(\tau|x)|x)f_X(x)} D_\beta(g_{\beta_0}(x) + c^h_{\beta_0,\tau})^t \mathbf{K}_{h_x}(x-X_i)$$

$$\bigg(\mathcal{K}_{h_y}((F^h_{Y|X})^{-1}(\tau|x) - h(Y_i)) - \frac{p^h((F^h_{Y|X})^{-1}(\tau|x),x)}{f_X(x)}\bigg)\,dx\,\mu(d\tau) + \mathcal{O}_p\Big(\frac{1}{\sqrt{n}}\Big)$$

$$= \beta_0 - \frac{1}{n}\sum_{i=1}^n \int\int \tilde{\kappa}(x)\mathbf{K}_{h_x}(x-X_i)\bigg(\mathcal{K}_{h_y}((F^h_{Y|X})^{-1}(\tau|x) - h(Y_i))$$

$$- \frac{p^h((F^h_{Y|X})^{-1}(\tau|x),x)}{f_X(x)}\bigg)\,dx\,\mu(d\tau) + \mathcal{O}_p\Big(\frac{1}{\sqrt{n}}\Big),$$

where

$$\tilde{\kappa}(x) = \Omega^{-1}\frac{v(x)}{f^h_{Y|X}((F^h_{Y|X})^{-1}(\tau|x)|x)f_X(x)} D_\beta(g_{\beta_0}(x) + c^h_{\beta_0,\tau})^t$$

is a (multidimensional) function with compact support. To show $||\bar{\beta}^{\hat{h}} - \beta_0|| = \mathcal{O}_p(\frac{1}{\sqrt{n}})$ it is sufficient to prove

$$E\Bigg[\bigg(\frac{1}{n}\sum_{i=1}^n \int\int \tilde{\kappa}_k(x)\mathbf{K}_{h_x}(x-X_i)$$

$$\bigg(\mathcal{K}_{h_y}((F^h_{Y|X})^{-1}(\tau|x) - h(Y_i)) - \frac{p^h((F^h_{Y|X})^{-1}(\tau|x),x)}{f_X(x)}\bigg)\,dx\,\mu(d\tau)\bigg)^2\Bigg] = \mathcal{O}\Big(\frac{1}{n}\Big).$$

for each component $\tilde{\kappa}_k$ of $\tilde{\kappa}$, $k = 1, ..., d_B$. This in turn leads to analysing

$$E\Bigg[\bigg(\int\int \tilde{\kappa}_k(x)\mathbf{K}_{h_x}(x-X_1)$$

$$\bigg(\mathcal{K}_{h_y}((F^h_{Y|X})^{-1}(\tau|x) - h(Y_1)) - \frac{p^h((F^h_{Y|X})^{-1}(\tau|x),x)}{f_X(x)}\bigg)\,dx\,\mu(d\tau)\bigg)^2\Bigg]$$

and

$$E\Bigg[\int\int \tilde{\kappa}_k(x)\mathbf{K}_{h_x}(x-X_1)$$

$$\bigg(\mathcal{K}_{h_y}((F^h_{Y|X})^{-1}(\tau|x) - h(Y_1)) - \frac{p^h((F^h_{Y|X})^{-1}(\tau|x),x)}{f_X(x)}\bigg)\,dx\,\mu(d\tau)\Bigg]^2.$$

For some sufficiently large $C > 0$ the first expectation can be bounded by

$$\begin{aligned}
&E\bigg[\bigg(\int\int \tilde{\kappa}_k(x)\mathbf{K}_{h_x}(x-X_1)\bigg(\mathcal{K}_{h_y}((F^h_{Y|X})^{-1}(\tau|x)-h(Y_1)) \\
&\quad -\frac{p^h((F^h_{Y|X})^{-1}(\tau|x),x)}{f_X(x)}\bigg)\,dx\,\mu(d\tau)\bigg)^2\bigg] \\
&\leq CE\bigg[\bigg(\int |\mathbf{K}_{h_x}(x-X_1)|\,dx\bigg)^2\bigg] \\
&\leq C\bigg(\int |\mathbf{K}(x)|\,dx\bigg)^2,
\end{aligned}$$

while the second expectation can be treated as in (2.48). Finally, $||\bar{\beta}^{\hat{h}}-\beta_0|| = \mathcal{O}_p\big(\frac{1}{\sqrt{n}}\big)$ has been proven. Additionally, due to (2.79) it was shown that

$$\begin{aligned}
&||\bar{\beta}^h-\beta_0|| \\
&= \bigg|\bigg|\Omega^{-1}\int\int v(x)\big((\hat{F}^h_{Y|X})^{-1}(\tau|x)-(F^h_{Y|X})^{-1}(\tau|x)\big)D_\beta\big(g_{\beta_0}(x)+c^h_{\beta_0,\tau}\big)^t\,dx\,\mu(d\tau)\bigg|\bigg| \\
&= \mathcal{O}_p\bigg(\frac{1}{\sqrt{n}}\bigg).
\end{aligned}$$

□

Putting Things together

Let $\beta = (\beta_n)_{n\in\mathbb{N}}$ be a sequence in B with $\beta-\beta_0 = \mathcal{O}_p\big(n^{-\frac{1}{2}}h_x^{-\frac{d_X}{4}}\big)$. Then, as in (2.59) a Taylor expansion of $\beta \mapsto \big((\hat{F}^{\hat{h}}_{Y|X})^{-1}-g_\beta(x)-c_{\beta,\tau}\big)^2$ and the binomial formula yield for some β^* between $\hat{\beta}$ and β_0

$$\begin{aligned}
&nh_x^{\frac{d_X}{2}}\int\int v(x)\big((\hat{F}^{\hat{h}}_{Y|X})^{-1}(\tau|x)-g_\beta(x)-\hat{c}^{\hat{h}}_{\beta,\tau}\big)^2\,dx\,\mu(d\tau) \\
&= nh_x^{\frac{d_X}{2}}\int\int v(x)\big((\hat{F}^{\hat{h}}_{Y|X})^{-1}(\tau|x)-g_{\beta_0}(x)-\hat{c}^{\hat{h}}_{\beta_0,\tau}\big)^2\,dx\,\mu(d\tau) \\
&\quad -2nh_x^{\frac{d_X}{2}}\int\int v(x)\big((\hat{F}^{\hat{h}}_{Y|X})^{-1}(\tau|x)-g_{\beta_0}(x)-\hat{c}^{\hat{h}}_{\beta_0,\tau}\big) \\
&\qquad D_\beta\big(g_{\beta_0}(x)+\hat{c}^{\hat{h}}_{\beta_0,\tau}\big)\,dx\,\mu(d\tau)(\beta-\beta_0)+nh_x^{\frac{d_X}{2}}(\beta-\beta_0)^t\Omega(\beta-\beta_0)+o_p(1) \\
&= nh_x^{\frac{d_X}{2}}\int\int v(x)\big((\hat{F}^{\hat{h}}_{Y|X})^{-1}(\tau|x)-g_{\beta_0}(x)-c^h_{\beta_0,\tau}\big)^2\,dx\,\mu(d\tau) \\
&\quad +nh_x^{\frac{d_X}{2}}\int\int v(x)(c^h_{\beta_0,\tau}-\hat{c}^{\hat{h}}_{\beta_0,\tau})^2\,dx\,\mu(d\tau) \\
&\quad +2nh_x^{\frac{d_X}{2}}\int\int v(x)\big((\hat{F}^{\hat{h}}_{Y|X})^{-1}(\tau|x)-g_{\beta_0}(x)-c^h_{\beta_0,\tau}\big)(c^h_{\beta_0,\tau}-\hat{c}^{\hat{h}}_{\beta_0,\tau})\,dx\,\mu(d\tau) \\
&\quad -2nh_x^{\frac{d_X}{2}}\int\int v(x)\big((\hat{F}^{\hat{h}}_{Y|X})^{-1}(\tau|x)-g_{\beta_0}(x)-c^h_{\beta_0,\tau}\big)
\end{aligned}$$

$$D_\beta\big(g_{\beta_0}(x) + c^h_{\beta_0,\tau}\big)\,dx\,\mu(d\tau)(\beta-\beta_0) + nh_x^{\frac{d_X}{2}}(\beta-\beta_0)^t\Omega(\beta-\beta_0) + o_p(1)$$
$$= \tilde{T}_n^h + G^{\hat{h}}(\beta) + o_p(1) \tag{2.80}$$

and

$$nh_x^{\frac{d_X}{2}} \int\int v(x)\big((\hat{F}^h_{Y|X})^{-1}(\tau|x) - g_\beta(x) - c^h_{\beta,\tau}\big)^2\,dx\,\mu(d\tau) = \tilde{T}_n^h + G^h(\beta) + o_p(1).$$

Note that in contrast to the proof of Theorem 2.3.4, equation (2.74) leads to asymptotic negligibility of the terms containing $c^h_{\beta_0,\tau} - \hat{c}^{\hat{h}}_{\beta_0,\tau} = \mathcal{O}_p\big(n^{-\frac{1}{2}}\big)$. Due to (2.75) and (2.76), one has

$$\begin{aligned}
T_n^{\hat{h}} &= \tilde{T}_n^h + G^{\hat{h}}(\hat{\beta}^{\hat{h}}) + o_p(1)\\
&\geq \tilde{T}_n^h + G^{\hat{h}}(\bar{\beta}^{\hat{h}}) + o_p(1)\\
&= nh_x^{\frac{d_X}{2}} \int\int v(x)\Big((\hat{F}^{\hat{h}}_{Y|X})^{-1}(\tau|x) - g_{\bar{\beta}^{\hat{h}}}(x) - \hat{c}^{\hat{h}}_{\bar{\beta}^{\hat{h}},\tau}\Big)^2\,dx\,\mu(d\tau) + o_p(1)\\
&\geq T_n^{\hat{h}} + o_p(1)
\end{aligned}$$

so that it suffices to consider the minimum of $\beta \mapsto \tilde{T}_n^h + G^{\hat{h}}(\beta)$ in (2.80). Therefore, $T_n^{\hat{h}}$ is asymptotically equivalent to

$$\begin{aligned}
&\tilde{T}_n^h + G^{\hat{h}}(\bar{\beta}^{\hat{h}})\\
&= \tilde{T}_n^h - 2nh_x^{\frac{d_X}{2}} \int\int v(x)\big((\hat{F}^{\hat{h}}_{Y|X})^{-1}(\tau|x) - g_{\beta_0}(x) - c^h_{\beta_0,\tau}\big)\\
&\qquad D_\beta(g_{\beta_0}(x) + c^h_{\beta_0,\tau})\,dx\,\mu(d\tau)(\bar{\beta}^{\hat{h}} - \beta_0) + nh_x^{\frac{d_X}{2}}(\bar{\beta}^{\hat{h}} - \beta_0)^t\Omega(\bar{\beta}^{\hat{h}} - \beta_0)\\
&= \tilde{T}_n^h - nh_x^{\frac{d_X}{2}}(\bar{\beta}^{\hat{h}} - \beta_0)^t\Omega(\bar{\beta}^{\hat{h}} - \beta_0)\\
&\overset{(2.76)}{=} \tilde{T}_n^h + \mathcal{O}_p\big(h_x^{\frac{d_X}{2}}\big).
\end{aligned}$$

Recall that H_0 was assumed, that is $\delta_n = \delta_{1,n} = \delta_{2,n} = \delta_{3,n} = 0$ with $\delta_{1,n}, \delta_{2,n}, \delta_{3,n}$ from Remark 2.3.5. Hence, a similar reasoning to that from above for T_n^h (with $\bar{\beta}^h$ instead of $\bar{\beta}^{\hat{h}}$) leads to $T_n^h = \tilde{T}_n^h + o_p(1)$, so that

$$T_n^{\hat{h}} = \tilde{T}_n^h + o_p(1) = T_n^h + o_p(1).$$

□

2.8.7 Proof of Theorem 2.4.3

It will be started with the first assertion. $\hat{F}^{\hat{h}}_{Y|X}(y|x)$ was defined as

$$\hat{F}^{\hat{h}}_{Y|X}(y|x) = \frac{\hat{p}^{\hat{h}}(y,x)}{\hat{f}_X(x)} \quad \text{with} \quad \hat{p}^{\hat{h}}(y,x) = \frac{1}{n}\sum_{i=1}^n \mathcal{K}_{h_y}(y - \hat{h}(Y_i))\mathbf{K}_{h_x}(x - X_i).$$

Let $\mathcal{K} = [k_1, k_2]$ be compact and $\delta > 0$. One has

$$\sup_{y \in [k_1-\delta, k_2+\delta]} |\hat{h}(y) - h(y)| = o_p(1).$$

Let $\delta_n \searrow 0$ be the monotonic sequence from Lemma 1.5.1 with

$$\sup_{y \in [k_1-\delta, k_2+\delta]} |\hat{h}(y) - h(y)| = o_p(\delta_n).$$

Then, the results of Hansen (2008) can be adjusted as later in (4.6.10) to obtain

$$\begin{aligned}
&\sup_{x \in \mathrm{supp}(v), y \in \mathcal{K}} |\hat{p}^{\hat{h}}(y,x) - \hat{p}^h(y,x)| \\
&= \sup_{x \in \mathrm{supp}(v), y \in \mathcal{K}} \left| \frac{1}{n} \sum_{i=1}^n \mathcal{K}_{h_y}(y - h(Y_i) + h(Y_i) - \hat{h}(Y_i)) - \mathcal{K}_{h_y}(y - h(Y_i))\mathbf{K}_{h_x}(x - X_i) \right| \\
&\leq \sup_{x \in \mathrm{supp}(v), y \in \mathcal{K}} \frac{1}{n} \sum_{i=1}^n \left| \mathcal{K}_{h_y}(y - h(Y_i) + h(Y_i) - \hat{h}(Y_i)) - \mathcal{K}_{h_y}(y - h(Y_i)) \right| |\mathbf{K}_{h_x}(x - X_i)| \\
&\leq \sup_{x \in \mathrm{supp}(v), y \in \mathcal{K}} \frac{1}{n} \sum_{i=1}^n \int_{\frac{y-\delta_n-h(Y_i)}{h_y}}^{\frac{y+\delta_n-h(Y_i)}{h_y}} |K(u)| \, du \, |\mathbf{K}_{h_x}(x - X_i)| + o_p(1) \\
&= \sup_{x \in \mathrm{supp}(v), y \in \mathcal{K}} E\left[\int_{\frac{y-\delta_n-h(Y_1)}{h_y}}^{\frac{y+\delta_n-h(Y_1)}{h_y}} |K(u)| \, du \, |\mathbf{K}_{h_x}(x - X_1)| \right] + o_p(1) \\
&= \sup_{x \in \mathrm{supp}(v), y \in \mathcal{K}} \int\int\int_{\frac{y-\delta_n-g_{\beta_0}(w)-c_0-e}{h_y}}^{\frac{y+\delta_n-g_{\beta_0}(w)-c_0-e}{h_y}} |K(u)| \, du \, |\mathbf{K}_{h_x}(x - w)| f_X(w) f_\varepsilon(e) \, de \, dw + o_p(1) \\
&= \sup_{x \in \mathrm{supp}(v), y \in \mathcal{K}} \int\int\int_{y-\delta_n-g_{\beta_0}(w)-c_0-h_y u}^{y+\delta_n-g_{\beta_0}(w)-c_0-h_y u} f_\varepsilon(e) \, de \, |K(u)| \, du \, |\mathbf{K}_{h_x}(x - w)| f_X(w) \, dw \\
&\quad + o_p(1) \\
&= \sup_{x \in \mathrm{supp}(v), y \in \mathcal{K}} \int\int \big(F_\varepsilon(y + \delta_n - g_{\beta_0}(x - h_x w) - c_0 - h_y u) \\
&\qquad - F_\varepsilon(y - \delta_n - g_{\beta_0}(x - h_x w) - c_0 - h_y u) \big) |K(u)| \, du \, |\mathbf{K}(w)| f_X(x - h_x w) \, dw + o_p(1) \\
&= o_p(1)
\end{aligned}$$

and thus

$$\sup_{x \in \mathrm{supp}(v), y \in \mathcal{K}} \left| \hat{F}^{\hat{h}}_{Y|X}(y|x) - \hat{F}^h_{Y|X}(y|x) \right| = o_p(1).$$

The rest of the proof of the first assertion was already given in the proof of Theorem 2.3.6. With the reasoning from above the proof of the second part directly follows from Theorem 2.4.1 and Slutsky's theorem. □

2.8.8 Proof of Theorem 2.5.1

Since the support of v is compact, the results of Hansen (2008) yield

$$\hat{f}_X(x) - f_X(x) = \mathcal{O}_p\left(\sqrt{\frac{\log(n)}{n h_x^{d_X}}} \right)$$

uniformly in $x \in \operatorname{supp}(v)$. Due to $f_X(x) > 0$ for all $x \in \operatorname{supp}(v)$, this leads to

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\frac{v(X_i)}{\hat f_X(X_i)^2} &= \frac{1}{n}\sum_{i=1}^{n}\frac{v(X_i)}{f_X(X_i)^2} + \mathcal{O}_p\bigg(\sqrt{\frac{\log(n)}{nh_x^{d_X}}}\bigg)\\ &= E\bigg[\frac{v(X_1)}{f_X(X_1)^2}\bigg] + \mathcal{O}_p\bigg(\sqrt{\frac{\log(n)}{nh_x^{d_X}}}\bigg)\\ &= \int \frac{v(w)}{f_X(w)}\,dw + o\big(h_x^{\frac{d_X}{2}}\big)\end{aligned}$$

and analogously to (2.33) and $\frac{1}{n}\sum_{i=1}^{n} v(X_i) = \int v(w) f_X(w)\,dw + o\big(h_x^{\frac{d_X}{2}}\big)$. The numerator in (2.35) can be treated similarly to the proof of Lemma 2.8.1. To this end, recall

$$(F_{Y|X}^h)^{-1}\Big(\frac{1}{2}|X_i\Big) - h(Y_i) = F_\varepsilon^{-1}\Big(\frac{1}{2}\Big) - \varepsilon_i.$$

The results of Hansen (2008) imply

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} v(X_i)K_{h_\varepsilon}\bigg((F_{Y|X}^h)^{-1}\Big(\frac{1}{2}|X_i\Big) - h(Y_i)\bigg) &= \frac{1}{n}\sum_{i=1}^{n} v(X_i)K_{h_\varepsilon}\bigg(F_\varepsilon^{-1}\Big(\frac{1}{2}\Big) - \varepsilon_i\bigg)\\ &= f_\varepsilon\bigg(F_\varepsilon^{-1}\Big(\frac{1}{2}\Big)\bigg)\int v(w)f_X(w)\,dw + o_p\big(h_x^{\frac{d_X}{2}}\big),\end{aligned}$$

that is, is suffices to show

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n} v(X_i)K_{h_\varepsilon}\bigg((\hat F_{Y|X}^{\hat h})^{-1}\Big(\frac{1}{2}|X_i\Big) - \hat h(Y_i)\bigg)\\ &= \frac{1}{n}\sum_{i=1}^{n} v(X_i)K_{h_\varepsilon}\bigg((F_{Y|X}^h)^{-1}\Big(\frac{1}{2}|X_i\Big) - h(Y_i)\bigg) + o_p\big(h_x^{\frac{d_X}{2}}\big).\end{aligned}$$

Since the set $\Big\{(\hat F_{Y|X}^h)^{-1}\Big(\frac{1}{2}|x\Big) : x \in \operatorname{supp}(v)\Big\}$ is bounded, there exists a compact set $\mathcal{K}$, such that

$$\begin{aligned}P\Bigg(&\frac{1}{n}\sum_{i=1}^{n} v(X_i)K_{h_\varepsilon}\bigg((\hat F_{Y|X}^{\hat h})^{-1}\Big(\frac{1}{2}|X_i\Big) - \hat h(Y_i)\bigg)\\ &= \frac{1}{n}\sum_{i=1}^{n} v(X_i)I_{\{Y_i\in\mathcal{K}\}}K_{h_\varepsilon}\bigg((\hat F_{Y|X}^{\hat h})^{-1}\Big(\frac{1}{2}|X_i\Big) - \hat h(Y_i)\bigg)\Bigg) \to 1\end{aligned}$$

and

$$\begin{aligned}P\Bigg(&\frac{1}{n}\sum_{i=1}^{n} v(X_i)K_{h_\varepsilon}\bigg((F_{Y|X}^{h})^{-1}\Big(\frac{1}{2}|X_i\Big) - h(Y_i)\bigg)\\ &= \frac{1}{n}\sum_{i=1}^{n} v(X_i)I_{\{Y_i\in\mathcal{K}\}}K_{h_\varepsilon}\bigg((F_{Y|X}^{h})^{-1}\Big(\frac{1}{2}|X_i\Big) - h(Y_i)\bigg)\Bigg) \to 1.\end{aligned}$$

Hence, one has for some appropriate $C > 0, y_i^* \in \mathbb{R}$

$$\bigg|\frac{1}{n}\sum_{i=1}^{n} v(X_i)K_{h_\varepsilon}\bigg((\hat F_{Y|X}^{\hat h})^{-1}\Big(\frac{1}{2}|X_i\Big) - \hat h(Y_i)\bigg) - \frac{1}{n}\sum_{i=1}^{n} v(X_i)K_{h_\varepsilon}\bigg((F_{Y|X}^{h})^{-1}\Big(\frac{1}{2}|X_i\Big) - h(Y_i)\bigg)\bigg|$$

$$
\begin{aligned}
&= \Big| \frac{1}{n} \sum_{i=1}^{n} v(X_i) I_{\{Y_i \in \mathcal{K}\}} \bigg(K_{h_\varepsilon}\Big((\hat{F}^{\hat{h}}_{Y|X})^{-1}\Big(\frac{1}{2}|X_i\Big) - \hat{h}(Y_i)\Big) - K_{h_\varepsilon}\Big((F^{h}_{Y|X})^{-1}\Big(\frac{1}{2}|X_i\Big) - h(Y_i)\Big)\bigg)\Big| \\
&\quad + o_p\big(h_x^{\frac{d_X}{2}}\big) \\
&\leq \Big| \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{r-1} \frac{v(X_i) I_{\{Y_i \in \mathcal{K}\}}}{j!} \frac{\partial^j}{\partial y^j} K_{h_\varepsilon}(y)\Big|_{y=(F^h_{Y|X})^{-1}(\frac{1}{2}|X_i)-h(Y_i)} \\
&\qquad \bigg((\hat{F}^{\hat{h}}_{Y|X})^{-1}\Big(\frac{1}{2}|X_i\Big) - (\hat{F}^{h}_{Y|X})^{-1}\Big(\frac{1}{2}|X_i\Big) + h(Y_i) - \hat{h}(Y_i)\bigg)^j \Big| \\
&\quad + \Big| \frac{1}{n} \sum_{i=1}^{n} \frac{v(X_i) I_{\{Y_i \in \mathcal{K}\}}}{r!} \frac{\partial^r}{\partial y^r} K_{h_\varepsilon}(y)\Big|_{y=y_i^*} \\
&\qquad \bigg((\hat{F}^{\hat{h}}_{Y|X})^{-1}\Big(\frac{1}{2}|X_i\Big) - (F^{h}_{Y|X})^{-1}\Big(\frac{1}{2}|X_i\Big) + h(Y_i) - \hat{h}(Y_i)\bigg)^r \Big| + o_p\big(h_x^{\frac{d_X}{2}}\big) \\
&\leq \sum_{j=1}^{r-1} \frac{C}{h_\varepsilon^j} \bigg(\Big| \sup_{x \in \mathrm{supp}(v)} \Big| (\hat{F}^{\hat{h}}_{Y|X})^{-1}\Big(\frac{1}{2}|x\Big) - (F^{h}_{Y|X})^{-1}\Big(\frac{1}{2}|x\Big)\Big|^j + \sup_{y \in \mathcal{K}} |\hat{h}(y) - h(y)|^j \bigg) \\
&\qquad \frac{1}{n h_\varepsilon^j} \sum_{i=1}^{n} v(X_i) \Big| \frac{\partial^j}{\partial y^j} K(y)\Big|_{y=\frac{F_\varepsilon^{-1}(\frac{1}{2})-\varepsilon_i}{h_\varepsilon}} \Big| \\
&\quad + \frac{C}{h_\varepsilon^{r+1}} \bigg(\sup_{x \in \mathrm{supp}(v)} \Big| (\hat{F}^{\hat{h}}_{Y|X})^{-1}\Big(\frac{1}{2}|x\Big) - (F^{h}_{Y|X})^{-1}\Big(\frac{1}{2}|x\Big)\Big|^r + \sup_{y \in \mathcal{K}} |\hat{h}(y) - h(y)|^r \bigg) + o_p\big(h_x^{\frac{d_X}{2}}\big) \\
&= \mathcal{O}_p\Big(\big(n h_\varepsilon^4\big)^{-\frac{1}{4}} + \big(n^r h_\varepsilon^{4(r+1)}\big)^{-\frac{1}{4}} \Big) \\
&= o_p\big(h_x^{\frac{d_X}{2}}\big).
\end{aligned}
$$

□

3

Identification in a Fully Nonparametric Transformation Model with Heteroscedasticity

The underlying question of this Chapter can be formulated quite easily: Given some real valued random variable Y and some $\mathbb{R}^{d_X}$-valued random variable X fulfilling the heteroscedastic transformation model

$$h(Y) = g(X) + \sigma(X)\varepsilon \tag{3.1}$$

with some error term ε fulfilling $\varepsilon \perp X, E[\varepsilon] = 0$ and $\mathrm{Var}(\varepsilon) = 1$, are the model components $h : \mathbb{R} \to \mathbb{R},\ g : \mathbb{R}^{d_X} \to \mathbb{R}, \sigma : \mathbb{R}^{d_X} \to (0, \infty)$ and the error distribution uniquely determined if the joint distribution of (Y, X) is known? This uniqueness is called identification of a model. Already Box and Cox (1964), Bickel and Doksum (1981) and Zellner and Revankar (1969) introduced some parametric classes of transformation models. Horowitz (1996) proved for a linear regression function g and homoscedastic errors that the model is identified, when $h(y_0) = 0$ is assumed for some $y_0 \in \mathbb{R}$ and the regression parameter is standardized so that the first component, which is different from zero, is equal to one. Later, the ideas of Horowitz (1996) were extended by Ekeland et al. (2004) to general smooth regression functions g. The arguably most general identification results so far were provided by Chiappori et al. (2015) and Vanhems and Van Keilegom (2019), who considered general regression functions and homoscedastic errors as well, but allowed endogenous regressors. Linton et al. (2008) used similar ideas to obtain identifiability of a model with parametric transformation functions as a special case. Results allowing heteroscedasticity are rare. Zhou et al. (2009) showed identifiability in some kind of single-index model with a linear regression function g and a known variance function σ. Neumeyer et al. (2016) assumed identifiability implicitly by their assumption (a7).

In contrast to the approaches mentioned above, it is tried here to avoid any parametric assumption on h, g or σ, which to the author's knowledge has not been done in the literature so far. Note that the validity of the model is unaffected by linear transformations. This means that for arbitrary constants $a > 0, b \in \mathbb{R}$ equation (3.1) still holds when replacing h,

g and σ by

$$\tilde{h}(y) = ah(y) + b,$$

$$\tilde{g}(x) = ag(x) + b,$$

$$\tilde{\sigma}(x) = a\sigma(x).$$

Of course, one could have chosen an arbitrary $a \in \mathbb{R}$ as well, but as in Section 1.4 the transformation function h will be restricted to be strictly increasing. Nevertheless, at least two conditions for fixing a and b are needed. Referring to the fact that these conditions will determine the linear transformation they are sometimes called location and scale constraints.

This chapter is organized as follows. First, some differences to the homoscedastic case (that is, $\sigma \in \mathbb{R}$ is constant) are pointed out, before the main identification result for heteroscedastic transformation models as in (3.1) is presented. The chapter is completed by a brief discussion in 3.3. The proof of the main result is given in 3.5 and some additional remarks are postponed to 3.6.

3.1 Differences to the Homoscedastic Case

Many of the homoscedastic identification approaches mentioned above are based on the same idea (see Ekeland et al. (2004), Horowitz (2009) and recently Chiappori et al. (2015)). Using the example of Chiappori et al. (2015) their method can be summarized in the following way (see Section 1.4 for details): Let $F_{Y|X}$ be the conditional distribution function of Y conditioned on X. Take the derivatives of $F_{Y|X}$ with respect to y and x, divide the first by the latter one and obtain the transformation function by integrating this quotient. After applying some identification constraints the transformation function is identified as it only depends on the joint distribution of (Y, X).

In heteroscedastic models the reasoning has to be changed since the conditional distribution function and its partial derivatives can be written as

$$\begin{aligned} F_{Y|X}(y|x) &= P(Y \leq y|X = x) \\ &= P\Big(\varepsilon \leq \frac{h(Y) - g(X)}{\sigma(X)}\Big| X = x\Big) \\ &= F_\varepsilon\Big(\frac{h(y) - g(x)}{\sigma(x)}\Big), \end{aligned}$$

$$\frac{\partial F_{Y|X}(y|x)}{\partial y} = f_\varepsilon\Big(\frac{h(y) - g(x)}{\sigma(x)}\Big)\frac{h'(y)}{\sigma(x)} > 0 \tag{3.2}$$

and

$$\frac{\partial F_{Y|X}(y|x)}{\partial x_i} = -f_\varepsilon\Big(\frac{h(y) - g(x)}{\sigma(x)}\Big)\frac{\sigma(x)\frac{\partial g(x)}{\partial x_i} + (h(y) - g(x))\frac{\partial \sigma(x)}{\partial x_i}}{\sigma(x)^2}.$$

Here, $h'(y)$ is an abbreviation for the derivative $\frac{\partial}{\partial y}h(y)$ and $F_\varepsilon, f_\varepsilon$ denote the distribution function and density of ε. Hence, the transformation function can not be obtained by choosing an appropriate $i = 1, ..., d_X$ and simply integrating the quotient

$$\frac{\frac{\partial F_{Y|X}(y|x)}{\partial y}}{\frac{\partial F_{Y|X}(y|x)}{\partial x_i}} = -\frac{h'(y)\sigma(x)}{\sigma(x)\frac{\partial g(x)}{\partial x_i} + (h(y) - g(x))\frac{\partial \sigma(x)}{\partial x_i}}, \tag{3.3}$$

since the denominator now also depends on the transformation function. Although obvious, note that for a constant σ the same quotient as in the homoscedastic case of Section 1.4 is obtained.
Remark that equation (3.3) only holds if

$$f_\varepsilon\left(\frac{h(y) - g(x)}{\sigma(x)}\right) > 0.$$

Therefore, $f_\varepsilon(y) > 0$ is assumed for all $y \in \mathbb{R}$. This condition can be replaced by considering appropriate subsets $\mathcal{Y}$ of the support of Y such that one of the sets

$$\tilde{\mathcal{A}}_{\mathcal{Y},i} = \left\{x : \frac{\partial F_{Y|X}(y|x)}{\partial x_i} > 0 \text{ for all } y \in \mathcal{Y}\right\}, \quad i = 1, ..., d_X, \tag{3.4}$$

is not empty. But since the consequent results would be restricted on $\mathcal{Y}$ as well this will not be examined further. Chiappori et al. (2015) circumvent this problem by their assumption A6, but its validity is not obvious and in general it can not be fulfilled for bounded support of f_ε and $\mathcal{Y} = \mathbb{R}$. See Section 3.6.1 for a short discussion on bounded support of f_ε.

3.2 The Transformation Function as a Solution of a Differential Equation

From now on, consider without loss of generality the case $i = 1$ and define

$$\lambda(y|x) := \frac{\frac{\partial F_{Y|X}(y|x)}{\partial x_1}}{\frac{\partial F_{Y|X}(y|x)}{\partial y}} = -\frac{\sigma(x)\frac{\partial g(x)}{\partial x_1} + (h(y) - g(x))\frac{\partial \sigma(x)}{\partial x_1}}{h'(y)\sigma(x)} \tag{3.5}$$

and

$$\lambda(y) := \int v(x)\lambda(y|x)\,dx = -\int v(x)\frac{\sigma(x)\frac{\partial g(x)}{\partial x_1} + (h(y) - g(x))\frac{\partial \sigma(x)}{\partial x_1}}{h'(y)\sigma(x)}\,dx = -\frac{A + Bh(y)}{h'(y)}, \tag{3.6}$$

where v is an appropriate weighting function and A and B are defined as

$$A := \int v(x)\left(\frac{\sigma(x)\frac{\partial g(x)}{\partial x_1} - g(x)\frac{\partial \sigma(x)}{\partial x_1}}{\sigma(x)}\right)dx \quad \text{and} \quad B := \int v(x)\frac{\frac{\partial \sigma(x)}{\partial x_1}}{\sigma(x)}\,dx. \tag{3.7}$$

Note that λ is uniquely determined by the joint distribution of (Y, X). From an identification point of view it is not necessary to consider a weighted version of λ, but since this will be needed later in Section 4.2 it is already introduced here. For all functions h, g, σ, which fulfil model (3.1), consider a weight function v such that $B \neq 0$. Note that if $B = 0$ for all weight functions v, this implies homoscedasticity of the error and identifiability of the model can be shown as in Chiappori et al. (2015). Some further observations are summarized in the following remark.

Remark 3.2.1 *Assume $h'(y) > 0$ for all $y \in \mathbb{R}$. Then, it holds that:*

(i) λ is well defined on $\mathbb{R}$.

(ii) λ has at most one root, which is denoted by y_0. Equation (3.6) leads to $h(y_0) = -\frac{A}{B}$.

(iii) The sign of B is equal to the sign of λ at $(-\infty, y_0)$. Therefore, assume the sign of B to be known and assume w.l.o.g B to be positive.

(iv) In Section 4.1.1 estimators for λ and y_0 will be derived. These can be used to test whether or not there exists a root y_0 of λ.

(v) A can be written as

$$A = \int v(x)\sigma(x) \frac{\partial}{\partial x_1}\left(\frac{g(x)}{\sigma(x)}\right) dx.$$

(vi) Possibly, one has $B = 0$. As mentioned above, one can proceed analogously to Chiappori et al. (2015) in this case. It might be sensible to first apply a test for the null hypothesis $H_0 : B = 0$ (see Remark 4.1.2).

Assume existence of the root y_0 defined as in Remark 3.2.1 above and assume $B > 0$. See Section 3.6.1 for the case, in which there is no such root. Consider a compact set $[z_a, z_b] \subseteq \mathbb{R}$ with $y_0 < z_a$. The condition $y_0 < z_a$ can be replaced by $z_b < y_0$ as well. Let $h'(y) > 0$ for all $y \in \mathbb{R}$. Then, validity of equation (3.6) on (y_0, ∞) is equivalent to the differential equation

$$h'(y) = -\frac{A + Bh(y)}{\lambda(y)} \tag{3.8}$$

for all $y \in (y_0, \infty)$. Now, an appropriate initial condition

$$h(y_1) = \lambda_1 \tag{3.9}$$

for some $y_1 > y_0$ and some $\lambda_1 > -\frac{A}{B}$ (see Remark 3.2.1) together with Theorem 3.6.6 in Section 3.6 can be applied to obtain uniqueness of any existing solution h on some compact interval $[z_a, z_b] \subseteq (y_0, \infty)$. Basically, this is one part of the proof of Lemma 3.2.3 below. There, the initial condition can be seen as the previous scale constraint.
This result will yield that for fixed A and B the transformation function h is identified on every compact interval $[z_a, z_b]$ such that $y_0 < z_a$ by the differential equation (3.8) combined with an appropriate initial value. Thus, to prove uniqueness the remaining task consists in extending any solution to the whole set of real numbers on the one hand and to show that A and B as well as this extension are unique on the other hand. When doing so, the assumptions (**A4**) and (**A7**) from Section 3.4 will play a key role.

An analytic Expression of the Transformation Function

Again, let $[z_a, z_b] \subseteq (y_0, \infty)$ be compact. Note that $A + Bh(y) > 0$ for all $y > y_0$ by definition of y_0 and $B > 0$. When having a closer look on the definition of λ, one may notice that equation (3.6) can be written as

$$\frac{1}{\lambda(y)} = -\frac{h'(y)}{A + Bh(y)} = -\frac{1}{B}\left(\frac{\partial \log(|A + Bh(y)|)}{\partial y}\right) \quad \text{for all } y \neq y_0. \tag{3.10}$$

As a consequence, condition (3.9) leads to

$$\exp\left(- B\int_{y_1}^{y} \frac{1}{\lambda(u)}\, du\right) = \frac{A + Bh(y)}{A + B\lambda_1} \quad \text{for all } y \in [z_a, z_b]$$

and

$$h(y) = \frac{(A + B\lambda_1)\exp\left(- B\int_{y_1}^{y} \frac{1}{\lambda(u)}\, du\right) - A}{B} \quad \text{for all } y \in [z_a, z_b] \tag{3.11}$$

for an arbitrary, but fixed $\lambda_1 > -\frac{A}{B}$. So far, λ_1 has not been fixed and no specific location constraint has been considered yet. Later, it will be shown that B is already determined by the independence of ε and X, so that apart from the initial condition (3.9), any (location) constraint which fixes A would be sufficient. This is consistent with the beginning of this chapter where it was mentioned that there are at least two conditions necessary to identify the transformation function since the model (3.1) is invariant under (monotonically growing) linear transformations. In the following, it will be proven that as in the homoscedastic case there will be exactly two conditions needed to identify the model.
An elegant way to choose the location constraint is to require

$$h(y_0) = 0, \tag{3.12}$$

since this in turn results due to Remark 3.2.1 in $A = 0$. An obvious choice for fixing λ_1 may consist in $\lambda_1 = 1$, so that

$$h(y) = \exp\left(- B\int_{y_1}^{y} \frac{1}{\lambda(u)}\, du\right) \quad \text{for all } y \in [z_a, z_b]. \tag{3.13}$$

Although this choice seems to be quite convenient, in principle any condition that fixes A is conceivable.

Example 3.2.2 *Another way to determine the location and scale constraints is requiring*

$$h(y_1) = 0 \quad \text{and} \quad h'(y_1) = 1. \tag{3.14}$$

Due to equation (3.6), this leads to $A = -\lambda(y_1)$. One possibility to solve the corresponding differential equation consists in writing

$$H(y) := \int_{y_1}^{y} \frac{1}{\lambda(u)}\, du > -\infty \quad \text{for all } y > y_0$$

and rewriting (3.6) as

$$\underbrace{h'(y)\exp(BH(y)) + h(y)\frac{B}{\lambda(y)}\exp(BH(y))}_{=\frac{\partial}{\partial y}h(y)\exp(BH(y))} + \frac{A}{\lambda(y)}\exp(BH(y)) = 0.$$

Integration results in

$$h(y)\exp(BH(y)) - \underbrace{h(y_1)\exp(BH(y_1))}_{=0} = -A\int_{y_1}^{y} \frac{1}{\lambda(y)}\exp(BH(u))\, du$$

and finally

$$\begin{aligned}
h(y) &= -A\exp(-BH(y))\int_{H(y_1)}^{H(y)} \exp(Bu)\,du \\
&= -\frac{A}{B}\exp(-BH(y))(\exp(BH(y)) - \underbrace{\exp(BH(y_1))}_{=1}) \\
&= \frac{A}{B}(\exp(-BH(y)) - 1) \\
&= \frac{\lambda(y_1)(1-\exp(-BH(y)))}{B}.
\end{aligned}$$

For uniqueness reasons this approach should lead to the same solution as the previous one and indeed (3.11) leads to

$$\begin{aligned}
h(y) &= \frac{(A + Bh(y_1))\exp\Big(-B\int_{y_1}^{y}\frac{1}{\lambda(u)}\,du\Big) - A}{B} \\
&= \frac{A}{B}(\exp(-BH(y)) - 1).
\end{aligned}$$

These preliminary thoughts are formalized in the following lemma.

Lemma 3.2.3 *Assume (**A1**)–(**A6**) from Section 3.4. Further, require condition (3.9) and let $y_2 < y_0 < y_1$ as well as $B \neq 0$.*

1. *For each $A \in \mathbb{R}$ such that $\lambda_1 > -\frac{A}{B}$, the unique solution to (3.8) on (y_0, ∞) is given by (3.11). It can be extended to a global unique solution to (3.8) by*

$$h(y) = \begin{cases} \frac{(A+B\lambda_1)\exp\left(-B\int_{y_1}^{y}\frac{1}{\lambda(u)}\,du\right) - A}{B} & y > y_0 \\ -\frac{A}{B} & y = y_0 \\ \frac{(A+B\lambda_2)\exp\left(-B\int_{y_2}^{y}\frac{1}{\lambda(u)}\,du\right) - A}{B} & y < y_0 \end{cases}, \tag{3.15}$$

where λ_2 is uniquely determined by requiring $\lim\limits_{y\searrow y_0} h'(y) = \lim\limits_{y\nearrow y_0} h'(y) = h'(y_0)$ as

$$\lambda_2 = -\frac{\lim\limits_{t\to 0}(A + B\lambda_1)\exp\left(B\Big(\int_{y_2}^{y_0-t}\frac{1}{\lambda(u)}\,du - \int_{y_1}^{y_0+t}\frac{1}{\lambda(u)}\,du\Big)\right) + A}{B}. \tag{3.16}$$

2. *If additionally (3.12) and $\lambda_1 = 1$ hold, one has*

$$h(y) = \begin{cases} \exp\Big(-B\int_{y_1}^{y}\frac{1}{\lambda(u)}\,du\Big) & y > y_0 \\ 0 & y = y_0 \\ \lambda_2\exp\Big(-B\int_{y_2}^{y}\frac{1}{\lambda(u)}\,du\Big) & y < y_0 \end{cases}, \tag{3.17}$$

where λ_2 is uniquely determined by requiring $\lim\limits_{y\searrow y_0} h'(y) = \lim\limits_{y\nearrow y_0} h'(y) = h'(y_0)$ as

$$\lambda_2 = -\lim_{t\to 0}\exp\left(B\Big(\int_{y_2}^{y_0-t}\frac{1}{\lambda(u)}\,du - \int_{y_1}^{y_0+t}\frac{1}{\lambda(u)}\,du\Big)\right). \tag{3.18}$$

The proof can be found in Section 3.5.
Given any fixed B, the uniqueness of A has not been treated so far apart from the fact that (3.12) or (3.14) imply $A = 0$ or $A = -\lambda(y_1)$, respectively. Nevertheless, this can be used to argue that any conditions (probably with identification constants different from zero or one) of type (3.9) and (3.12) or type (3.14) determine A as well, since all solutions can be linearly transformed to the already identified case. Therefore, only the identification constraints (3.9) and (3.12) for $\lambda_1 = 1$ are considered from now on so that $A = 0$.
Up to now, uniqueness of a solution for any fixed B was shown. In the last part of this section, the uniqueness of B is discussed in order to derive an identification result from the previous arguments. Assume the existence of two continuously differentiable solutions $h, \tilde{h}$ to the differential equation (3.8) as in (3.15) with corresponding parameters (B, λ_2) and $(\tilde{B}, \tilde{\lambda}_2)$ such that $B, \tilde{B} > 0$. Assume $B \neq \tilde{B}$ and without loss of generality assume $\tilde{B} > B$. Then, equation (3.15) leads to

$$\tilde{h}(y) = \frac{(A + \tilde{B}\lambda_1)\big(\frac{Bh(y)+A}{A+B\lambda_1}\big)^{\frac{\tilde{B}}{B}} - A}{\tilde{B}} \quad \text{for all } y > y_0$$

and consequently

$$\tilde{h}'(y) = \frac{(A + \tilde{B}\lambda_1)B}{(A + B\lambda_1)\tilde{B}} \left(\frac{Bh(y) + A}{A + B\lambda_1}\right)^{\frac{\tilde{B}}{B}-1} h'(y)$$

for all $y > y_0$. Continuous differentiability of h and $h'(y_0) \in (0, \infty)$ imply $\tilde{h}'(y_0) = 0$, which is a contradiction to assumption (**A4**) from Section 3.4. Hence, the assumption $B \neq \tilde{B}$ has to be rejected and one has $h = \tilde{h}$. Finally, the main identification result of this section directly follows from Lemma 3.2.3.

Theorem 3.2.4 *Assume (**A1**)–(**A6**) from Section 3.4 as well as conditions (3.9) for $\lambda_1 = 1$ and (3.12). Further, let $y_2 < y_0 < y_1$ as well as $B \neq 0$. Then, the unique solution of the model equation (3.1) is given by*

$$h(y) = \begin{cases} \exp\Big(- B \int_{y_1}^{y} \frac{1}{\lambda(u)}\, du\Big) & y > y_0 \\ 0 & y = y_0 \\ \lambda_2 \exp\Big(- B \int_{y_2}^{y} \frac{1}{\lambda(u)}\, du\Big) & y < y_0 \end{cases},$$

where λ_2 is uniquely determined as (3.18) via $\lim_{y \searrow y_0} h'(y) = \lim_{y \nearrow y_0} h'(y) = h'(y_0) > 0$. Further, B is uniquely determined and one has

$$g(x) = E[h(Y)|X = x] \quad \text{and} \quad \sigma(x) = \sqrt{\mathrm{Var}(h(Y)|X = x)}.$$

So far, uniqueness of B was proven by using $h'(y_0) > 0$. In Section 3.6.3, it is tried to relax this assumption. Assumption (**A7**) from Section 3.4 will play a major role there instead. Now, an example is given in which neither (**A4**) nor (**A7**) are fulfilled and the model there is not identified.

Example 3.2.5 *(see also Remark 3.4.1 below). Consider the model*

$$Y = X + X\varepsilon,$$

that is $g(x) = \sigma(x) = x$*, for some random variables* $X \perp \varepsilon$ *with* $E[\varepsilon] = 0$ *and* $E[\varepsilon^6] < \infty$*. This model does not fulfil assumption* (***A7***) *or* $h'(y) > 0$ *for all* $y \in \mathbb{R}$*. Indeed, the transformation function is no longer identified since for example*

$$\begin{aligned} Y^3 &= X^3(1+\varepsilon)^3 \\ &= X^3 E[(1+\varepsilon)^3] + X^3\big((1+\varepsilon)^3 - E[(1+\varepsilon)^3]\big) \\ &= \tilde{g}(X) + \tilde{\sigma}(X)\tilde{\varepsilon} \end{aligned}$$

for $\tilde{g}(x) = x^3 E[(1+\varepsilon)^3], \tilde{\sigma}(x) = x^3\sqrt{\mathrm{Var}((1+\varepsilon)^3)}$ *and* $\tilde{\varepsilon} = \frac{(1+\varepsilon)^3 - E[(1+\varepsilon)^3]}{\sqrt{\mathrm{Var}((1+\varepsilon)^3)}}$.

Under the additional assumption that h is twice continuously differentiable, an analytic expression for B can be obtained. Equation (3.6) yields

$$\lambda(y) = -\frac{Bh(y)}{h'(y)},$$

so that the derivative of λ can be written as

$$\frac{\partial}{\partial y}\lambda(y) = -B\frac{h'(y)^2 - h''(y)h(y)}{h'(y)^2}$$

with $h''(y) = \frac{\partial^2}{\partial y^2}h(y)$. Applying (3.12), that is $h(y_0) = 0$, results in

$$\frac{\partial}{\partial y}\lambda(y)\Big|_{y=y_0} = -B. \tag{3.19}$$

Equation (3.19) will be important in Section 4.1.2, since it yields an analytic expression for B, which can be used to construct a plug in estimator. This section is completed by an additional remark.

Remark 3.2.6 *Some possible extensions are listed in the following. It is conjectured, that they can be shown with slightly more effort.*

1. *The identification result should hold for any other combination of location and scale constraints as long as they fix* A *and* λ_1 *in equation (3.11).*

2. *It should be possible to extend Theorem 3.2.4 to all monotonic functions* h *with* $\lim_{y\to-\infty} h(y) = -\infty$ *and* $\lim_{y\to\infty} h(y) = \infty$ *although one has to change argumentation (see corollary 3.6.4 below). For example, uniqueness of* λ_2 *possibly can be shown in a similar way to the proof of Lemma 3.6.3 below.*

3. *There might be cases when* h *can be defined as shown here even when the model is not fulfilled. The author does not know yet if applying the resulting transformation could still be advisable in terms of simplifying the model or making other procedures better applicable.*

4. *It is conjectured that an extension to models with endogenous regressors can be deduced similarly to Chiappori et al. (2015) and Vanhems and Van Keilegom (2019).*

3.3 Discussion

The so far most general identification result in the theory of transformation models has been provided. While doing so, the techniques of Ekeland et al. (2004) and Chiappori et al. (2015) have been used to reduce the problem of identifiability to that of solving an ordinary differential equation. Most of the previous results are contained as special cases. The main contribution consists in allowing heteroscedastic errors, which justifies the common practice to assume identifiability like for example in the paper of Neumeyer et al. (2016).
Moreover, the result is constructive in the sense that it does not only guarantee identification of the model, but even supplies an analytic expression of the transformation function depending on the joint distribution function of the data and some parameter B. This parameter is identified, too, and can be expressed as in (3.19) under the additional assumption of a twice continuously differentiable transformation function.
Future research could consist in successively generalizing the result as has been suggested in Remark 3.2.6. Moreover, it would be desirable to develop conditions on the joint distribution function of (Y, X) under which model (3.1) is fulfilled. In contrast to the thoughts on identifiability here, such a question addresses the solvability of (3.1), that is, the issue of existence of a solution instead of uniqueness.

3.4 Assumptions

In the following, assumptions needed for the presented results are listed.

(**A1**) Let Y, ε and X be real valued and $\mathbb{R}^{d_X}$-valued random variables, respectively, with

$$h(Y) = g(X) + \sigma(X)\varepsilon.$$

Let h be continuously differentiable.

(**A2**) ε is a centred random variable independent of X with $E[\varepsilon] = 0$ and $\mathrm{Var}(\varepsilon) = 1$.

(**A3**) The density f_ε is continuous with $f_\varepsilon(y) > 0$ for all $y \in \mathbb{R}$.

(**A4**) The conditional distribution function $F_{Y|X}(y|x)$ is continuously differentiable with respect to y and x with $\frac{\partial}{\partial y} F_{Y|X}(y|x) > 0$ for all $y \in \mathbb{R}, x \in \mathbb{R}^{d_X}$.

(**A5**) g and σ are continuously differentiable and $\sigma(x) > 0$ for all $x \in \mathbb{R}^{d_X}$.

(**A6**) v is a weighting function such that A, B and λ are well defined by equations (3.6) and (3.7) and such that $B \neq 0$.

(**A7**) Define for $c > 0$ the sets $M_{>c} := \{x : g(x) > c\}$ and $M_{<-c} := \{x : g(x) < -c\}$. There is some $c > 0$ such that $P(X \in M_{>c}) > 0$ (or $P(X \in M_{<-c}) > 0$) holds and $x \mapsto \frac{\sigma(x)}{g(x)}$ is not constant on $M_{>c}$ (or M_{-c}).

Remark 3.4.1 *1. Let*

$$\begin{aligned} \varphi_\alpha : \quad \mathbb{R} &\to \quad \mathbb{R} \\ y &\mapsto \quad \mathrm{sign}(y)|y|^\alpha. \end{aligned}$$

*Assumption (**A7**) is needed to avoid cases like $g(x) = g\sigma(x)$ for some $g \in \mathbb{R}$, since in this case for arbitrary $\alpha > 0$ one has*

$$\begin{aligned}\varphi_\alpha(g(X) + \sigma(X)\varepsilon) &= \sigma(X)^\alpha \varphi_\alpha(g + \varepsilon)\\ &= \underbrace{\sigma(X)^\alpha E[\varphi_\alpha(g+\varepsilon)]}_{=\tilde{g}(X)}\\ &\quad + \underbrace{\sigma(X)^\alpha \sqrt{\mathrm{Var}(\varphi_\alpha(g+\varepsilon))}}_{=\tilde{\sigma}(X)} \underbrace{\frac{\varphi_\alpha(g+\varepsilon) - E[\varphi_\alpha(g+\varepsilon)]}{\sqrt{\mathrm{Var}(\varphi_\alpha(g+\varepsilon))}}}_{=\tilde{\varepsilon}}\\ &= \tilde{g}(X) + \tilde{\sigma}(X)\tilde{\varepsilon},\end{aligned}$$

so that by definition (3.7)

$$B = B(\alpha) = \int v(x) \frac{\frac{\partial \tilde{\sigma}(x)}{\partial x_1}}{\tilde{\sigma}(x)}\, dx = \alpha B(1)$$

could attain any value, such that the second moment of $\varphi_\alpha(g + \varepsilon)$ exists.

2. *Due to equation (3.2), assumption (**A4**) ensures $h'(y) > 0$ for all $y \in \mathbb{R}$.*

3.5 Proof of Lemma 3.2.3

First, consider a compact interval $\mathcal{K} = [k_1, k_2] \subseteq (y_0, \infty)$ and recall equation (3.11). Assumption (**A4**) ensures $h' > 0$. First, it is shown that h as defined in (3.11) is the unique solution to (3.8) on $[k_1, k_2]$. For the moment assume $k_1 = y_1$ and define

$$G = [y_1, k_2] \times [\lambda_1, \infty) \quad \text{and} \quad D : G \to \mathbb{R},\ D(y, h) = -\frac{A + Bh}{\lambda(y)}.$$

With $a = y_1, b = k_2$ and $\theta_0 = \lambda_1$, Theorem 3.6.6 below ensures uniqueness of the solution of

$$h'(y) = -\frac{A + Bh(y)}{\lambda(y)} \quad \text{for all } y \in [y_1, k_2].$$

In Section 3.2 it was shown that (3.11) indeed is a solution of (3.8). Since $f_\varepsilon(y) > 0$ for all $y \in \mathbb{R}$, this solution holds for arbitrarily large k_2. Hence, by letting k_2 tend to infinity uniqueness of h on $[y_1, \infty)$ is obtained.

Now, consider an arbitrary value $\tilde{y} \in (y_0, y_1)$. Then, with the previous initial condition replaced by

$$\tilde{h}(\tilde{y}) = \tilde{\lambda}$$

for some $\tilde{\lambda} > 0$ and the same arguments as before, the differential equation

$$h'(y) = -\frac{A + Bh(y)}{\lambda(y)} \quad \text{for all } y \in [\tilde{y}, k_2]$$

is uniquely solved by

$$\tilde{h}(y) = \frac{(A + B\tilde{\lambda}) \exp\Big(-B \int_{\tilde{y}}^{y} \frac{1}{\lambda(u)}\, du\Big) - A}{B}$$

$$= \frac{\frac{(A+B\tilde{\lambda})}{(A+B\lambda_1)} \exp\Big(- B \int_{\tilde{y}}^{y_1} \frac{1}{\lambda(u)}\, du \Big)(A + Bh(y)) - A}{B}$$

for all $y \in [\tilde{y}, \infty)$, where the last equation follows from (3.11). To fulfil the previous scale constraint $\tilde{h}(y_1) = \lambda_1$ it is required that

$$\tilde{\lambda} = \frac{(A + \lambda_1 B) \exp\Big(B \int_{\tilde{y}}^{y_1} \frac{1}{\lambda(u)}\, du \Big) - A}{B}$$

Since this in turn results in $\tilde{h}(y) = h(y)$ for all $y \in [\tilde{y}, \infty)$, h is identified for all $y \in [\tilde{y}, \infty)$. Choosing $\tilde{y}$ arbitrarily close to y_0 results in

$$h(y) = \frac{(A + B\lambda_1) \exp\Big(- B \int_{y_1}^{y} \frac{1}{\lambda(u)}\, du \Big) - A}{B} \qquad \text{for all } y > y_0.$$

When proceeding analogously for $y < y_0$ with the initial condition

$$h(y_2) = \lambda'$$

for some $\lambda' < -\frac{A}{B}$, one has

$$h(y) = \frac{(A + B\lambda') \exp\Big(- B \int_{y_2}^{y} \frac{1}{\lambda(u)}\, du \Big) - A}{B} \qquad \text{for all } y < y_0.$$

Recall $h'(y) > 0$ for all $y \in \mathbb{R}$ and let $t > 0$. Due to the continuous differentiability of h in y_0 one has

$$\lim_{t \to 0} \frac{\frac{h(y_0 + t) - h(y_0)}{t}}{\frac{h(y_0 - t) - h(y_0)}{-t}} \to 1.$$

On the other hand, it holds that

$$\begin{aligned} \frac{\frac{h(y_0 + t) - h(y_0)}{t}}{\frac{h(y_0 - t) - h(y_0)}{-t}} &= - \frac{(A + B\lambda_1) \exp\Big(- B \int_{y_1}^{y_0+t} \frac{1}{\lambda(u)}\, du \Big)}{(A + B\lambda') \exp\Big(- B \int_{y_2}^{y_0-t} \frac{1}{\lambda(u)}\, du \Big)} \\ &= - \frac{(A + B\lambda_1)}{(A + B\lambda')} \exp\bigg(B \Big(\int_{y_2}^{y_0-t} \frac{1}{\lambda(u)}\, du - \int_{y_1}^{y_0+t} \frac{1}{\lambda(u)}\, du \Big) \bigg), \end{aligned}$$

so that

$$\lambda' = - \frac{\lim_{t \to 0} (A + B\lambda_1) \exp\bigg(B \Big(\int_{y_2}^{y_0-t} \frac{1}{\lambda(u)}\, du - \int_{y_1}^{y_0+t} \frac{1}{\lambda(u)}\, du \Big) \bigg) + A}{B} = \lambda_2.$$

This leads to the uniqueness of Solution (3.15). Inserting $A = 0$ yields the second part of the assertion. □

3.6 Miscellaneous

In this section all results, proofs or remarks are collected, that are not directly necessary to follow the golden thread of the chapter, but nevertheless important and interesting additions that complete a comprehensive discussion of the topic.

3.6.1 Bounded Support of f_ε

In the previous discussion, the density f_ε of the error term was always assumed to be greater than zero on the whole set of real numbers. In this Subsection it is tried to relax this assumption and to show how the derived approaches can be adapted to this case. It is assumed that $h' > 0$ although it is conjectured that the ideas here can be extended to general monotonic transformations h by similar ideas as in Section 3.6.3 below.
Identifiability of the model (3.1) on compact intervals is considered first. Hence, let $\mathcal{Y} \subseteq \mathbb{R}$ be a compact interval. The main drawback when allowing one-sided or even bounded support of f_ε consists in the fact that in general the set

$$\mathcal{A}_{\mathcal{Y}} = \left\{ x : \frac{\partial F_{Y|X}(y|x)}{\partial y} > 0 \text{ for all } y \in \mathcal{Y} \right\},$$

which is defined similar to equation (3.4), no longer consists of every $x \in \mathbb{R}^{d_X}$.

Example 3.6.1 *Consider the following model*

$$Y = X + 1 + \varepsilon$$

with one-sided error $\varepsilon = \eta - 1$ *and* $\eta \sim \mathrm{Exp}(1)$. *For the choice of* $\mathcal{Y} = [0,1]$ *one has* $\mathcal{A}_{\mathcal{Y}} = (-\infty, 0)$.

Chiappori et al. (2015) introduced an assumption similar to $\mathcal{A}_{\mathcal{Y}} \neq \emptyset$, where $\mathcal{Y}$ is equal to the support of Y. Although they considered the partial derivative with respect to an appropriate x_i instead of y, the underlying problem remains the same. As in example 3.6.1, the weighting with respect to x has to be restricted to $\mathcal{A}_{\mathcal{Y}}$. Although not a big problem from an identification perspective this issue becomes crucial when estimating since neither g nor $\mathcal{A}_{\mathcal{Y}}$ are known a priori. From an identification point of view, this “weighting” can even be implemented using a dirac measure.
The argumentation becomes even more complicated when considering an error term with bounded (from both sides) support or a heteroscedastic model, but at least the following corollary can be stated.

Corollary 3.6.2 *Assume the support* $\mathcal{Y}$ *of* Y *is an interval and can be partitioned into countably many bounded subintervals* $(\mathcal{Y}_n)_{n\in\mathbb{N}}$ *such that*

$$\mathcal{A}_{\mathcal{Y}_n} \neq \emptyset \quad \textit{and} \quad \max\{y : y \in \mathcal{Y}_n\} = \min\{y : y \in \mathcal{Y}_{n+1}\} \quad \textit{for all } n \in \mathbb{N}.$$

Then, the transformation function h *from model (3.1) is identified on* $\mathcal{Y}$ *via* $h(y_1) = \lambda_1$ *and* $h(y_2) = \lambda_2$ *for arbitrary* $y_1, y_2 \in \mathcal{Y}, \lambda_1 < \lambda_2$.

Proof: Assume existence of a root y_0 of λ and let $y_0 \in \mathcal{Y}_m$ for some $m \in \mathbb{N}$. If there does not exist such a root, one can proceed as later in Subsection 3.6.2. The basic reasoning in this case would be the same, so that only the first case with a root is considered in the remaining proof.
Let $y_{1,m}$ and $y_{2,m}$ be the lower and upper bound of the interval $\mathcal{Y}_m$. For any $\lambda_{1,m} < \lambda_{2,m}$, identification of h on $\mathcal{Y}_m$ via $h(y_{1,m}) = \lambda_{1,m}$ and $h(y_{2,m}) = \lambda_{2,m}$ can be shown as before in

Section 3.2.
Now look at the two consecutive intervals $\mathcal{Y}_m$ and $\mathcal{Y}_{m+1}$, that is $y_{2,m} = y_{1,m+1}$. Since h is continuously differentiable in $y_{2,m}$, the limits

$$\lim_{y \to y_{2,m}} h(y) \quad \text{and} \quad \lim_{y \to y_{2,m}} h'(y)$$

exist and especially have to be independent of any sequence $y_l \overset{l\to\infty}{\longrightarrow} y_{2,m}$. Therefore, the location and scale constraints for $\mathcal{Y}_{m+1}$ are determined by the continuous differentiability of h. Hence, h is identified on $\mathcal{Y}_{m+1}$ as well. After proceeding analogously for all previous and consecutive intervals, one obtains a version of h which only depends on the chosen values of $\lambda_{1,m}$ and $\lambda_{2,m}$ from the first step. Due to the continuous differentiability of h, this is even the case if there are accumulation points in the sequence $(y_{2,n})_{n\in\mathbb{N}}$. The two constants $\lambda_{1,m}$ and $\lambda_{2,m}$ are directly linked to the global location and scale constraints of h and are uniquely determined by $h(y_1) = \lambda_1$ and $h(y_2) = \lambda_2$.
□

If one has chosen a set $\mathcal{Y}$ and an appropriate weighting function v once, the conditions necessary to ensure identifiability of the model are thus the same as in the case with unbounded support.

3.6.2 The Case without a Root y_0

Again, the notations from Section 3.2 are used, that is, y_0 is defined as the root of the map $y \mapsto A + Bh(y)$ with A and B from (3.7). Recall the model (3.1). B is assumed to be different from zero, so that the only possibility that there does not exist any root y_0 of λ is the case where the image of h and consequently the support of f_ε is bounded from at least one side. Therefore, this is a special case of Subsection 3.6.1, but has not been treated in detail there.
Identifiability of the model (3.1) on compact intervals is considered in the following. Hence, let $\mathcal{Y}$ be an interval and let v be weighting function, such that $\operatorname{supp}(v) \subseteq \mathcal{A}_\mathcal{Y} \neq \emptyset$ and $h(y) \neq 0$ for all $y \in \mathcal{Y}$. Introduce the location and scale constraints

$$h(y_1) = \lambda_1 \quad \text{and} \quad h(y_2) = \lambda_2$$

for arbitrary values $y_1 < y_2 \in \mathcal{Y}$ and $\lambda_1 < \lambda_2 \in \mathbb{R}$. Then, the corresponding solution to (3.8) on $\mathcal{Y}$ is given by

$$h(y) = (\lambda_2 - \lambda_1)\frac{\exp\Big(-B\int_{y_1}^{y}\frac{1}{\lambda(u)}\,du\Big) - 1}{\exp\Big(-B\int_{y_1}^{y_2}\frac{1}{\lambda(u)}\,du\Big) - 1} + \lambda_1, \quad y \in \mathcal{Y}.$$

Indeed, the transformation function expressed in this way fulfils the differential equation as well as the boundary constraints. Uniqueness of h can be shown as in Section 3.2. Since $h(y_1) < h(y_2)$ and the map $y \mapsto \exp\big(-B\int_{y_1}^{y}\frac{1}{\lambda(u)}\,du\big)$ is monotone, h is increasing as required. Estimating the transformation would become insofar easier later as there is no longer the need to estimate y_0 and λ_2 as in Chapter 4.

3.6.3 Vanishing Derivatives of h

Up to now, it was assumed that the derivative of h is positive. Keeping the assumption of monotonicity, one potential generalisation could consist in allowing $h'(y) = 0$ at least for some real numbers $y \in \mathbb{R}$. In the following, another reasoning for identifying B is presented, which does not require $h'(y_0) > 0$. For this purpose, define for any $\alpha \in \mathbb{R}\backslash\{0\}$ the function

$$\begin{aligned} \varphi_\alpha : \ \mathbb{R} &\to \ \mathbb{R} \\ y &\mapsto \ \operatorname{sign}(y)|y|^\alpha. \end{aligned}$$

Lemma 3.6.3 *Let X, ε, g and σ be as in assumptions (**A1**)–(**A3**) and (**A7**) from Section 3.4. Further, let $\tilde{\varepsilon}$ be a centred random variable independent of X with* $\operatorname{Var}(\tilde{\varepsilon}) = 1$. *Let $\tilde{g}$ and $\tilde{\sigma}$ be functions, such that for some $\alpha > 0$ with $E[\varphi_\alpha(g(X) + \sigma(X)\varepsilon)^2] < \infty$ one has*

$$\varphi_\alpha(g(X) + \sigma(X)\varepsilon) = \tilde{g}(X) + \tilde{\sigma}(X)\tilde{\varepsilon} \quad \text{almost surely.}$$

Then, it holds that $\alpha = 1$.

Proof: According to assumption (**A7**), assume w.l.o.g. that $x \mapsto \frac{g(x)}{\sigma(x)}$ is not almost surely constant on $M_{>c}$ for an appropriate $c > 0$ and that $P(X \in M_{>c}) > 0$ holds (the other case can be treated analogously). Let $M_X \subseteq M_{>c}$ be a bounded subset such that

(i) $x \mapsto \frac{g(x)}{\sigma(x)}$ is not almost surely constant on M_X,

(ii) $P(X \in M_X) > 0$,

(iii) $\sup\limits_{x \in M_X} \left|\frac{\sigma(x)}{g(x)}\right| < \infty$.

Consequently, there exists some $\delta > 0$ such that

$$\inf_{x \in M_X,\, e \in (-\delta,\delta)} g(x) + \sigma(x)e > 0 \quad \text{and} \quad \sup_{x \in M_X,\, e \in (-\delta,\delta)} \left|\frac{\sigma(x)e}{g(x)}\right| < 1.$$

The generalized Binomial Theorem provides

$$I_{M_X}(X) I_{(-\delta,\delta)}(\varepsilon)(g(X) + \sigma(X)\varepsilon)^\alpha = I_{M_X}(X) I_{(-\delta,\delta)}(\varepsilon) g(X)^\alpha \sum_{k=0}^\infty \binom{\alpha}{k} \left(\frac{\sigma(X)}{g(X)}\right)^k \varepsilon^k.$$

Since the conditional expectation, conditional variance and the remaining error term can be written as

$$\tilde{g}(X) = E\big[\varphi_\alpha(g(X) + \sigma(X)\varepsilon)|X\big], \quad \tilde{\sigma}^2(X) = \operatorname{Var}\big(\varphi_\alpha(g(X) + \sigma(X)\varepsilon)|X\big)$$

and

$$\tilde{\varepsilon} = \frac{\varphi_\alpha(g(X) + \sigma(X)\varepsilon) - \tilde{g}(X)}{\tilde{\sigma}(X)},$$

it holds that

$$I_{M_X}(X) I_{(-\delta,\delta)}(\varepsilon)\tilde{\varepsilon} = I_{M_X}(X) I_{(-\delta,\delta)}(\varepsilon) \frac{\varphi_\alpha(g(X) + \sigma(X)\varepsilon) - \tilde{g}(X)}{\tilde{\sigma}(X)}$$

$$= I_{M_X}(X) I_{(-\delta,\delta)}(\varepsilon) \frac{\sum_{k=0}^{\infty} \binom{\alpha}{k} \sigma(X)^k g(X)^{\alpha-k} \varepsilon^k - \tilde{g}(X)}{\tilde{\sigma}(X)}$$

$$= I_{M_X}(X) I_{(-\delta,\delta)}(\varepsilon) \underbrace{\left(\frac{g(X)^\alpha - \tilde{g}(X)}{\tilde{\sigma}(X)} + \sum_{k=1}^{\infty} \binom{\alpha}{k} \frac{\sigma(X)^k g(X)^{\alpha-k}}{\tilde{\sigma}(X)} \varepsilon^k \right)}_{\text{independent of } X \in M_X}.$$

This can be alternatively expressed as

$$I_{M_X}(X) I_{(-\delta,\delta)}(\varepsilon) \tilde{\varepsilon} = I_{M_X}(X) I_{(-\delta,\delta)}(\varepsilon) \sum_{k=0}^{\infty} \beta_k(X) \varepsilon^k$$

with

$$\beta_0(X) = \frac{g(X)^\alpha - \tilde{g}(X)}{\tilde{\sigma}(X)} \quad \text{and} \quad \beta_k(X) = \binom{\alpha}{k} \frac{\sigma(X)^k g(X)^{\alpha-k}}{\tilde{\sigma}(X)} \quad \text{for all } k \geq 1.$$

Due to the independence of $\tilde{\varepsilon}$ and X, the coefficients $\beta_k, k \geq 0$, are not allowed to depend on X. Since $\frac{\sigma(X)}{g(X)}$ by assumption depends on X, $\frac{g(X)^\alpha \sigma(X)^k}{\tilde{\sigma}(X) g(X)^k}$ is at most for one $k \in \mathbb{N}_0$ independent of X. Thus, $\binom{\alpha}{k} \neq 0$ for at most one k. Due to $\alpha > 0$ one has $\alpha = 1$. □

It is conjectured that similar techniques to the proof above can be used to identify model 3.1 without assuming $h' > 0$. Nevertheless, the following conjecture has not been completely proven so far, so that only a sketch of a possible proof is given.

Conjecture 3.6.4 *Assume **(A1)**–**(A3)**,**(A5)**–**(A7)** from Section 3.4 as well as conditions (3.9) for $\lambda_1 = 1$ and (3.12). Let the conditional distribution function $F_{Y|X}(y|x)$ be continuously differentiable with respect to y and x and assume $h'(y) > 0$ for all $y \neq y_0$. Then, the unique solution to the model equation (3.1) is given by*

$$h(y) = \begin{cases} \exp\Big(-B \int_{y_1}^{y} \frac{1}{\lambda(u)}\, du \Big) & y > y_0 \\ 0 & y = y_0 \\ \lambda_2 \exp\Big(-B \int_{y_2}^{y} \frac{1}{\lambda(u)}\, du \Big) & y < y_0 \end{cases},$$

where it is set $\frac{1}{0} := \infty$ as well as $\frac{1}{\infty} := 0$ and λ_2 is uniquely determined. Further, one has

$$g(x) = E[h(Y)|X = x] \quad \text{and} \quad \sigma(x) = \sqrt{\mathrm{Var}(h(Y)|X = x)}.$$

Sketch of the Proof: The case $h'(y_0) > 0$ has been considered in Theorem 3.2.4 so that $h(y_0) = 0$ is assumed. Moreover, assume w.l.o.g. $B > 0$. For $B = 0$ the approach of Chiappori et al. (2015) can be adjusted in the same way as follows.
Let h and $\tilde{h}$ fulfil model (3.1) with (3.9) for $\lambda_1 = 1$ and (3.12), that is, h and $\tilde{h}$ are solutions to the differential equalities

$$h'(y) = -\frac{A + Bh(y)}{\lambda(y)} \quad \tilde{h}'(y) = -\frac{\tilde{A} + \tilde{B}\tilde{h}(y)}{\lambda(y)}$$

for some $B, \tilde{B} > 0, A, \tilde{A} \in \mathbb{R}$ and all $y \neq y_0$. By the same reasoning as before, h and $\tilde{h}$ can be written as

$$h(y) = \begin{cases} \exp\Big(-B \int_{y_1}^{y} \frac{1}{\lambda(u)}\, du \Big) & y > y_0 \\ 0 & y = y_0 \\ \lambda_2 \exp\Big(-B \int_{y_2}^{y} \frac{1}{\lambda(u)}\, du \Big) & y < y_0 \end{cases}$$

and

$$\tilde{h}(y) = \begin{cases} \exp\Big(- \tilde{B} \int_{y_1}^{y} \frac{1}{\lambda(u)}\, du \Big) & y > y_0 \\ 0 & y = y_0 \\ \tilde{\lambda}_2 \exp\Big(- \tilde{B} \int_{\tilde{y}_2}^{y} \frac{1}{\lambda(u)}\, du \Big) & y < y_0 \end{cases},$$

where $A = \tilde{A} = 0$ is implied by the location constraint (3.12). Due to assumption (**A3**) one has

$$\lim_{y \to -\infty} h(y) = -\infty \quad \text{and} \quad \lim_{y \to \infty} h(y) = \infty.$$

Therefore, the transformation functions h and $\tilde{h}$ from above can be written as

$$h(y) = \begin{cases} \exp\Big(- B \int_{y_1}^{y} \frac{1}{\lambda(u)}\, du \Big) & y > y_0 \\ 0 & y = y_0 \\ \exp\Big(- B \int_{y_3}^{y} \frac{1}{\lambda(u)}\, du \Big) & y < y_0 \end{cases} \tag{3.20}$$

and

$$\tilde{h}(y) = \begin{cases} \exp\Big(- \tilde{B} \int_{y_1}^{y} \frac{1}{\lambda(u)}\, du \Big) & y > y_0 \\ 0 & y = y_0 \\ \exp\Big(- \tilde{B} \int_{\tilde{y}_3}^{y} \frac{1}{\lambda(u)}\, du \Big) & y < y_0 \end{cases}$$

for some appropriate $y_3, \tilde{y}_3 < y_0$, which are uniquely determined by λ, B, λ_2 and $\lambda, \tilde{B}, \tilde{\lambda}_2$, respectively. These expressions in turn yield

$$\varphi_{\frac{\tilde{B}}{B}}(h(y)) = \begin{cases} \exp\Big(- \tilde{B} \int_{y_1}^{y} \frac{1}{\lambda(u)}\, du \Big) & y > y_0 \\ 0 & y = y_0 \\ \exp\Big(- \tilde{B} \int_{y_3}^{y} \frac{1}{\lambda(u)}\, du \Big) & y < y_0 \end{cases}$$

$$= \tilde{h}(y) + \tilde{h}(y) I_{\{y<y_0\}} \bigg(\exp\Big(- \tilde{B} \int_{y_3}^{\tilde{y}_3} \frac{1}{\lambda(u)}\, du \Big) - 1 \bigg),$$

that is,

$$\tilde{h}(y) = \varphi_{\frac{\tilde{B}}{B}}(h(y)) - \tilde{h}(y) I_{\{y<y_0\}} \bigg(\exp\Big(- \tilde{B} \int_{y_3}^{\tilde{y}_3} \frac{1}{\lambda(u)}\, du \Big) - 1 \bigg).$$

It is conjectured that similar arguments to the proof of Lemma 3.6.3 lead to

$$\exp\Big(- \tilde{B} \int_{y_3}^{\tilde{y}_3} \frac{1}{\lambda(u)}\, du \Big) = 1$$

and thus to $\tilde{y}_3 = y_3$. Finally, Lemma 3.6.3 ensures $\tilde{B} = B$ and consequently $\tilde{h} = h$.

It is conjectured that this result can be extended to a more general class of monotonic functions.

3.6.4 Uniqueness of Solutions to Ordinary Differential Equations

Finally, two basic results about ordinary differential equations and uniqueness of possible solutions are given. Theorem 3.6.6 is slightly modified compared to the version of Forster (1999, p. 102) so that the proof is presented as well.

Lemma 3.6.5 *(Gronwall's Inequality, see Grönwall (1919) or Bellman (1953) for details) Let $I = [a, b] \subseteq \mathbb{R}$ be a compact interval. Let $u, v : I \to \mathbb{R}$ and $q : I \to [0, \infty)$ be continuous functions. Further, let*

$$u(y) \leq v(y) + \int_a^y q(z)u(z)\,dz$$

for all $y \in I$. Then, one has

$$u(y) \leq v(y) + \int_a^y v(z)q(z)\exp\left(\int_z^y q(t)\,dt\right) dz \quad \text{for all } y \in I.$$

Theorem 3.6.6 *(see Forster (1999, p. 102) for a related version) Let $b > a > y_0$ and $G \subseteq (y_0, \infty) \times \mathbb{R}^+$ be a set such that $[a, b] \times \mathbb{R}^+ \subseteq G$. Moreover, let $D : G \to \mathbb{R}, (y, h) \mapsto D(y, h)$, be continuous with respect to both components and continuously differentiable with respect to the second component. Then, for all $\theta_0 > 0$ any solution $h \in \mathcal{C}([a, b], \mathbb{R}^+)$ of the initial value problem*

$$h'(y) = D(y, h(y)), \quad h(a) = \theta_0$$

is unique.

Proof: Let $h_1, h_2 : [a, b] \to \mathbb{R}^+$ be two solutions of the mentioned initial value problem. Since

$$K := \{(y, \theta) \in [a, b] \times \mathbb{R}^+ : y \in [a, b], \theta \in \{h_1(y), h_2(y)\}\}$$

is compact, there exists some $L > 0$ such that $|D(y, \theta) - D(y, \psi)| \leq L|\theta - \psi|$ for all $(y, \theta), (y, \psi) \in K$. Consider the distance $d(y) := |h_1(y) - h_2(y)|$. Then for all $y \in [a, b]$

$$\begin{aligned} d(y) &= |h_1(y) - h_1(a) - (h_2(y) - h_2(a))| \\ &= \left| \int_a^y (D(z, h_1(z)) - D(z, h_2(z)))\,dz \right| \\ &\leq \int_a^y |D(z, h_1(z)) - D(z, h_2(z))|\,dz \\ &\leq L \int_a^y |h_1(z) - h_2(z)|\,dz \\ &= L \int_a^y d(z)\,dz \end{aligned}$$

Gronwall's Inequality leads to $d \leq 0$ (set $u = d, v \equiv 0, q \equiv L$). □

4

Nonparametric Estimation of the Transformation Function in a Heteroscedastic Model

After identifiability of model (3.1) under conditions (3.9) and (3.12) was proven in the last chapter, the question arises how its components can be estimated appropriately. To the author's knowledge, there is no estimating approach in such a general model as (3.1) so far. To mention only some methods in the literature, Chiappori et al. (2015) provided an estimator for homoscedastic models, while Neumeyer et al. (2016) extended the ideas of Linton et al. (2008) to the case of heteroscedastic errors, but only for parametric transformation functions. In the context of a linear regression function, Horowitz (2009) discussed several approaches for a parametric/ nonparametric transformation function and a parametric/ nonparametric distribution function of the error term.
In the following, the analytical expressions of the model components in (3.1) are used to construct corresponding estimators in Section 4.1. Afterwards, the asymptotic behaviour of these estimators is examined in Section 4.2. When doing so, equation (3.17) and the ideas of Horowitz (1996) will play key roles in defining estimators and deriving the asymptotic behaviour. Some simulations are conducted in Section 4.3 and the chapter is concluded by a short discussion in 4.4. The proofs can be found in Section 4.6.
Throughout this chapter, assume (**A1**)–(**A7**) from Section 3.4 as well as $B > 0$ (see Remark 3.2.1). Moreover, assume the location and scale constraints (3.9) and (3.12) for some $y_1 > y_0$ with $\lambda_1 = 1$ and let $(Y_i, X_i), i = 1, ..., n$, be independent and identically distributed observations from model (3.1).

4.1 Definition of the Estimator

An estimation technique is presented, which combines parts of the work of Chiappori et al. (2015) and that of Horowitz (1996) and Linton et al. (2008). First, some preliminary definitions are needed before the final estimator is defined in Subsection 4.1.3.

4.1.1 Estimation of $\boldsymbol{\lambda}$ and $\boldsymbol{y_0}$

As in Section 3.2, λ is defined as

$$\lambda(y) = \int v(x) \frac{\frac{\partial F_{Y|X}(y|x)}{\partial x_1}}{\frac{\partial F_{Y|X}(y|x)}{\partial y}}\, dx$$

for some weight function v and y_0 is defined by the equation $\lambda(y_0) = 0$. In this thesis, a Plug-In-approach is used to estimate first λ by some kernel estimator $\hat{\lambda}$ and then estimate y_0 by the root of $\hat{\lambda}$. To be precise, the conditional distribution function $F_{Y|X}$ is estimated for some kernel function K and some bandwidth sequences $h_y \searrow 0$ and $h_x \searrow 0$ by

$$\hat{F}_{Y|X}(y|x) = \frac{\hat{p}(y, x)}{\hat{f}_X(x)}$$

with $\hat{f}_X$ and $\hat{p}$ as defined in equations (1.2) and (1.5). Then, this estimator is plugged into the expression for λ yielding

$$\hat{\lambda}(y) = \int v(x) \frac{\frac{\partial \hat{F}_{Y|X}(y|x)}{\partial x_1}}{\frac{\partial \hat{F}_{Y|X}(y|x)}{\partial y}}\, dx. \tag{4.1}$$

Note that by construction and assumption (**B2**) from Section 4.5 the estimated conditional distribution function $\hat{F}_{Y|X}$ is continuously differentiable. Once λ is estimated an estimator for y_0 can be defined as the solution to $\hat{\lambda}(y) = 0$. In Section 4.1.4 it will be shown that for arbitrary large compact sets $\mathcal{K} \subseteq \mathbb{R}$ there will be at most one solution with probability converging to one for $n \to \infty$. Since for finite sample sizes it might be the case that there is more than one solution, an estimator is defined by

$$\hat{y}_0 = \operatorname*{arg\,min}_{y:\hat{\lambda}(y)=0} |y|. \tag{4.2}$$

Assumption (**A3**) from Section 3.4 and $B \neq 0$ ensure that there exists a root of λ, since h is surjective under (**A3**). Hence, due to the uniform convergence of $\hat{\lambda}$ to λ, which is proven in Lemma 4.2.1 below, $\hat{\lambda}$ possesses a root (that is close to y_0) as well with probability converging to one. Details will be given in Subsection 4.1.4.

4.1.2 Estimation of $\boldsymbol{B}$

Recall

$$h(y) = \exp\Big(- B \int_{y_1}^{y} \frac{1}{\lambda(u)}\, du\Big) \quad \text{for all } y > y_0$$

and some $y_1 > y_0$, that is, once y_1 is fixed and λ and y_0 are estimated appropriately it remains to estimate B, at least to estimate h on (y_0, ∞). Due to $B \in \mathbb{R}$ this can be seen as a parametric problem. Two approaches to estimate B will be provided in this section. Unfortunately, it will be seen that without further conditions the already existing methods in (semi-)parametric transformation models (e.g. of Linton et al. (2008) or Colling and Van Keilegom (2018)) can not be applied in the scenario here. The reason is that they rely on appropriate estimators for conditional mean and variance or require an appropriate

nonparametric estimator. See Section 1.3 for details on these procedures.
Nevertheless, proceeding similarly to Horowitz (1996), an estimator for B can be deduced, which converges under several conditions to B with a $\sqrt{n}$-rate as will be seen in Section 4.2.

Estimation of B via the Derivative of λ

Since the convergence rate of the estimator presented later relies on some additional assumptions, first a less sophisticated estimator is provided, which is based on equation (3.19) and less computationally demanding, but achieves a slower convergence rate compared to the second estimator, which is presented later. Under the conditions (3.9) and (3.12), it was shown in Section 3.2 that

$$\lambda(y) = -\frac{Bh(y)}{\frac{\partial}{\partial y}h(y)}$$

and

$$\frac{\partial}{\partial y}\lambda(y)\Big|_{y=y_0} = -B.$$

Plugging the estimators for λ and y_0 given in Section 4.1.1 into the previous equation, leads to the estimator

$$\tilde{B} := -\frac{\partial}{\partial y}\hat{\lambda}(y)\Big|_{y=\hat{y}_0}. \tag{4.3}$$

Later, asymptotic normality of this estimator will be shown in Subsection 4.2.1.

The Mean Square Distance from Independence Approach

Now, a more sophisticated approach for estimating B will be presented. Apart from using conditional quantiles instead of the conditional mean, this estimator will be related to the Mean-Square-Distance-From-Independence estimator of Linton et al. (2008). Let c be a parameter that needs to be examined. The basic idea of the estimator is that for all parameters c some appropriately defined residuals are independent of X if and only if c is equal to the true parameter, which will be B in this Section. This idea and the definition of the residuals will be explained in detail below.
To examine the estimator, let U, V be some random variables, where U is real valued, $\tau \in (0,1)$ and denote the τ-quantile of U conditional on $V = v$ by

$$F_{U|V}^{-1}(\tau|v) = \inf\{u \in \mathbb{R} : F_{U|V}(u|v) \geq \tau\}.$$

F_ε and f_ε denote the distribution function and density of ε. Since h is assumed to be strictly increasing and

$$h(Y) = g(X) + \sigma(X)\varepsilon$$

with ε independent of X, it holds that (with $F_\varepsilon^{-1}(\tau) = F_{\varepsilon|X}^{-1}(\tau|X)$)

$$h(F_{Y|X}^{-1}(\tau|X)) = F_{h(Y)|X}^{-1}(\tau|X) = g(X) + \sigma(X)F_{\varepsilon|X}^{-1}(\tau|X).$$

Especially, one has

$$h(Y) - h(F_{Y|X}^{-1}(\tau|X)) = g(X) + \sigma(X)\varepsilon - g(X) - \sigma(X)F_\varepsilon^{-1}(\tau)$$

$$= \sigma(X)(\varepsilon - F_\varepsilon^{-1}(\tau)). \tag{4.4}$$

To obtain a random variable independent of X one has to adjust for the standard error $\sigma(X)$. This can be done in several ways: Consider for some $\beta \in (0,1)$

(i) $F^{-1}_{\sigma(X)^2(\varepsilon - F_\varepsilon^{-1}(\tau))^2|X}(\beta|X) = \sigma(X)^2 F^{-1}_{(\varepsilon - F_\varepsilon^{-1}(\tau))^2}(\beta)$,

(ii) $F^{-1}_{\sigma(X)|\varepsilon - F_\varepsilon^{-1}(\tau)|\,|X}(\beta|X) = \sigma(X) F^{-1}_{|\varepsilon - F_\varepsilon^{-1}(\tau)|}(\beta)$,

(iii) $F^{-1}_{\sigma(X)(\varepsilon - F_\varepsilon^{-1}(\tau))|X}(\beta|X) = \sigma(X)\big(F_\varepsilon^{-1}(\beta) - F_\varepsilon^{-1}(\tau)\big)$.

Note that due to $\sigma, f_\varepsilon > 0$ all of these expressions are different from zero (in the third case consider $\beta \neq \tau$) so that the quotients

$$\frac{h(Y) - h(F^{-1}_{Y|X}(\tau|X))}{\sqrt{F^{-1}_{\sigma(X)^2(\varepsilon - F_\varepsilon^{-1}(\tau))^2|X}(\beta|X)}} = \frac{\varepsilon - F_\varepsilon^{-1}(\tau)}{\sqrt{F^{-1}_{(\varepsilon - F_\varepsilon^{-1}(\tau))^2}(\beta)}}, \tag{4.5}$$

$$\frac{h(Y) - h(F^{-1}_{Y|X}(\tau|X))}{F^{-1}_{\sigma(X)|\varepsilon - F_\varepsilon^{-1}(\tau)|\,|X}(\beta|X)} = \frac{\varepsilon - F_\varepsilon^{-1}(\tau)}{F^{-1}_{|\varepsilon - F_\varepsilon^{-1}(\tau)|}(\beta)}, \tag{4.6}$$

$$\frac{h(Y) - h(F^{-1}_{Y|X}(\tau|X))}{F^{-1}_{\sigma(X)(\varepsilon - F_\varepsilon^{-1}(\tau))|X}(\beta|X)} = \frac{\varepsilon - F_\varepsilon^{-1}(\tau)}{F_\varepsilon^{-1}(\beta) - F_\varepsilon^{-1}(\tau)} \quad =: \tilde{\varepsilon} \tag{4.7}$$

are well defined. Principally, all of these standardisations can be used to construct an estimator. Nevertheless, only the third approach is considered in the following. Note that $\tilde{\varepsilon}$ is independent of X if and only if ε is independent of X.
Assume a quantile, for which lower and upper bounds are known, needs to be estimated. As in the paper of Horowitz (1996) the idea is used that the exact value of an observation should not influence an appropriate estimator of the quantile if the observation exceeds one of these bounds. This property will turn out to be the crucial advantage of using the estimated conditional quantile instead of the mean, like for example in the paper of Linton et al. (2008). Since parametric classes of transformation functions were considered there, the problem of estimating the mean after transforming Y could be solved by assuming A.5 (Linton et al., 2008, p. 700), a uniform (with respect to the transformation parameter) integrability condition of the derivatives with respect to the parameter.
Assume $B \in [B_1, B_2]$ for some $0 < B_1 < B_2$ and define

$$h_c(y) = \exp\Big(-c\int_{y_1}^{y} \frac{1}{\lambda(u)}\,du\Big) \quad \text{and} \quad \hat{h}_c(y) = \exp\Big(-c\int_{y_1}^{y} \frac{1}{\hat{\lambda}(u)}\,du\Big) \tag{4.8}$$

for $y > y_0$ and (compare to (4.4) and (4.7))

$$\tilde{\varepsilon}_c = \frac{h_c(Y) - h_c(F^{-1}_{Y|X}(\tau|X))}{h_c(F^{-1}_{Y|X}(\beta|X)) - h_c(F^{-1}_{Y|X}(\tau|X))}.$$

Moreover, define for an estimator $\bar{h}_1$ of h_1 and $c \in [B_1, B_2]$

$$\bar{h}_c(y) = \operatorname{sign}(\bar{h}_1(y))|\bar{h}_1(y)|^c. \tag{4.9}$$

Consequently, $h_B = h$ is the true transformation function. As will be seen later, it suffices to consider the case $Y > y_0$ here. In Chapter 3, it was shown that $c = B$ is the only value such that $\tilde{\varepsilon}_c$ is independent of X (see Lemma 3.6.3). As in Chapter 2, $F_{Y|X}$ and consequently $F_{Y|X}^{-1}$ as well as $\tilde{\varepsilon}_c$ can be estimated by replacing $F_{Y|X}$ with

$$\hat{F}_{Y|X}(y|x) = \frac{\sum_{i=1}^{n} \mathcal{K}_{h_y}(y - Y_i)\mathbf{K}_{h_x}(x - X_i)}{\sum_{i=1}^{n} \mathbf{K}_{h_x}(x - X_i)}. \tag{4.10}$$

Uniform convergence of $\hat{F}_{Y|X}^{-1}$ to $F_{Y|X}^{-1}$ was shown in Lemma 2.8.1. Consider a given interval $[z_a, z_b] \subseteq (y_0, \infty)$ and let $\tau < \beta \in (0,1)$, $[e_a, e_b] \subseteq \mathbb{R}$ and M_X be a non-random interval such that

(M1) $M_X \subseteq \operatorname{supp}(v)$ and $f_X(x) > 0$ for all $x \in M_X$,

(M2) $x \mapsto \frac{g(x)}{\sigma(x)}$ is not almost surely constant on M_X,

(M3) $F_{Y|X}^{-1}(\tau|x), F_{Y|X}^{-1}(\beta|x) \in (z_a, z_b)$ for all $x \in M_X$,

(M4) $\sup_{x \in M_X, e \in [e_a,e_b], c \in [B_1,B_2]} h_c(F_{Y|X}^{-1}(\tau|x)) + e(h_c(F_{Y|X}^{-1}(\beta|x)) - h_c(F_{Y|X}^{-1}(\tau|x))) < h_c(z_b)$ and

(M5) $\inf_{x \in M_X, e \in [e_a,e_b], c \in [B_1,B_2]} h_c(F_{Y|X}^{-1}(\tau|x)) + e(h_c(F_{Y|X}^{-1}(\beta|x)) - h_c(F_{Y|X}^{-1}(\tau|x))) > h_c(z_a)$.

Since M_X is an interval, the boundary of M_X has Lebesgue-measure equal to zero. See example 4.1.3 for a (admittedly, rather technical) way to construct a set M_X fulfilling these assumptions.

Remark 4.1.1 *1. It holds that*

$$M_X \subseteq \bigcap_{e \in [e_a,e_b], c \in [B_1,B_2]} \Big\{ x : h_c(z_a) \leq h_c(F_{Y|X}^{-1}(\tau|x)) + e(h_c(F_{Y|X}^{-1}(\beta|x)) - h_c(F_{Y|X}^{-1}(\tau|x))) \leq h_c(z_b) \Big\}.$$

2. *Condition **(M1)** can be relaxed to the case, where there exists some subset $\tilde{M}_X \subseteq M_X$ that fulfils **(M1)** (and **(M2)**–**(M5)**).*

3. *If $[e_a, e_b] \subseteq [0,1]$, conditions **(M4)** and **(M5)** are implied by*

$$\sup_{x \in M_X, c \in [B_1,B_2]} \max\big(h_c(F_{Y|X}^{-1}(\tau|x)), h_c(F_{Y|X}^{-1}(\beta|x))\big) < h_c(z_b)$$

and

$$\inf_{x \in M_X, c \in [B_1,B_2]} \min\big(h_c(F_{Y|X}^{-1}(\tau|x)), h_c(F_{Y|X}^{-1}(\beta|x))\big) > h_c(z_a).$$

4. *Let $\bar{h}_1, \hat{f}_{m_\tau}$ and $\hat{f}_{m_\beta}$ be some estimators, such that $\bar{h}_1(y)$ converges uniformly in $y \in [z_a, z_b]$ to $h_1(y)$ and $\hat{f}_{m_\tau}(x), \hat{f}_{m_\beta}(x)$ converge uniformly in $x \in M_X$ to $F_{Y|X}^{-1}(\tau|x)$ and $F_{Y|X}^{-1}(\beta|x)$. Then, conditions **(M1)**–**(M5)** imply $P(\tilde{\varepsilon}_B \leq e) = P(\tilde{\varepsilon}_B \leq e | X \in M_X)$ as well as*

$$M_X \subseteq \bigcap_{e \in [e_a,e_b], c \in [B_1,B_2]} \Big\{ x : \bar{h}_c(z_a) \leq \bar{h}_c(\hat{f}_{m_\tau}(x)) + e(\bar{h}_c(\hat{f}_{m_\beta}(x)) - \bar{h}_c(\hat{f}_{m_\tau}(x))) \leq \bar{h}_c(z_b) \Big\}$$

with probability converging to one, where $\bar{h}_c$ is defined as in (4.9).

When estimating $P(\tilde{\varepsilon}_c \leq e)$ the problem arises that $\tilde{\varepsilon}_c$ can not be observed directly, but has to be estimated as well. Since nonparametric estimators such as $\hat{h}_c$ usually only converge to h_c with $\sqrt{n}$-rate on compact subsets of (y_0, ∞), $\tilde{\varepsilon}$ can not be estimated with $\sqrt{n}$-rate in general. Here, the advantage of using the conditional quantiles instead of the conditional mean in (4.5)–(4.7) becomes clear:
After conditioning on $X \in M_X$, $P(\tilde{\varepsilon}_c \leq e|X \in M_X)$ can be estimated by

$$\hat{P}(\tilde{\varepsilon}_c \leq e|X \in M_X) = \frac{\frac{1}{n}\sum_{i=1}^n I_{\{\hat{\tilde{\varepsilon}}_{c,i} \leq e\}} I_{\{X_i \in M_X\}}}{\frac{1}{n}\sum_{i=1}^n I_{\{X_i \in M_X\}}},$$

where

$$\hat{\tilde{\varepsilon}}_{c,i} = \frac{\hat{h}_c(Y_i) - \hat{h}_c(\hat{F}^{-1}_{Y|X}(\tau|X_i))}{\hat{h}_c(\hat{F}^{-1}_{Y|X}(\beta|X_i)) - \hat{h}_c(\hat{F}^{-1}_{Y|X}(\tau|X_i))}.$$

Although $\hat{h}_c$ might not be a $\sqrt{n}$-consistent estimator for h_c on $\mathbb{R}$, it is still strictly monotonic. Since $X \in M_X$ implies

$$\hat{h}_c(z_a) \leq \hat{h}_c(\hat{F}^{-1}_{Y|X}(\tau|X)) + e(\hat{h}_c(\hat{F}^{-1}_{Y|X}(\beta|X)) - \hat{h}_c(\hat{F}^{-1}_{Y|X}(\tau|X))) \leq \hat{h}_c(z_b),$$

one has

$$\frac{\hat{h}_c(z_a) - \hat{h}_c(\hat{F}^{-1}_{Y|X}(\tau|X))}{\hat{h}_c(\hat{F}^{-1}_{Y|X}(\beta|X)) - \hat{h}_c(\hat{F}^{-1}_{Y|X}(\tau|X))} \leq \hat{\tilde{\varepsilon}}_c + e - \hat{\tilde{\varepsilon}}_c \leq \frac{\hat{h}_c(z_b) - \hat{h}_c(\hat{F}^{-1}_{Y|X}(\tau|X))}{\hat{h}_c(\hat{F}^{-1}_{Y|X}(\beta|X)) - \hat{h}_c(\hat{F}^{-1}_{Y|X}(\tau|X))}.$$

Consequently, monotonicity of $\hat{h}$ leads to

$$\begin{aligned} Y < z_a \quad &\Rightarrow \quad \hat{\tilde{\varepsilon}}_c < \frac{\hat{h}_c(z_a) - \hat{h}_c(\hat{F}^{-1}_{Y|X}(\tau|X))}{\hat{h}_c(\hat{F}^{-1}_{Y|X}(\beta|X)) - \hat{h}_c(\hat{F}^{-1}_{Y|X}(\tau|X))} \quad \Rightarrow \quad \hat{\tilde{\varepsilon}}_c < e, \\ Y > z_b \quad &\Rightarrow \quad \hat{\tilde{\varepsilon}}_c > \frac{\hat{h}_c(z_b) - \hat{h}_c(\hat{F}^{-1}_{Y|X}(\tau|X))}{\hat{h}_c(\hat{F}^{-1}_{Y|X}(\beta|X)) - \hat{h}_c(\hat{F}^{-1}_{Y|X}(\tau|X))} \quad \Rightarrow \quad \hat{\tilde{\varepsilon}}_c > e, \end{aligned}$$

if $X \in M_X$. Therefore, $\hat{\tilde{\varepsilon}}_c$ only has to be calculated when $Y \in [z_a, z_b]$, which means that all results about uniform convergence on compact sets like 4.2.2 can be applied without worsening convergence rates. See Horowitz (1996, p. 107) for a similar reasoning.
Consider $\mathfrak{h}, f_{m_\tau}, f_{m_\beta}$ belonging to specific function sets specified later and define $s = (\mathfrak{h}, f_{m_\tau}, f_{m_\beta})^t$ as well as $s_0 = (h_1, F^{-1}_{Y|X}(\tau|\cdot), F^{-1}_{Y|X}(\beta|\cdot))^t$ with h_1 from (4.8) and

$$\tilde{\varepsilon}_c(s) = \frac{\mathfrak{h}(Y)^c - \mathfrak{h}(f_{m_\tau}(X))^c}{\mathfrak{h}(f_{m_\beta}(X))^c - \mathfrak{h}(f_{m_\tau}(X))^c},$$

$$\begin{aligned} G_{MD}(c,s)(x,e) &= P\big(X \leq x, \tilde{\varepsilon}_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta}) \leq e|X \in M_X\big) \\ &\quad - P\big(X \leq x|X \in M_X\big)P\big(\tilde{\varepsilon}_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta}) \leq e|X \in M_X\big), \qquad (4.11) \\ G_{nMD}(c,s)(x,e) &= \hat{P}\big(X \leq x, \tilde{\varepsilon}_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta}) \leq e|X \in M_X\big) \\ &\quad - \hat{P}\big(X \leq x|X \in M_X\big)\hat{P}\big(\tilde{\varepsilon}_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta}) \leq e|X \in M_X\big) \qquad (4.12) \end{aligned}$$

with

$$\hat{P}\big(X \leq x, \tilde{\varepsilon}_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta}) \leq e | X \in M_X\big) = \frac{\frac{1}{n}\sum_{i=1}^n I_{\{\tilde{\varepsilon}_{c,i}(s) \leq e\}} I_{\{X_i \leq x\}} I_{\{X_i \in M_X\}}}{\frac{1}{n}\sum_{i=1}^n I_{\{X_i \in M_X\}}}$$

$$\hat{P}\big(X \leq x | X \in M_X\big) = \frac{\frac{1}{n}\sum_{i=1}^n I_{\{X_i \leq x\}} I_{\{X_i \in M_X\}}}{\frac{1}{n}\sum_{i=1}^n I_{\{X_i \in M_X\}}}$$

$$\hat{P}\big(\tilde{\varepsilon}_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta}) \leq e | X \in M_X\big) = \frac{\frac{1}{n}\sum_{i=1}^n I_{\{\tilde{\varepsilon}_{c,i}(s) \leq e\}} I_{\{X_i \in M_X\}}}{\frac{1}{n}\sum_{i=1}^n I_{\{X_i \in M_X\}}}.$$

Moreover, define

$$A(c,s) := \sqrt{\int_{M_X} \int_{[e_a,e_b]} G_{MD}(c,s)(x,e)^2 \, de \, dx} = ||G_{MD}(c,s)||_2, \tag{4.13}$$

where $||.||_2$ denotes the $\mathcal{L}^2$-norm on $M_X \times [e_a, e_b]$. Then, Lemma 3.6.3 implies $A(c, s_0) = 0$ if and only if $c = B$.
For some estimator $\hat{s}$ of s_0 the function $c \mapsto A(c, s_0)$ can be estimated by

$$\hat{A}(c,\hat{s}) := \sqrt{\int_{M_X} \int_{[e_a,e_b]} G_{nMD}(c,\hat{s})(x,e)^2 \, de \, dx} = ||G_{nMD}(c,\hat{s})||_2.$$

From now on, $\hat{s}$ will be defined as

$$\hat{s} = \big(\hat{h}_1, \hat{F}^{-1}_{Y|X}(\tau|X), \hat{F}^{-1}_{Y|X}(\beta|X)\big)^t \tag{4.14}$$

in this section, where $\hat{h}_1$ is defined as in (4.8) and $\hat{F}^{-1}_{Y|X}$ denotes the inverse of the estimator of the conditional distribution function as in (4.10). Minimizing $\hat{A}(c, \hat{s})$ with respect to c leads to the estimator

$$\hat{B} = \underset{c \in [B_1, B_2]}{\arg\min}\, \hat{A}(c, \hat{s}). \tag{4.15}$$

Remark 4.1.2 *Without further examination, some thoughts on testing for $H_0 : B = 0$ are given together with two possible testing approaches. Assume $B = 0$. Then, equation (3.6) implies*

$$\lambda(y) = -\frac{A}{\frac{\partial}{\partial y} h(y)}.$$

1. *Due to $\frac{\partial}{\partial y} h(y) > 0$, λ is well defined and has either no root (when $A \neq 0$) or infinitely many roots (when $A = 0$). This can be used to reject H_0, if there is only one root in a given interval $[z_a, z_b] \subseteq \mathbb{R}$.*

2. *The underlying estimating approach of $\hat{B}$ is based on the fact that the residuals corresponding to every $c \in \mathbb{R}$ are independent of X if and only if $c = B$. Hence, it could be possible to proceed as in the paper of Chiappori et al. (2015) and to test for independence of X and the residuals.*

Example 4.1.3 *Constructing an appropriate set M_X. Let $[z_a, z_b] \subseteq (y_0, \infty)$ be a given interval, $\{\tilde{X}_1, ..., \tilde{X}_q\} = \{X_1, ..., X_n : X_i \in \operatorname{supp}(v)\}$ for some appropriate $q \in \mathbb{N}$ the set of observations falling in the support of v and $x^* = \hat{\tilde{X}}_q$ the empirical mean of these observations. Define for each $k \in \mathbb{N}$ the (possibly empty) set*

$$Q_k := \left\{(\iota, \xi) : \iota < \xi, \iota, \xi \in \left\{\frac{1}{k}, ..., \frac{k-1}{k}\right\}, F_{Y|X}^{-1}(\iota|E[\tilde{X}]), F_{Y|X}^{-1}(\xi|E[\tilde{X}]) \in \left(z_a + \frac{1}{k}, z_b - \frac{1}{k}\right)\right\}$$

and for each $e \in \mathbb{R}, c \in [B_1, B_2], m \in \mathbb{N}$ and $\tau < \beta \in (0,1)$ the set

$$\Omega_{e,c,m}^{\tau,\beta} := \left\{x : h_c\left(z_a + \frac{1}{m}\right) < h_c(F_{Y|X}^{-1}(\tau|x)) + e\big(h_c(F_{Y|X}^{-1}(\beta|X)) - h_c(F_{Y|X}^{-1}(\tau|x))\big) < h_c\left(z_b - \frac{1}{m}\right)\right\}.$$

Further, for all $k \in \mathbb{N}$ define $(\tau_k, \beta_k) := \underset{(\iota,\xi)\in Q_k}{\arg\max}\{\xi - \iota\}$ and choose τ_k minimal if the maximizing values are not unique. Moreover, define

$$m_k = \min\left\{m \in \mathbb{N} : \bigcap_{e\in[-\frac{1}{m}, \frac{1}{m}],\, c\in[B_1,B_2]} \Omega_{e,c,m}^{\tau_k,\beta_k} \neq \emptyset\right\}$$

if the set of appropriate m is not empty (otherwise set $m_k = \infty$). When choosing $k^* =$ $\min\{k \in \mathbb{N} : \ m_k < \infty\}$, the interior of the set

$$M_X^* := \bigcap_{e\in[-\frac{1}{m_{k^*}}, \frac{1}{m_{k^*}}], c\in[B_1,B_2]} \Omega_{e,c,2m_{k^*}}^{\tau_{k^*},\beta_{k^*}} \neq \emptyset$$

is not empty, since $(y, c) \mapsto h_c(y)$ is uniformly continuous on compact sets. Now choose $l \in \mathbb{N} \cup \{\frac{1}{n} : n \in \mathbb{N}\}$ minimal such that

$$M_X := \frac{1}{l}\left[\begin{pmatrix} i_1 \\ \vdots \\ i_d \end{pmatrix}, \begin{pmatrix} i_1 \\ \vdots \\ i_d \end{pmatrix} + \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}\right] \subseteq \overset{\circ}{M_X^*}$$

holds for appropriate $i_1, ..., i_d \in \mathbb{Z}$, where $\overset{\circ}{M_X^*}$ denotes the interior of M_X^*.
Up to now, M_X is unknown in general and thus has to be approximated. Let $t_n = \frac{1}{log(n)}$ and define

$$\hat{Q}_k := \left\{(\iota, \xi) : \iota < \xi, \iota, \xi \in \left\{\frac{1}{k}, ..., \frac{k-1}{k}\right\}, \hat{F}_{Y|X}^{-1}(\iota|x^*), \hat{F}_{Y|X}^{-1}(\xi|x^*) \in \left(z_a + \frac{1}{k} + t_n, z_b - \frac{1}{k} - t_n\right)\right\},$$

$$(\hat{\tau}_k, \hat{\beta}_k) := \underset{(\iota,\xi)\in\hat{Q}_k}{\arg\max}\{\xi - \iota\},$$

$$\hat{\Omega}_{e,c,m}^{\tau,\beta} := \left\{x : \hat{h}_c\left(z_a + \frac{1}{m}\right) + t_n < \hat{h}_c(\hat{F}_{Y|X}^{-1}(\tau|x)) + e\big(\hat{h}_c(\hat{F}_{Y|X}^{-1}(\beta|X)) - \hat{h}_c(\hat{F}_{Y|X}^{-1}(\tau|x))\big) < \hat{h}_c\left(z_b - \frac{1}{m}\right) - t_n\right\}$$

as well as

$$\hat{m}_k = \min\left\{m \in \mathbb{N} : \bigcap_{e\in[-\frac{1}{m}, \frac{1}{m}], c\in[B_1,B_2]} \hat{\Omega}_{e,c,m_k}^{\hat{\tau}_k,\hat{\beta}_k} \neq \emptyset\right\}.$$

and $\hat{k}^* = \min\{k \in \mathbb{N} : \ \hat{m}_k < \infty\}$. In a similar way, estimators $\hat{l}$ and $\hat{M}_X$ for l and M_X can be defined. One has

$$x^* - E[\tilde{X}] = o_p(1), \quad \hat{h}_c(y) - h_c(y) = o_p(1) \quad \text{and} \quad \hat{F}_{Y|X}^{-1}(\tau|x^*) - F_{Y|X}^{-1}(\tau|E[\tilde{X}]) = o_p(1),$$

where the last two convergences hold uniformly on compact sets. Therefore,

$$P(\hat{k}^* = k^*) \to 1,$$

$$P(\hat{m}_{k^*} = m_{k^*}) \to 1,$$

$$P(\hat{\tau}_{k^*} = \tau_{k^*}) \to 1,$$

$$P(\hat{\beta}_{k^*} = \beta_{k^*}) \to 1,$$

$$P(\hat{l} = l) \to 1,$$

and consequently $P(\hat{M}_X = M_X) \to 1$, *which means that* M_X *can be viewed as known and not random.*

4.1.3 Putting Things together

So far, estimators of all the components in (3.17) apart from λ_2 have been presented. These estimators are now combined to obtain an estimator of the transformation function h on (y_0, ∞). While doing so, it is assumed that some $y_1 \in (y_0, \infty)$ and a compact set $\mathcal{K} \subseteq (y_0, \infty)$, on which the transformation function h needs to be estimated, are given. The extension to $(-\infty, y_0)$ as well as the estimation of λ_2 are postponed to Section 4.1.4.
In (4.8) an estimator for h_c was already given. Note that $h = h_B$. Insert each of the estimators $\tilde{B}$ and $\hat{B}$ for B from (4.3) or (4.15) to get

$$\hat{h}(y) = \exp\Big(- \hat{B} \int_{y_1}^{y} \frac{1}{\hat{\lambda}(u)}\, du\Big), \quad y \in \mathcal{K}, \tag{4.16}$$

and

$$\tilde{h}(y) = \exp\Big(- \tilde{B} \int_{y_1}^{y} \frac{1}{\hat{\lambda}(u)}\, du\Big), \quad y \in \mathcal{K}. \tag{4.17}$$

4.1.4 Extending the Estimator to $(-\infty, y_0)$

So far, the estimator was only considered on compact sets $\mathcal{K} \subseteq (y_0, \infty)$. Now, the estimator is extended to arbitrary values $y \in \mathbb{R}$. Doing so requires estimators for y_0 and λ_2. While an estimator for y_0 was already defined in (4.2) an estimator for λ_2 is given first, before these are combined to an estimator $\hat{h}$ on $\mathbb{R}$ and the asymptotic behaviour is examined.

An Estimator for λ_2

The presented approach for estimating λ_2 will be similar to estimating B by $\tilde{B}$ in (4.3). Recall the analytic expression (3.17) for h, that is

$$h(y) = \begin{cases} \exp\Big(- B \int_{y_1}^{y} \frac{1}{\lambda(u)}\, du\Big) & y > y_0 \\ 0 & y = y_0 \\ \lambda_2 \exp\Big(- B \int_{y_2}^{y} \frac{1}{\lambda(u)}\, du\Big) & y < y_0 \end{cases}$$

for some arbitrary fixed value $y_2 < y_0$. It is known that λ_2 is uniquely determined by

$$\lim_{y \searrow y_0} \frac{\partial}{\partial y} h(y) = \lim_{y \nearrow y_0} \frac{\partial}{\partial y} h(y) \qquad \Bigg(= \quad \frac{\partial}{\partial y} h(y_0) > 0 \Bigg)$$

as

$$\lambda_2 = -\lim_{t\to 0} \exp\Bigg(B\bigg(\int_{y_2}^{y_0-t} \frac{1}{\lambda(u)}\,du - \int_{y_1}^{y_0+t} \frac{1}{\lambda(u)}\,du\bigg)\Bigg).$$

Since estimators for λ, B and y_0 are already available, these can be plugged in to obtain the estimator

$$\tilde{\lambda}_2 = -\exp\Bigg(\tilde{B}\bigg(\int_{y_2}^{\hat{y}_0-t_n} \frac{1}{\hat{\lambda}(u)}\,du - \int_{y_1}^{\hat{y}_0+t_n} \frac{1}{\hat{\lambda}(u)}\,du\bigg)\Bigg) \tag{4.18}$$

for an appropriate sequence $t_n \searrow 0$. Similarly, an estimator

$$\hat{\lambda}_2 = -\exp\Bigg(\hat{B}\bigg(\int_{y_2}^{\hat{y}_0-t_n} \frac{1}{\hat{\lambda}(u)}\,du - \int_{y_1}^{\hat{y}_0+t_n} \frac{1}{\hat{\lambda}(u)}\,du\bigg)\Bigg) \tag{4.19}$$

is obtained, when estimating B by $\hat{B}$ as in (4.15).

A Global Estimator

Having a look at equation (3.17) again, note that estimators for all of its components have been provided in the previous sections. Hence, these can be used to define an estimator of the transformation function h that can be applied globally for all $y \in \mathbb{R}$. Because h is continuous in its root y_0, one has

$$B\int_{y_1}^{y} \frac{1}{\lambda(u)}\,du \overset{y\searrow y_0}{\to} \infty \quad \text{and} \quad B\int_{y_2}^{y} \frac{1}{\lambda(u)}\,du \overset{y\nearrow y_0}{\to} \infty.$$

Therefore, to estimate h in a neighbourhood of y_0, it might not be a good idea to do so by estimating B and the integrals directly. To motivate the estimators in (4.20) and (4.21) below, one can write for an appropriate sequence $y_n \searrow y_0$ (e.g. $y_n = y_0 + t_n$ with t_n as above)

$$\begin{aligned}
h(y) &= \exp\bigg(-B\int_{y_1}^{y} \frac{1}{\lambda(u)}\,du\bigg)\\
&= \exp\bigg(-B\int_{y_n}^{y} \frac{1}{\lambda(u)-\lambda(y_0)}\,du - B\int_{y_1}^{y_n} \frac{1}{\lambda(u)}\,du\bigg)\\
&= \exp\bigg(-B\int_{y_n}^{y} \frac{1}{\frac{\partial}{\partial y}\lambda(y)\big|_{y=y_0}(u-y_0)+o(u-y_0)}\,du\bigg)h(y_n)\\
&\approx \exp\bigg(\underbrace{-\frac{B}{\frac{\partial}{\partial y}\lambda(y)\big|_{y=y_0}}}_{=1}\int_{y_n}^{y} \frac{1}{u-y_0}\,du\bigg)h(y_n)\\
&= \exp\big(\log(y-y_0)-\log(y_n-y_0)\big)h(y_n)\\
&= \frac{y-y_0}{y_n-y_0}h(y_n),
\end{aligned}$$

so that it makes sense to estimate h in a neighbourhood of y_0 by a linearisation of $\hat{h}$.
For some sequence $t_n \searrow 0$ the resulting estimator on $\mathbb{R}$ can be defined by

$$\hat{h}(y) = \begin{cases} \exp\Big(-\hat{B}\int_{y_1}^{y}\frac{1}{\hat{\lambda}(u)}\,du\Big), & y \geq \hat{y}_0 + t_n, \\ \frac{y-\hat{y}_0}{t_n}\hat{h}(\hat{y}_0+t_n), & y \in (\hat{y}_0, \hat{y}_0+t_n), \\ 0, & y = \hat{y}_0, \\ \frac{\hat{y}_0-y}{t_n}\hat{h}(\hat{y}_0-t_n), & y \in (\hat{y}_0-t_n, \hat{y}_0), \\ \hat{\lambda}_2\exp\Big(-\hat{B}\int_{y_2}^{y}\frac{1}{\hat{\lambda}(u)}\,du\Big), & y \leq \hat{y}_0 - t_n. \end{cases} \tag{4.20}$$

If $(t_n)_{n\in\mathbb{N}}$ is chosen appropriately, this estimator is uniformly consistent on compact sets $\mathcal{K} \subseteq \mathbb{R}$ (details can be found in part 4.2.1). Again, a similar estimator

$$\tilde{h}(y) = \begin{cases} \exp\Big(-\tilde{B}\int_{y_1}^{y}\frac{1}{\hat{\lambda}(u)}\,du\Big), & y \geq \hat{y}_0 + t_n, \\ \frac{y-\hat{y}_0}{t_n}\tilde{h}(\hat{y}_0+t_n), & y \in (\hat{y}_0, \hat{y}_0+t_n), \\ 0, & y = \hat{y}_0, \\ \frac{\hat{y}_0-y}{t_n}\tilde{h}(\hat{y}_0-t_n), & y \in (\hat{y}_0-t_n, \hat{y}_0), \\ \tilde{\lambda}_2\exp\Big(-\tilde{B}\int_{y_2}^{y}\frac{1}{\hat{\lambda}(u)}\,du\Big), & y \leq \hat{y}_0 - t_n. \end{cases} \tag{4.21}$$

is obtained when estimating B and λ_2 as in (4.3) and (4.18).

4.2 Asymptotic Behaviour of the Estimator

Here, the asymptotic behaviour of the estimators presented in Section 4.1 is examined. First, the main results on estimating h_1 (defined as in (4.8)), B and h are given. Afterwards, some minor adjustments taking care of the estimation of y_0 and λ_2 as well as some technical part preparing the proof of the main results are presented. The assumptions can be found in Section 4.5.

4.2.1 Asymptotic Behaviour of the Estimated Transformation Function

In this part, the main results regarding the convergence rates of the estimator of h provided in the previous Section 4.1 are given. These are separated into Theorem 4.2.2 and Theorem 4.2.4 considering on the one hand the estimation of the integral in (4.8) and on the other hand the estimation of B. In Theorem 4.2.2, uniform convergence on compact sets $\mathcal{K} \subseteq (y_0, \infty)$ is shown.

Due to the plug in type approach, first an expression for $\hat{\lambda} - \lambda$ is proven. This expression is used in a second step to prove some convergence results on the integral over λ and to connect this later via (3.17) to the estimation of h. To handle $\hat{\lambda}(u) - \lambda(u)$, one can proceed similarly to Chiappori et al. (2015) and define with the notations from Section 1.1 (e.g. see (1.3)–(1.4)) and $f = f_X$

$$D_{p,0} = -\frac{f_x}{\Phi_y f^2}, \qquad D_{p,y} = -\frac{\Phi_x}{\Phi_y^2 f}, \qquad D_{p,x} = \frac{1}{\Phi_y f},$$

$$D_{f,0} = \frac{2pf_x}{\Phi_y f^3} - \frac{p_x}{\Phi_y f^2} + \frac{p_y\Phi_x}{\Phi_y^2 f^2}, \qquad D_{f,x} = -\frac{p}{\Phi_y f^2}$$

and corresponding estimators of Φ, Φ_x and Φ_y with (1.2), (1.5)–(1.7). Here and in the following, the convention $f = f_X$ is used to denote the density of X.

Lemma 4.2.1 *Assume **(A1)**–**(A7)** and **(B1)**–**(B5)** and let $\mathcal{K} \subseteq \mathbb{R}$ be compact. Then, with*

$$\lambda(y|x) = \frac{\frac{\partial F_{Y|X}(y|x)}{\partial x_1}}{\frac{\partial F_{Y|X}(y|x)}{\partial y}} \quad \text{and} \quad \hat{\lambda}(y|x) = \frac{\frac{\partial \hat{F}_{Y|X}(y|x)}{\partial x_1}}{\frac{\partial \hat{F}_{Y|X}(y|x)}{\partial y}}$$

one has

$$\begin{aligned}\hat{\lambda}(y|x) - \lambda(y|x) = {} & (\hat{p}(y,x) - p(y,x))D_{p,0}(y,x) + (\hat{p}_x(y,x) - p_x(y,x))D_{p,x}(y,x) \\ & + (\hat{p}_y(y,x) - p_y(y,x))D_{p,y}(y,x) + (\hat{f}(x) - f(x))D_{f,0}(y,x) \\ & + (\hat{f}_x(x) - f_x(x))D_{f,x}(y,x) + o_p\left(\frac{1}{\sqrt{n}}\right)\end{aligned}$$

uniformly in $x \in \operatorname{supp}(v)$ and $y \in \mathcal{K}$. Moreover, with $\hat{\lambda}(y) = \int v(x)\hat{\lambda}(y|x)\,dx$ from (4.1) it holds that

$$\begin{aligned}\hat{\lambda}(u) - \lambda(u) = {} & \frac{1}{n}\sum_{i=1}^{n}\bigg(v(X_i)D_{p,0}(u,X_i)\mathcal{K}_{h_y}(u - Y_i) - \frac{\partial\big(v(X_i)D_{p,x}(u,X_i)\big)}{\partial x_1}\mathcal{K}_{h_y}(u - Y_i) \\ & + v(X_i)D_{p,y}(u,X_i)K_{h_y}(u - Y_i) + v(X_i)D_{f,0}(u,X_i) - \frac{\partial\big(v(X_i)D_{f,x}(u,X_i)\big)}{\partial x_1}\bigg) \\ & + o_p\left(\frac{1}{\sqrt{n}}\right) \\ = {} & \mathcal{O}_p\left(\sqrt{\frac{\log(n)}{nh_y}}\right)\end{aligned} \tag{4.22}$$

uniformly in $y \in \mathcal{K}$.

The proof can be found in Section 4.6.1. Lemma 4.2.1 gives an expression which makes it possible to extend (or even improve) the convergence rate of $\hat{\lambda} - \lambda$ to the one of the integral. Again, similar techniques as in the paper of Chiappori et al. (2015) are applied.

Theorem 4.2.2 *Assume **(A1)**–**(A7)** and **(B1)**–**(B5)**. Then, for all compact intervals $[u_1, u_2] \subseteq (y_0, \infty)$ the process $(Z_n(y))_{y \in [u_1, u_2]}$ defined by*

$$Z_n(y) := \sqrt{n}\int_{y_1}^{y}\left(\frac{1}{\hat{\lambda}(u)} - \frac{1}{\lambda(u)}\right)du$$

converges weakly to a centred Gaussian process Z_λ with covariance function κ_Z which can be found in the proof in Section 4.6.2.

Asymptotic Behaviour of $\tilde{B}$ and $\hat{B}$

Before stating Theorem 4.2.4 some further notations and assumptions are needed. Recall that under the mean square distance from independence approach in 4.1.2 the estimator $\hat{B}$ was defined as the minimizer with respect to c of

$$\hat{A}(c, \hat{h}_1, \hat{F}_{Y|X}^{-1}(\tau|X), \hat{F}_{Y|X}^{-1}(\beta|X))$$

$$= \int_{M_X} \int_{[e_a,e_b]} \Big(\hat{P}\big(X \le x, \tilde{\varepsilon}_c(\hat{h}_1, \hat{F}^{-1}_{Y|X}(\tau|X), \hat{F}^{-1}_{Y|X}(\beta|X)) \le e | X \in M_X\big)$$

$$- \hat{P}\big(X \le x | X \in M_X\big) \hat{P}\big(\tilde{\varepsilon}_c(\hat{h}_1, \hat{F}^{-1}_{Y|X}(\tau|X), \hat{F}^{-1}_{Y|X}(\beta|X)) \le e | X \in M_X\big)\Big)^2 \, de \, dx.$$

For appropriate functions $h : \mathbb{R} \to \mathbb{R}, f_{m_\tau}, f_{m_\beta} : \mathbb{R}^{d_X} \to \mathbb{R}$ write again $s = (\mathfrak{h}, f_{m_\tau}, f_{m_\beta})$ and $s_0 = (h_1, F^{-1}_{Y|X}(\tau|\cdot), F^{-1}_{Y|X}(\beta|\cdot))^t$ with h_1 from (4.8). Denote the supremum norms of f_{m_τ}, f_{m_β} and h on M_X and $[z_a, z_b]$ by $||.||_{M_X}$ and $||.||_{[z_a,z_b]}$, respectively. Let $C > 0$ such that $\sup\limits_{u \in [z_a,z_b]} \big|\frac{\partial^2}{\partial u^2} h_1(u)\big| \le \frac{C}{2}$ and define the set of functions

$$\mathcal{H} = \Bigg\{ s = (\mathfrak{h}, f_{m_\tau}, f_{m_\beta})^t : \mathfrak{h} \in \mathcal{C}^2([z_a, z_b]), f_{m_\tau}, f_{m_\beta} \in \mathcal{C}^2(M_X), f_{m_\tau}(M_X) \subseteq (z_a, z_b),$$

$$f_{m_\beta}(M_X) \subseteq (z_a, z_b), \left|\frac{\partial^2}{\partial u^2}\mathfrak{h}(u)\right| \le C, 2 \inf_{u \in [z_a,z_b]} \frac{\partial}{\partial u}\mathfrak{h}(u) > \inf_{u \in [z_a,z_b]} \frac{\partial}{\partial u} h_1(u) \Bigg\} \qquad (4.23)$$

endowed with the supremum norm

$$||s||_{\mathcal{H}} = \max\big(||h||_{[z_a,z_b]}, ||f_{m_\tau}||_{M_X}, ||f_{m_\beta}||_{M_X}\big).$$

Following Section 2.7.1 of Van der Vaart and Wellner (1996), consider for some $\gamma, R > 0$ the (Hölder-)class C_R^γ of all functions on M_X such that all partial derivatives up to order $\lfloor\gamma\rfloor$ are uniformly bounded by R and the partial derivatives of highest order are Lipschitz of order $\gamma - \lfloor\gamma\rfloor$. More precisely, define for any multi-index $j = (j_1, ..., j_{d_X})$ the differential

$$D_j = \frac{\partial^j}{\partial x_1^{j_1} ... \partial x_d^{j_d}}$$

as well as the norm

$$||f||_\gamma = \max_{j \le \lfloor\gamma\rfloor} \sup_{x \in M_X} |D_j f(x)| + \max_{j = \lfloor\gamma\rfloor} \sup_{x \ne y \in M_X} \frac{|D_j f(x) - D_j f(y)|}{||x-y||^{\gamma - \lfloor\gamma\rfloor}},$$

where the inequality $j \le \lfloor\gamma\rfloor$ has to be read in the sense of $\sum_{i=1}^{d_X} j_i \le \lfloor\gamma\rfloor$ for every multi-index $j = (j_1, ..., j_{d_X})$. In the case of $d_X = 1$ the norm can be written as

$$||f||_\gamma = \max_{j=1,...,\lfloor\gamma\rfloor} \sup_{x \in M_X} \left|\frac{\partial^j}{\partial x^j} f(x)\right| + \sup_{x \ne y \in M_X} \frac{\big|\frac{\partial^{\lfloor\gamma\rfloor}}{\partial x^{\lfloor\gamma\rfloor}} f(x) - \frac{\partial^{\lfloor\gamma\rfloor}}{\partial x^{\lfloor\gamma\rfloor}} f(y)\big|}{||x-y||^{\gamma - \lfloor\gamma\rfloor}}.$$

Further, define for some $R > 0$ the set $C_R^\gamma(M_X)$ as the set of all (sufficiently often differentiable) functions f with $||f||_\gamma \le R$ and

$$\tilde{\mathcal{H}} = \Bigg\{ s \in \mathcal{H} : h \in C_{R_h}^{\gamma_h}([z_a, z_b]), f_{m_\tau} \in C_{R_{f_{m_\tau}}}^{\gamma_{f_{m_\tau}}}(M_X), f_{m_\beta} \in C_{R_{f_{m_\beta}}}^{\gamma_{f_{m_\beta}}}(M_X) \Bigg\} \qquad (4.24)$$

for some constants $\gamma_h > 1, \gamma_{f_{m_\tau}}, \gamma_{f_{m_\beta}} > d_X$ and $R_h, R_{f_{m_\tau}}, R_{f_{m_\beta}} < \infty$. For all $(x, e) \in M_X \times [e_a, e_b]$ let $\Gamma_1(c, s_0)(x, e)$ denote the ordinary derivative of $G_{MD}(c, s_0)(x, e)$ with respect to c and let $\Gamma_2(c, s_0)(x, e)[s - s_0]$ denote the directional derivative of $G_{MD}(c, s_0)(x, e)$ with respect to s, that is

$$\Gamma_2(c, s_0)(x, e)[s - s_0] := \lim_{t \to 0} \frac{G_{MD}(c, s_0 + t(s - s_0))(x, e) - G_{MD}(c, s_0)(x, e)}{t}.$$

Furthermore, let

$$D_h G_{MD}(c, s_0)(x, e)[\mathfrak{h} - h_1],$$

$$D_{f_{m_\tau}} G_{MD}(c, s_0)(x, e)\big[f_{m_\tau} - F^{-1}_{Y|X}(\tau|\cdot)\big],$$

$$D_{f_{m_\beta}} G_{MD}(c, s_0)(x, e)\big[f_{m_\beta} - F^{-1}_{Y|X}(\beta|\cdot)\big]$$

denote the directional derivatives with respect to $\mathfrak{h}, f_{m_\tau}$ and f_{m_β}, respectively. Now, some further properties can be formulated. These will allow to proceed in the proof of Theorem 4.2.4 below similarly to Linton et al. (2008). Most of them are already implied by assumptions (**A1**)-(**A7**),(**B1**)–(**B5**) (see Lemma 4.2.3 below). Let G_{MD} and G_{nMD} be defined as in (4.11) and (4.12) and let $\hat{s}$ be an estimator of s_0. In the proof of Theorem 4.2.4, it will be shown that the assumptions (**C1**)–(**C3**) and (**C5**) below are already implied by (**A1**)–(**A7**) in Section 3.4 and (**M1**)–(**M5**) in Section 4.1.2, while validity of (**C4**),(**C6**) and (**C6'**) for $\hat{s}$ as in (4.14) is treated in Lemma 4.2.3 below.
Let M_X and $[e_a, e_b]$ be defined as in the definition of $\hat{B}$ in Section 4.1.

(**C1**) One has $G_{MD}(B, s_0) \equiv 0$ and $\hat{B} - B = o_p(1)$.

(**C2**) For all $(x, e) \in M_X \times [e_a, e_b]$ the ordinary derivative (with respect to c) $\Gamma_1(c, s_0)(x, e)$ of $G_{MD}(c, s_0)(x, e)$ exists in a neighbourhood of B and is continuous at $c = B$. $\Gamma_1(B, s_0)(x, e)$ is different from zero on a set with positive $\lambda_{M_X \times [e_a, e_b]}$-measure.

(**C3**) The directional derivative $\Gamma_2(c, s_0)(x, e)[s - s_0]$ of $G_{MD}(c, s_0)(x, e)$ with respect to s exists for all $c \in B_\delta, (x, e) \in \lambda_{M_X \times [e_a, e_b]}$ and in all directions $[s - s_0]$ with $s \in \tilde{\mathcal{H}}$ and $\tilde{\mathcal{H}}$ as in (4.24). Moreover, for any $\delta > 0$ let B_δ be the δ-neighbourhood of B in $[B_1, B_2]$ and $\tilde{\mathcal{H}}_\delta = \{s \in \tilde{\mathcal{H}} : ||s - s_0||_{\mathcal{H}} < \delta\}$. Consider a positive sequence $\delta_n \to 0$ and $(c, s) \in B_{\delta_n} \times \tilde{\mathcal{H}}_{\delta_n}$. Then,

(i) for an appropriate constant $C \geq 0$ (independent of c and s) it holds that

$$||G_{MD}(c, s) - G_{MD}(c, s_0) - \Gamma_2(c, s_0)[s - s_0]||_2$$

$$\leq C\big(||h - h_1||^{\frac{3}{2}}_{[z_a, z_b]} + ||f_{m_\tau} - F^{-1}_{Y|X}(\tau|\cdot)||^2_{M_X} + ||f_{m_\beta} - F^{-1}_{Y|X}(\beta|\cdot)||^2_{M_X}\big).$$

(ii) one has $||\Gamma_2(c, s_0)[\hat{s} - s_0] - \Gamma_2(B, s_0)[\hat{s} - s_0]|| = o_p(|c - B|) + o_p\big(n^{-\frac{1}{2}}\big)$ uniformly in $c \in B_{\delta_n}$.

(**C4**) $\hat{s} = (\bar{h}_1, \hat{f}_{m_\tau}, \hat{f}_{m_\beta})$ is some estimator of s_0 with $\hat{s} \in \tilde{\mathcal{H}}$ with probability converging to one, $||\bar{h}_1 - h_1||^{\frac{3}{2}}_{[z_a, z_b]} = o_p(n^{-\frac{1}{2}})$ and

$$||\hat{f}_{m_\tau} - F^{-1}_{Y|X}(\tau|\cdot)||_{M_X}, ||\hat{f}_{m_\beta} - F^{-1}_{Y|X}(\beta|\cdot)||_{M_X} = o_p(n^{-\frac{1}{4}}).$$

(**C5**) $$\sup_{||c-B|| \leq \delta_n, ||s - s_0|| \leq \delta_n} ||G_{nMD}(c, s) - G_{MD}(c, s) - G_{nMD}(B, s_0)||_2 = o_p(n^{-\frac{1}{2}}).$$

(**C6**) For some $\sigma_A^2 > 0$, it holds that

$$\sqrt{n} \int_{M_X} \int_{[e_a, e_b]} \Gamma_1(B, s_0)(x, e)\big(G_{nMD}(B, s_0)(x, e) + \Gamma_2(B, s_0)(x, e)[\hat{s} - s_0]\big)\, de\, dx$$

$$\xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_A^2).$$

(**C6'**) There exist a function ψ_{Γ_2} with $E[\psi_{\Gamma_2}(Y,X)] = o\big(n^{-\frac{1}{2}}\big)$ and $E[\psi_{\Gamma_2}(Y,X)^2] \in (0,\infty)$ such that

$$\sqrt{n}\int_{M_X}\int_{[e_a,e_b]} \Gamma_1(B,s_0)(x,e)\Big(G_{nMD}(B,s_0)(x,e) + \Gamma_2(B,s_0)(x,e)[\hat{s}-s_0]\Big)\,de\,dx$$
$$-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{\Gamma_2}(Y_i,X_i)$$
$$= o_p(1). \tag{4.25}$$

Lemma 4.2.3 *Assume **(A1)**–**(A7)** from Section 3.4 and **(M1)**–**(M5)** for a compact interval $[z_a,z_b] \subseteq (y_0,\infty)$ as well as **(B1)**–**(B5)** and **(B3')**. If $d_X = 1$ and $\hat{s}$ is defined as in (4.14), one has $\hat{s} \in \tilde{\mathcal{H}}$ with probability converging to one for $\gamma_h = \gamma_{f_{m_\tau}} = \gamma_{f_{m_\beta}} = 2$ and some sufficiently large constants $R_h, R_{f_{m_\tau}}, R_{f_{m_\beta}} > 0$. The second part of **(C4)** as well as **(C6)** and **(C6')** are valid, but instead of (4.26) below one has*

$$||\Gamma_2(B,s_0)[\hat{s}-s_0]||_2 = \mathcal{O}_p\bigg(\frac{1}{\sqrt{nh_x}}\bigg),$$

and consequently

$$\hat{B} - B = o_p\bigg(\frac{1}{\sqrt{nh_x}}\bigg).$$

The proof is given in Section 4.6.4. For arbitrary d_X, the estimators (or the assumptions) have to be adjusted, such that on the one hand the differentiability and boundedness conditions, that ensure $\hat{s} \in \tilde{\mathcal{H}}$, are met and on the other hand the convergence assumption in (**C4**) remains valid. This issue is not considered further in this thesis.

Theorem 4.2.4 *Assume **(A1)**–**(A7)** from Section 3.4 and **(M1)**–**(M5)** for some compact interval $[z_a,z_b] \subseteq (y_0,\infty)$. Further, let $\hat{s}$ be an estimator of s_0 such that **(C4)** and **(C6)** are valid.*

(i) Let

$$||\Gamma_2(B,s_0)[\hat{s}-s_0]||_2 = \mathcal{O}_p\bigg(\frac{1}{\sqrt{n}}\bigg). \tag{4.26}$$

Then,

$$\sqrt{n}(\hat{B}-B) \xrightarrow{\mathcal{D}} Z_B,$$

where $Z_B \sim \mathcal{N}(0,\sigma_B^2)$ is a centred, normally distributed random variable with variance

$$\sigma_B^2 = \frac{\sigma_A^2}{||\Gamma_1(B,s_0)||_2^4}$$

*and $\sigma_A > 0$ from **(C6)**.*

(ii) If

$$||\Gamma_2(B,s_0)[\hat{s}-s_0]||_2 = \mathcal{O}_p(a_n) \tag{4.27}$$

holds for some sequence $a_n \searrow 0$ with $a_n^{-1} = o\big(\sqrt{n}\big)$, one has $\hat{B} - B = o_p(a_n)$.

The proof can be found in Section 4.6.3. Note that the convergence rate of $\hat{B}$ to B is strongly linked to that of the estimated conditional quantile function $\hat{F}^{-1}_{Y|X}$ as a component of $\hat{s}$.
Another approach presented in 4.1.2 consisted in estimating B via $\tilde{B}$ from (4.3).

Theorem 4.2.5 *Assume* ***(A1)***–***(A7)*** *from Section 3.4 and* ***(B1)***–***(B5)***,***(B3')*** *from Section 4.5. Then,*

$$\sqrt{nh_y^3}(\tilde{B} - B) \xrightarrow{\mathcal{D}} Z_{\tilde{B}} \tag{4.28}$$

for $Z_{\tilde{B}} \sim \mathcal{N}(0, \sigma^2_{\tilde{B}})$ *and*

$$\sigma^2_{\tilde{B}} = \bigg(\int \Big(\frac{\partial}{\partial z} K(z)\Big)^2 dz\bigg)\bigg(\int v(w)^2 D_{p,y}(y_0, w)^2 f_{Y,X}(y_0, w)\, dw\bigg).$$

The proof is given in Section 4.6.5.

Combining the Results

Now, everything is prepared to state the main convergence result for the estimator of the transformation function given in part 4.1.3. First, a convergence result on compact subsets of (y_0, ∞) will be provided. An extension to compact subsets of $\mathbb{R}$ is given later in Theorem 4.2.11.

Theorem 4.2.6 *If not specified further below, let* $\bar{s} = (\bar{h}_1, \bar{f}_{m_\tau}, \bar{f}_{m_\beta})$ *be some estimator of* $s_0 = (h_1, F^{-1}_{Y|X}(\tau|x), F^{-1}_{Y|X}(\beta|x))$ *and define an estimator of* h *on* (y_0, ∞) *as in (4.9) by* $\bar{h}(y) = \bar{h}_{\hat{B}}(y)$ *with* $\hat{B}$ *from (4.15).*

(i) Let $\bar{h}_1$ *fulfil*

$$\bar{h}_1(y) - h_1(y) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_h(Y_i, X_i, y) + o_p\Big(\frac{1}{\sqrt{n}}\Big), \tag{4.29}$$

which holds uniformly on compact sets $\mathcal{K} \subseteq (y_0, \infty)$*. Moreover, let the function class* $\{(v, x) \mapsto \psi_h(v, x, y) : y \in \mathcal{K}\}$ *be Donsker with respect to the distribution law of* (Y, X)*. Let* $\mathcal{K} \subseteq (y_0, \infty)$ *be compact. Then, under the assumptions in Theorem 4.2.4 (i) and* ***(C6')*** *for* $\hat{s} = \bar{s}$ *the stochastic process* $(H_n(y))_{y\in\mathcal{K}}$ *defined by*

$$H_n(y) := \sqrt{n}(\bar{h}(y) - h(y))$$

converges weakly on compact sets to a centred Gaussian process $(Z_h(y))_{y\in\mathcal{K}}$ *with covariance function*

$$\kappa_h(u, v) = h(u)h(v)E\bigg[\bigg(\frac{B\psi_h(Y_1, X_1, u)}{h_1(u)} + \log(h_1(u))\psi_{\Gamma_2}(Y_1, X_1)\bigg)$$
$$\bigg(\frac{B\psi_h(Y_1, X_1, v)}{h_1(v)} + \log(h_1(v))\psi_{\Gamma_2}(Y_1, X_1)\bigg)\bigg]$$

and ψ_{Γ_2} *as in* ***(C6')****. In particular, if* ***(B1)***–***(B5)*** *are valid and* s_0 *is estimated by* $\hat{s}$ *from (4.14), equation (4.29) is valid and it holds that*

$$\kappa_h(u, v) = h(u)h(v)E\bigg[\bigg(B\eta_1(u) + \int_{y_1}^{u} \frac{1}{\lambda(y)}\, dy\ \psi_{\Gamma_2}(Y_1, X_1)\bigg)$$

$$\left(B\eta_1(v) + \int_{y_1}^{v} \frac{1}{\lambda(y)}\,dy\ \psi_{\Gamma_2}(Y_1, X_1)\right)\Bigg]$$

with η_1 as in (4.36).

(ii) Under the assumptions of (ii) in Theorem 4.2.4 one has

$$\bar{h}(y) - h(y) = o_p(a_n)$$

uniformly on compact sets $\mathcal{K} \subseteq (y_0, \infty)$.

(iii) Let $\hat{s}$ be defined as in (4.14). For every compact set $\mathcal{K} \subseteq (y_0, \infty)$ and under the assumptions of Theorem 4.2.5, the process $(\tilde{H}_n(y))_{y\in\mathcal{K}}$ defined by $\tilde{H}_n(y) = \sqrt{nh_y^3}(\tilde{h}(y) - h(y))$ with $\tilde{h}$ as in (4.17) converges weakly to the centred Gaussian process

$$(Z_{\tilde{h}}(y))_{y\in\mathcal{K}} = \left(h(y)\int_{y_1}^{y} \frac{1}{\lambda(u)}\,du\ Z_{\tilde{B}}\right)_{y\in\mathcal{K}}$$

with $Z_{\tilde{B}}$ from Theorem 4.2.5.

The proof can be found in Section 4.6.6.

Remark 4.2.7 *Colling and Van Keilegom (2019) adjusted the estimator of Chiappori et al. (2015) by first transforming Y with the empirical distribution function $\hat{F}_Y$ as explained in Section 1.4. In the context of the heteroscedastic model presented here, such a pretransformation is conceivable as well. Although not done by the author, the same techniques as used in the paper of Colling and Van Keilegom (2019) should work to obtain a result similar to Theorem 4.2.2. Note that the estimation of $F_{Y|X}^{-1}$, if based on the original data, remains unaffected by this pretransformation*

Asymptotic Behaviour of the Estimators of y_0 and λ_2

Let $\hat{h}$ be defined as in (4.20). There, estimators for y_0 and λ_2 with an asymptotic behaviour which has not been examined yet occur. This will be the subject of the following passages, before these findings are used to derive a result concerning the asymptotic behaviour of $\hat{h}$. Assume h to be two times continuously differentiable.

Theorem 4.2.8 *(**A1**)–(**A7**) in Section 3.4 and (**B1**)–(**B5**) in Section 4.5. Let $\hat{y}_0$ be defined as in (4.2). Then,*

$$\sqrt{nh_y}(\hat{y}_0 - y_0) \overset{\mathcal{D}}{\to} \mathcal{N}(0, \sigma_{y_0}^2)$$

for

$$\sigma_{y_0}^2 = \frac{\int K(z)^2\,dz}{B^2} \int v(w)^2 D_{p,y}(y_0, w)^2 f_{Y,X}(y_0, w)\,dw.$$

The proof can be found in Section 4.6.7. To state a general asymptotic result for $\hat{h}$, the only thing missing is the asymptotic behaviour of $\hat{\lambda}_2$ and $\tilde{\lambda}_2$, respectively. Although based on the same idea, both of the estimators depend on the respective methods applied to estimate λ, y_0 and B. In the following, the convergence rates of $\tilde{B} - B$ and $\hat{B} - B$ are used to define suitable rates for the sequence $(t_n)_{n\in\mathbb{N}}$ in (4.21) and (4.20).

Theorem 4.2.9 *Assume **(A1)**–**(A7)** in Section 3.4 and **(B1)**–**(B5)** as well as **(B3')** in Section 4.5. Let $t_n \searrow 0$ be a sequence with $\frac{\log(n)}{t_n\sqrt{nh_y}} = o(1)$. Then,*

$$\tilde{h}(\hat{y}_0 + t_n) - h(y_0 + t_n) = \mathcal{O}_p\left(\frac{\log(t_n)t_n}{\sqrt{nh_y^3}} + \frac{\log(n)}{\sqrt{nh_y}}\right).$$

Further, for the estimator defined in (4.18) one has

$$\tilde{\lambda}_2 - \lambda_2 = \frac{\lambda_2 \frac{\partial^2}{\partial y^2}h(y_0)t_n}{\frac{\partial}{\partial y}h(y_0)} + \mathcal{O}_p\left(\frac{\log(t_n)}{\sqrt{nh_y^3}} + \frac{\log(n)}{t_n\sqrt{nh_y}}\right) + o_p(t_n). \tag{4.30}$$

For $t_n \sim \left(\frac{\log(n)^2}{nh_y}\right)^{\frac{1}{4}}$ the fastest convergence rate in (4.30) is reached and

$$\tilde{\lambda}_2 - \lambda_2 = \frac{\lambda_2 \frac{\partial^2}{\partial y^2}h(y_0)t_n}{\frac{\partial}{\partial y}h(y_0)} + o_p(t_n).$$

The proof is given in Section 4.6.8.

Remark 4.2.10 *If an estimator $\hat{B}$ with $\hat{B} - B = \mathcal{O}_p(a_n)$ for some sequence $(a_n)_{n\in\mathbb{N}}$ with $a_n \searrow 0$ is used to define $\hat{\lambda}_2$ as in (4.19), one can show similarly that*

$$\hat{h}(\hat{y}_0 + t_n) - h(y_0 + t_n) = \mathcal{O}_p\left(\log(t_n)t_n a_n + \frac{\log(n)}{\sqrt{nh_y}}\right)$$

and

$$\hat{\lambda}_2 - \lambda_2 = \frac{\lambda_2 \frac{\partial^2}{\partial y^2}h(y_0)t_n}{\frac{\partial}{\partial y}h(y_0)} + \mathcal{O}_p\left(\log(t_n)a_n + \frac{\log(n)}{t_n\sqrt{nh_y}}\right) + o_p(t_n).$$

Asymptotic Behaviour of the Global Estimator

By now, all ingredients have been presented that are necessary to state a uniform convergence result for $\hat{h}$ on compact sets.

Theorem 4.2.11 *Assume **(A1)**–**(A7)** in Section 3.4 and **(B1)**–**(B5)**,**(B3')** in Section 4.5. Let $t_n \searrow 0$ be a decreasing sequence.*

*(i) Assume **(M1)**–**(M5)**, (4.26) and (4.29). Let B and h be estimated as in (4.15) and (4.20) with $\hat{s}$ as in (4.14).*

(a) If $\mathcal{K} \subseteq (y_0, \infty)$ is compact, one has

$$\sup_{y\in\mathcal{K}} |\hat{h}(y) - h(y)| = \mathcal{O}_p\Big(\frac{1}{\sqrt{n}}\Big).$$

(b) If $\mathcal{K} \subseteq [y_0, \infty)$ is compact and $t_n \sim \frac{\log(n)}{\sqrt{nh_y}}$, one has

$$\sup_{y\in\mathcal{K}} |\hat{h}(y) - h(y)| = \mathcal{O}_p\Big(\frac{\log(n)}{\sqrt{nh_y}}\Big).$$

(ii) Let B and h be estimated as in (4.3) and (4.21). If $\mathcal{K} \subseteq \mathbb{R}$ is compact and $t_n \sim \left(\frac{\log(n)^2}{nh_y}\right)^{\frac{1}{4}}$, one has

$$\sup_{y\in\mathcal{K}} |\tilde{h}(y) - h(y)| = \mathcal{O}_p\left(\left(\frac{\log(n)^2}{nh_y}\right)^{\frac{1}{4}}\right).$$

The proof can be found in Section 4.6.9.

4.2.2 Uniform Convergence Rates for Kernel Estimators

This section deals with kernel estimators and their convergence rates. The notations are taken from Sections 1.1 and 4.2.1, especially recall the definitions (1.1)–(1.7). Instead of the general idea that was already captured in Section 1.1 more specific results needed to examine similar convergence rates for estimates of the conditional distribution and quantile function are given. Recall that all estimators proposed in Chapter 4 somehow depend on kernel estimators.

For this purpose, some ideas of Hansen (2008) are borrowed. To unify notations as consistently as possible, define

$$\hat{\Psi}(x) = \frac{1}{nh_x^{d_X}} \sum_{i=1}^n Z_i \mathbf{K}\left(\frac{x - X_i}{h_x}\right)$$

for some (here independent) random pairs $(Z_i, X_i), i = 1, ..., n$, some bandwidth sequence $h_x \searrow 0$, some product kernel $\mathbf{K}$ with corresponding kernel function K and $x \in \mathbb{R}^{d_X}$. In the proofs of the following results, it will be sometimes referred to K as the “kernel” of $\hat{\Psi}$ and to the dimension of the domain of K (here d_X) is as the “dimension” of $\hat{\Psi}$. These terminologies only influence the proofs of the following results and will be explained in detail there.

In the paper of Hansen (2008), the Z_i are replaced by random variables Y_i, which possibly depend on X_i. Here, Z_i will be replaced by $\mathcal{K}_{h_y}(y - Y_i), K_{h_y}(y - Y_i)$ or simply by 1. In contrast to Hansen (2008), not weakly dependent, but instead independent random pairs $(Z_i, X_i), i = 1, ..., n$, are considered. It was shown there, that under appropriate conditions

$$\hat{\Psi}(x) - E[\hat{\Psi}(x)] = \mathcal{O}_p(a_n) \tag{4.31}$$

uniformly on compact (even on appropriately growing) sets, where

$$a_n = \left(\frac{\log(n)}{nh_x^{d_X}}\right)^{\frac{1}{2}}.$$

Since the most general result is not needed and the proof later due to some minor assumptions will be sketched anyway, the details of how to obtain equation (4.31) are omitted here. Note that if assumption (**B3**) is valid, one has $a_n = o_p(n^{-\frac{1}{4}})$. This results in the following lemma.

Lemma 4.2.12 *Assume (**A1**)–(**A7**) and (**B1**)–(**B5**) and let $[z_a, z_b] \subseteq \mathbb{R}$ be an arbitrary, compact interval. Then, one has*

$$\sup_{y\in[z_a,z_b],x\in\mathbb{R}^{d_X}} v(x)|\hat{p}(y, x) - p(y, x)| = o_p\big(n^{-\frac{1}{4}}\big),$$

$$\sup_{y\in[z_a,z_b],x\in\mathbb{R}^{d_X}} v(x)|\hat{p}_x(y,x)-p_x(y,x)| = o_p\big(n^{-\frac{1}{4}}\big),$$

$$\sup_{y\in[z_a,z_b],x\in\mathbb{R}^{d_X}} v(x)|\hat{p}_y(y,x)-p_y(y,x)| = o_p\big(n^{-\frac{1}{4}}\big),$$

$$\sup_{x\in\mathbb{R}^{d_X}} v(x)|\hat{f}(x)-f(x)| = o_p\big(n^{-\frac{1}{4}}\big),$$

$$\sup_{x\in\mathbb{R}^{d_X}} v(x)|\hat{f}_x(x)-f_x(x)| = o_p\big(n^{-\frac{1}{4}}\big).$$

The proof can be found in Section 4.6.10.

Remark 4.2.13 *For all sequences $(c_n)_{n\in\mathbb{N}}$ with at most polynomial growth, the convergence results in Lemma 4.2.12 can be extended to $y\in[-c_n,c_n]$ similarly to the proof of Theorem 2 of Hansen (2008). Note that the constant q there can be chosen arbitrarily for independent data.*

In the proof of Theorem 4.2.2, these convergence results are used to rewrite the difference between the quotient of the conditional distribution function of Y conditioned on X and its empirical counterpart. The main tool for doing so is Lemma 1.1.2. Remember equation (1.8), that is,

$$\frac{\hat{a}}{\hat{b}}-\frac{a}{b}=\frac{1}{b}(\hat{a}-a)-\frac{a}{b^2}(\hat{b}-b)-\frac{\hat{b}-b}{\hat{b}b}\left(\hat{a}-a-\frac{a(\hat{b}-b)}{b}\right)$$

for arbitrary $a,b,\hat{a},\hat{b}\in\mathbb{R}, b,\hat{b}\neq 0$. Using the same techniques as in the proof of Lemma 4.2.12, one can show

Corollary 4.2.14 *Assume **(A1)**–**(A7)** and **(B1)**–**(B5)** and let $[z_a,z_b]\subseteq\mathbb{R}$ be an arbitrary compact interval. Then, one has*

$$\sup_{x\in\mathbb{R}^{d_X}} v(x)\left|\frac{\partial}{\partial x_1}\hat{f}_x(x)-\frac{\partial}{\partial x_1}f_x(x)\right| = o_p(1),$$

$$\sup_{u\in[z_a,z_b],x\in\mathbb{R}^{d_X}} v(x)\left|\frac{\partial}{\partial u}\hat{p}_x(u,x)-\frac{\partial}{\partial u}p_x(u,x)\right| = o_p(1),$$

$$\sup_{u\in[z_a,z_b],x\in\mathbb{R}^{d_X}} v(x)\left|\frac{\partial}{\partial u}\hat{p}_y(u,x)-\frac{\partial}{\partial u}p_y(u,x)\right| = o_p(1),$$

$$\sup_{u\in[z_a,z_b],x\in\mathbb{R}^{d_X}} v(x)\left|\frac{\partial}{\partial u}\hat{\Phi}_x(u,x)-\frac{\partial}{\partial u}\Phi_x(u,x)\right| = o_p(1),$$

$$\sup_{u\in[z_a,z_b],x\in\mathbb{R}^{d_X}} v(x)\left|\frac{\partial}{\partial u}\hat{\Phi}_y(u,x)-\frac{\partial}{\partial u}\Phi_y(u,x)\right| = o_p(1),$$

$$\sup_{u\in[z_a,z_b]}\left|\frac{\partial}{\partial u}\hat{\lambda}(u)-\frac{\partial}{\partial u}\lambda(u)\right| = o_p(1),$$

$$\sup_{u\in[z_a,z_b]}\left|\frac{\partial^2}{\partial u^2}\hat{\lambda}(u)-\frac{\partial^2}{\partial u^2}\lambda(u)\right| = o_p(1).$$

Further, the assertion still holds if the weighting function v is omitted as long as x is restricted to belong to a compact set.

The proof can be found in Section 4.6.11. Once a convergence result is proven for an estimator of the conditional distribution function, it can be extended relatively easily to the corresponding estimator of the conditional quantile function

$$\hat{F}_{Y|X}^{-1}(\tau|x) = \inf\{u \in \mathbb{R} : \hat{F}_{Y|X}(u|x) \geq \tau\}$$

as is done in the following Lemma.

Lemma 4.2.15 *Assume **(A1)**–**(A7)** and **(B1)**–**(B5)** and let* $[\tau_a, \tau_b] \subseteq (0,1), [z_a, z_b] \subseteq \mathbb{R}$ *be compact intervals. Then, one has*

$$\sup_{y\in[z_a,z_b],x\in\mathbb{R}^{d_X}} v(x)|\hat{F}_{Y|X}(y|x) - F_{Y|X}(y|x)| = o_p\big(n^{-\frac{1}{4}}\big),$$

$$\sup_{y\in[\tau_a,\tau_b],x\in\mathbb{R}^{d_X}} v(x)|\hat{F}_{Y|X}^{-1}(\tau|x) - F_{Y|X}^{-1}(\tau|x)| = o_p\big(n^{-\frac{1}{4}}\big).$$

The proof can be found in Section 4.6.12.

4.3 Simulations

In this section, the behaviour of the estimators for y_0, B and h given in (4.2), (4.3) and (4.17), respectively, for finite sample sizes is examined. For this purpose, observations of independent real valued random variables $X \sim \mathcal{U}([0,1])$ and $\varepsilon \sim \mathcal{U}([-1,1])$ are generated. Afterwards, Y is defined by

$$Y = \frac{\big(1 + X + \frac{(1+X)^2}{2}\varepsilon\big)^3}{8} + \frac{7\big(1 + X + \frac{(1+X)^2}{2}\varepsilon\big)}{8},$$

that is, model (3.1) is fulfilled with

$$h^{-1}(y) = \frac{y^3}{8} + \frac{7y}{8}, \quad g(x) = 1 + x \quad \text{and} \quad \sigma(x) = \frac{(1+x)^2}{2}.$$

The transformation function h is chosen such that it is strictly monotonic. Furthermore, it fulfils the identification conditions $h(0) = 0$ and $h(1) = 1$ and thus needs to be linearly transformed later when comparing it to the estimator $\tilde{h}$. Note that ε does not fulfil assumption (**A3**). Nevertheless, since the equation (3.3) is based on the idea that the factor $f_\varepsilon\big(\frac{h(y)-g(x)}{\sigma(x)}\big)$ in (3.2) cancels out when dividing $\frac{\partial F_{Y|X}(y|x)}{\partial x_i}$ by $\frac{\partial F_{Y|X}(y|x)}{\partial y}$, it is tried to keep f_ε constant.

The simulations are conducted with the language *R* (R Core Team (2017)). Some of the already implemented commands such as *integrate* and *h.select* are applied and an interface for C++ is used to reduce the computation time. The weighting function v is chosen to be the indicator function of $[0,1]$. Instead of integrating

$$x \mapsto v(x)\frac{\frac{\partial \hat{F}_{Y|X}(y|x)}{\partial x_1}}{\frac{\partial \hat{F}_{Y|X}(y|x)}{\partial y}}$$

as in (4.1), the mean of $N_x = 100$ evaluations of the integrand at equidistant points between the minimum and the maximum of the observations of X was taken. This is similar to

what Colling and Van Keilegom (2019) did and to what will be done later in Section 5.5.1 to estimate the transformation function in a homoscedastic model nonparametrically. To calculate the bandwidths h_y and h_x, cross validation and the normal reference rule, respectively, have been applied (Silverman (1986)). The kernel K is chosen to be the Epanechnikov kernel.
According to equation (3.7), B can be expressed as

$$B = \int_0^1 \frac{2}{1+x}\,dx = \log(4) \approx 1.39.$$

Solving for $\lambda(y) = 0$ leads to the equation $1 = \log(4)h(y)$ and thus to

$$y_0 = h^{-1}\left(\frac{1}{\log(4)}\right) = \frac{1}{8\log(4)^3} + \frac{7}{8\log(4)} \approx 0.68.$$

Observations are simulated for sample sizes of $n \in \{100, 200, 500, 1000, 2000, 5000, 10000\}$. For computational reasons, the number m of simulation runs for each of the scenarios decreases with the sample size and can be found in Table 4.1. Figure 4.1 shows a realization

Sample Size	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 2000$	$n = 5000$	$n = 10000$
Number of Sim. Runs	$m = 500$	$m = 500$	$m = 200$	$m = 200$	$m = 100$	$m = 50$	$m = 20$

Table 4.1: The sample sizes and the corresponding number of the simulation runs.

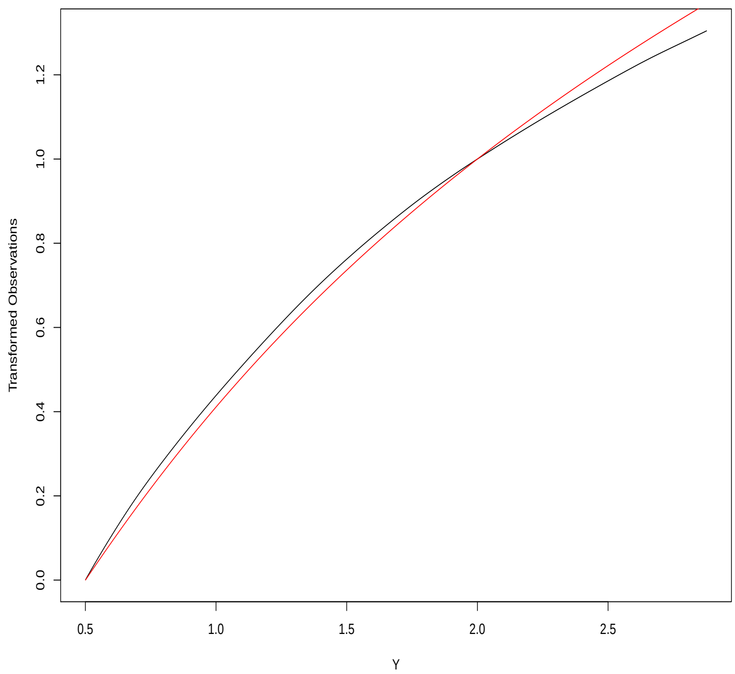

Figure 4.1: One realization of the estimated transformation function (black curve) and the true transformation function (red curve) are shown for $n = 500$.

of the estimator $\tilde{h}$ in (4.17) which is based on $n = 500$ observations in black and the true transformation function h in red, both for $y > \hat{y}_0$. The estimator $\hat{\lambda}_2$ of λ_2 has not been simulated. Here and in the following, the scale constraint in (3.9) for $y_1 = 2$ and $\lambda_1 = 1$ is used, that is $\tilde{h}(2) = 1 = h(2)$. Moreover, to compare the estimator to the true transformation function, h is linearly transformed to fulfil $h(\hat{y}_0) = 0$ (compare to (3.12)). Therefore, both functions have to intersect at least in $\hat{y}_0$ and $y_1 = 2$. The approximation in Figure 4.1 seems to be quite good, although the estimator for values below $y_1 = 2$ slightly overestimates the true transformation function, whereas the opposite holds for values above $y_1 = 2$. As can be seen in Graphic 4.2, this phenomenon carries over to all of the simulated scenarios. There, the difference $\tilde{h} - h$ of the estimator and the true transformation function,

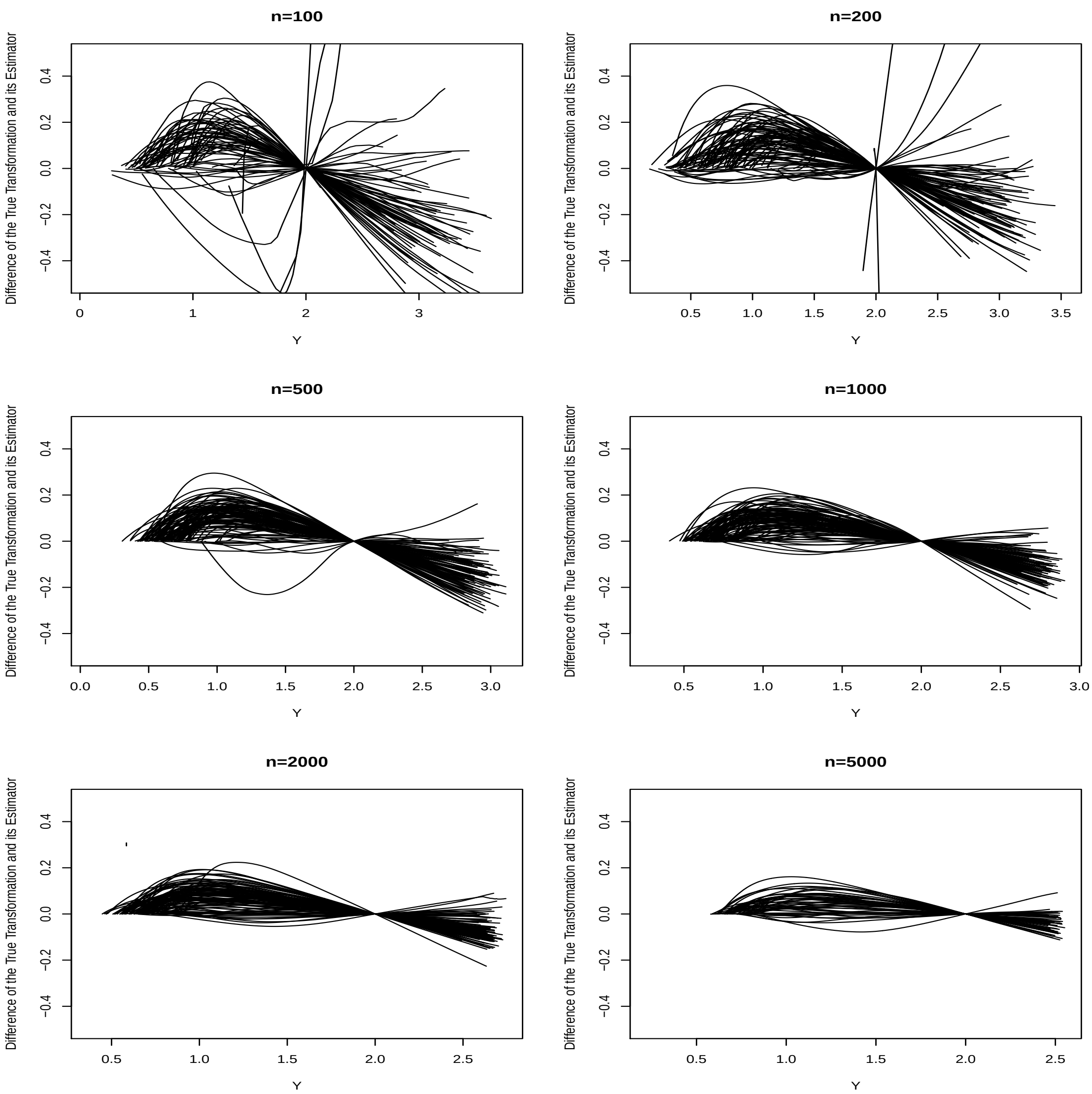

Figure 4.2: The difference of the true transformation function and its estimator under the same identification conditions is shown for the sample sizes of $n = 100, n = 200, n = 500, n = 1000, n = 2000, n = 5000, n = 10000$.

again based on the same identification conditions, for different sample sizes is shown. Up to a sample size of $n = 2000$, $m = 100$ curves are displayed, whereas for $n = 5000$ only $m = 50$ curves are shown. As expected, the difference decreases with a growing sample size. Since $h(y) \in (0, 1)$ for all $y \in (y_0, 2)$ and $h(y) > 1$ for all $y > 2$, the phenomenon of overestimating

h for values below $y_1 = 2$ and underestimating h for values above $y_1 = 2$ might indicate an underestimation of B. This theory is supported by the data listed in Table 4.2. Whereas

Sample Size	Mean of $\hat{y}_0$	Mean of $\tilde{B}$	Est. MISE
$n = 100$	1.14	0.80	33.19
$n = 200$	0.88	0.76	12.90
$n = 500$	0.64	0.81	2.38
$n = 1000$	0.65	0.85	2.25
$n = 2000$	0.66	0.99	2.19
$n = 5000$	0.71	1.10	2.30
$n = 10000$	0.66	1.16	1.92
True Values	0.68	1.39	

Table 4.2: Means of the estimators $\hat{y}_0$ and $\tilde{B}$ as well as the estimated MISE of the estimated transformation function for the sample sizes of $n = 100, n = 200, n = 500, n = 1000, n = 2000, n = 5000, n = 10000$.

$\hat{y}_0$ already seems to be unbiased for $n = 500$, the value of $\tilde{B}$ is even for $n = 10000$ below the true value of $B = 1.39$, although the gap between $\tilde{B}$ and B decreases with a growing sample size. The reason for this might be the estimation of B via the derivative of λ in y_0, since

$$\hat{\lambda}(y) = \int v(x) \frac{\frac{\partial \hat{F}_{Y|X}(y|x)}{\partial x_1}}{\frac{\partial \hat{F}_{Y|X}(y|x)}{\partial y}} \, dx$$

is already based on derivatives. Consequently, estimating B by $\tilde{B}$ leads to the issue of estimating second order derivatives. In this context, kernel estimators sometimes perform rather poorly.

Finally, some QQ-plots for $\hat{y}_0$ and $\tilde{B}$ are given in Figure 4.3. There, the empirical quantiles of the estimators are compared to those of standard normally distributed random variables. While the distribution of $\hat{y}_0$ seems to be almost normal already for a sample size of $n = 500$, the corresponding curve for $\tilde{B}$ has a small bump for $n = 500$, but at least seems to be linear for $n = 5000$. This indicates a slower convergence to the asymptotic distribution than for $\hat{y}_0$. One possibility to overcome especially the bias problem might consist in the usage of $\hat{B}$ instead of $\tilde{B}$, but this is not explored further in this thesis.

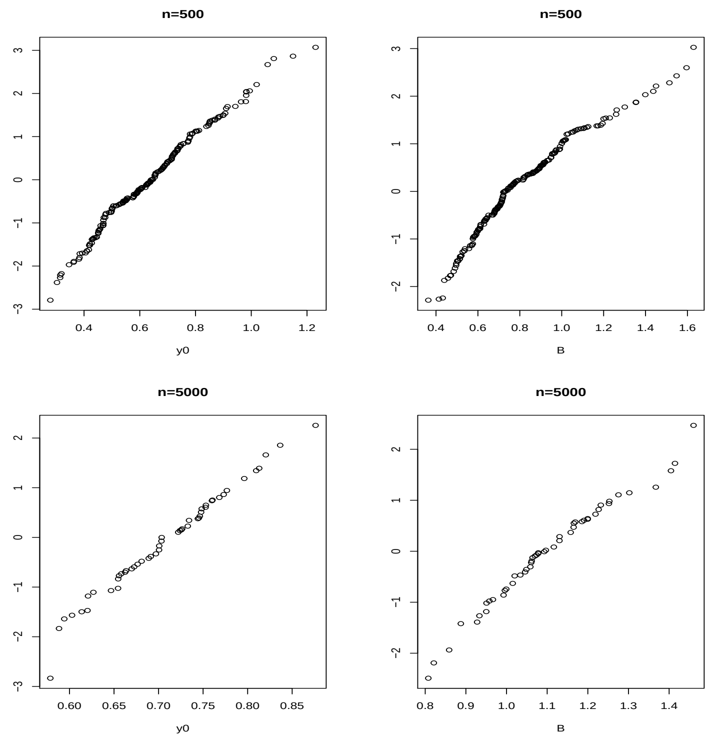

Figure 4.3: Normal-QQ-Plots of the estimators $\hat{y}_0$ and $\tilde{B}$ for the sample sizes of $n = 500$ and $n = 5000$.

4.4 Discussion

The so far most general approach for estimating the transformation function in the heteroscedastic model (3.1) has been developed. Depending on the chosen approach for estimating B, two estimators $\hat{h}$ and $\tilde{h}$ related to that of Chiappori et al. (2015) have been provided and the advantages and disadvantages of both have been discussed briefly. Consistency results for the proposed estimators and its components have been provided. A weak convergence result of the stochastic process $(\sqrt{n}(\hat{h}(y) - h(y)))_{y \in \mathcal{K}}$ on compact sets to a centred Gaussian process has been given.
Since the procedure is quite sophisticated, future research could consist in simplifying the estimation of at least some components of $\hat{h}$ or $\tilde{h}$. Moreover, recall that the estimator is based on equation (3.3), which in turn is based on the fact that

$$\frac{\partial F_{Y|X}(y|x)}{\partial y} = f_\varepsilon\left(\frac{h(y) - g(x)}{\sigma(x)}\right)\frac{h'(y)}{\sigma(x)} > 0.$$

From a theoretical point of view, the part depending on the density f_ε does not influence the estimator and cancels out when $\frac{\partial F_{Y|X}(y|x)}{\partial x_1}$ is divided by $\frac{\partial F_{Y|X}(y|x)}{\partial y}$. Nevertheless, the estimated values of $\frac{\partial F_{Y|X}(y|x)}{\partial x_1}$ and $\frac{\partial F_{Y|X}(y|x)}{\partial y}$ might be extremely small in applications, which might lead to numerical problems. Additionally, an examination of the behaviour of $\hat{B}$ for finite sample sizes would be worthwhile, since the usage of $\tilde{B}$ seems to be accompanied with a bias.

4.5 Assumptions

In the following, the assumptions of this chapter are listed. Additionally, assume (**A1**)–(**A7**) from Chapter 3.
Let $m \in \mathbb{N}$ and let v be a weight function with a compact support.

(**B1**) Let $(Y, X), (Y_1, X_1), ..., (Y_n, X_n)$ be independent and identically distributed observations from model (3.1). Let the density $f_{Y,X}$ of the joint distribution of (Y, X) be $(m+1)$-times continuously differentiable. Assume $f_{Y,X}$ to be bounded and f_X to be bounded away from zero on the support of v.

(**B2**) Let K be a continuously differentiable kernel of order m with compact support.

(**B3**) Let $\sqrt{n}h_y^m \to 0, \sqrt{n}h_x^m \to 0, \frac{\sqrt{n}h_y h_x^{d_X}}{\log(n)} \to \infty$ and $\frac{\sqrt{n}h_x^{d_X+2}}{\log(n)} \to \infty$.

(**B4**) Let v be $(m+1)$-times continuously differentiable.

(**B5**) Let there exist some $c > 0$ such that $\tilde{M}_{>c} = M_{>c} \cap \operatorname{supp}(v)$ or $\tilde{M}_{<-c} = M_{<-c} \cap \operatorname{supp}(v)$ fulfil assumption (**A7**).

In Section 4.1.2, the assumption (**B3'**) below, which is slightly stronger than assumption (**B3**), will be used.

(**B3'**) Let $\frac{nh_y^5 h_x}{\log(n)} \to \infty, \frac{nh_x^{d_X+4}}{\log(n)} \to \infty$ and $\frac{nh_x^3 h_y^3}{\log(n)} \to \infty$.

4.6 Proofs

This section contains the proofs of this chapter. The proofs are organized in a similar order as before, that is, the main results are proven first before the auxiliary assertions from Section 4.2.2 are considered.

4.6.1 Proof of Lemma 4.2.1

Proof: The proof follows the same line as the one of Lemma 1 of Chiappori et al. (2015). For reasons of clarity, the arguments of the occurring functions are omitted.
Recall equation (1.8):

$$\frac{\hat{a}}{\hat{b}} - \frac{a}{b} = \frac{1}{b}(\hat{a} - a) - \frac{a}{b^2}(\hat{b} - b) - \frac{\hat{b} - b}{\hat{b}b}\left(\hat{a} - a - \frac{a(\hat{b} - b)}{b}\right)$$

for arbitrary $a, b, \hat{a}, \hat{b} \in \mathbb{R}, b, \hat{b} \neq 0$. Since

$$\Phi_y = \frac{p_y}{f} \quad \text{and} \quad \Phi_x = \frac{p_x}{f} - \frac{p f_x}{f^2},$$

this results in

$$\frac{\hat{\Phi}_x}{\hat{\Phi}_y} - \frac{\Phi_x}{\Phi_y} = \frac{1}{\Phi_y}(\hat{\Phi}_x - \Phi_x) - \frac{\Phi_x}{\Phi_y^2}(\hat{\Phi}_y - \Phi_y) - \frac{\hat{\Phi}_y - \Phi_y}{\hat{\Phi}_y \Phi_y}\left(\hat{\Phi}_x - \Phi_x - \frac{\Phi_x(\hat{\Phi}_y - \Phi_y)}{\Phi_y}\right), \quad (4.32)$$

as well as

$$\hat{\Phi}_y - \Phi_y = \frac{1}{f}(\hat{p}_y - p_y) - \frac{p_y}{f^2}(\hat{f} - f) - \underbrace{\frac{\hat{f} - f}{\hat{f} f}\left(\hat{p}_y - p_y - \frac{p_y(\hat{f} - f)}{f}\right)}_{= o_p\left(\frac{1}{\sqrt{n}}\right)}. \tag{4.33}$$

Here, uniform convergence results like

$$\sup_{x \in \operatorname{supp}(v)} v(x)|\hat{f}(x) - f(x)| = o_p\big(n^{-\frac{1}{4}}\big)$$

or

$$\sup_{y \in [u_1, u_2],\, x \in \operatorname{supp}(v)} |\hat{p}_y(y, x) - p_y(y, x)| = o_p\big(n^{-\frac{1}{4}}\big)$$

are guaranteed by Lemma 4.2.12. To rewrite $\hat{\Phi}_x - \Phi_x$, note that

$$\begin{aligned}
\frac{\hat{p}\hat{f}_x}{\hat{f}^2} - \frac{p f_x}{f^2} &= \frac{1}{f^2}(\hat{p}\hat{f}_x - p f_x) - \frac{p f_x}{f^4}(\hat{f}^2 - f^2) + o_p\left(\frac{1}{\sqrt{n}}\right) \\
&= \frac{1}{f^2}((\hat{p} - p)\hat{f}_x + p(\hat{f}_x - f_x)) - \frac{p f_x}{f^4}(\hat{f} - f)(\hat{f} + f) + o_p\left(\frac{1}{\sqrt{n}}\right) \\
&= \frac{f_x}{f^2}(\hat{p} - p) + \frac{p}{f^2}(\hat{f}_x - f_x) - \frac{2p f_x}{f^3}(\hat{f} - f) + o_p\left(\frac{1}{\sqrt{n}}\right)
\end{aligned}$$

and therefore (again with Lemma 4.2.12),

$$\hat{\Phi}_x - \Phi_x = -\frac{f_x}{f^2}(\hat{p} - p) + \frac{1}{f}(\hat{p}_x - p_x) + \left(\frac{2p f_x}{f^3} - \frac{p_x}{f^2}\right)(\hat{f} - f) - \frac{p}{f^2}(\hat{f}_x - f_x) + o_p\left(\frac{1}{\sqrt{n}}\right). \tag{4.34}$$

Inserting this into equation (4.32) leads to the first assertion about the expression for $\hat{\lambda}(y|x) - \lambda(y|x)$. Remark that

$$\begin{aligned}
&D_{p,0} p + D_{p,x} p_x + D_{p,y} p_y + D_{f,0} f + D_{f,x} f_x \\
&\quad = -\frac{\Phi f_x}{\Phi_y f} + \frac{p_x}{\Phi_y f} - \frac{\Phi_x}{\Phi_y} + \frac{2\Phi f_x}{\Phi_y f} - \frac{p_x}{\Phi_y f} + \frac{\Phi_x}{\Phi_y} - \frac{\Phi f_x}{\Phi_y f} \\
&\quad = 0
\end{aligned}$$

and thus

$$\begin{aligned}
\hat{\lambda}(u) - \lambda(u) &= \int \big(\hat{\lambda}(u|x) - \lambda(u|x)\big) v(x)\, dx \\
&= \int \big(D_{p,0}(u,x)\hat{p}(u,x) + D_{p,x}(u,x)\hat{p}_x(u,x) + D_{p,y}(u,x)\hat{p}_y(u,x) \\
&\qquad + D_{f,0}(u,x)\hat{f}(x) + D_{f,x}(u,x)\hat{f}_x(x)\big) v(x)\, dx + o_p\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

Inserting the definition of $\hat{p}, \hat{p}_x, \hat{p}_y, \hat{f}, \hat{f}_x$ one obtains

$$\begin{aligned}
&\hat{\lambda}(u) - \lambda(u) \\
&= \frac{1}{n}\sum_{i=1}^n \int \Bigg(D_{p,0}(u,x)\mathcal{K}_{h_y}(u - Y_i)\mathbf{K}_{h_x}(x - X_i) + D_{p,x}(u,x)\mathcal{K}_{h_y}(u - Y_i)\frac{\partial \mathbf{K}_{h_x}(x - X_i)}{\partial x_1}
\end{aligned}$$

$$+ D_{p,y}(u,x)K_{h_y}(u-Y_i)\mathbf{K}_{h_x}(x-X_i) + D_{f,0}(u,x)\mathbf{K}_{h_x}(x-X_i)$$
$$+ D_{f,x}(u,x)\frac{\partial \mathbf{K}_{h_x}(x-X_i)}{\partial x_1}\Big)v(x)\,dx + o_p\Big(\frac{1}{\sqrt{n}}\Big).$$

Due to the assumptions (**B2**) and (**B3**), a Taylor expansion leads to

$$\int l(x)\mathbf{K}_{h_x}(x-X_i)\,dx = \int l(X_i+h_x x)\mathbf{K}(x)\,dx = l(X_i) + o\Big(\frac{1}{\sqrt{n}}\Big)$$

for every m-times continuously differentiable function l with bounded support. Moreover, integration by parts yields

$$\begin{aligned}
-\int l(x)\frac{\partial}{\partial x_1}\mathbf{K}_{h_x}(x-X_i)\,dx &= \int \mathbf{K}_{h_x}(x-X_i)\frac{\partial}{\partial x_1}l(x)\,dx \\
&= \int \mathbf{K}(x)\frac{\partial}{\partial x_1}l(x)\Big|_{x=X_i+h_x x}\,dx \\
&= \frac{\partial}{\partial x_1}l(x)\Big|_{x=X_i} + o\Big(\frac{1}{\sqrt{n}}\Big)
\end{aligned}$$

for every $(m+1)$-times continuously differentiable function l with bounded support. Due to the compactness of $\operatorname{supp}(v)$ and $[u_1,u_2]$ all derivatives of $D_{p,0},...,D_{f,x}$ are bounded, so that

$$\begin{aligned}
\hat\lambda(u)-\lambda(u) = \frac{1}{n}\sum_{i=1}^n \Big(& v(X_i)D_{p,0}(u,X_i)\mathcal{K}_{h_y}(u-Y_i) - \frac{\partial v(X_i)D_{p,x}(u,X_i)}{\partial x_1}\mathcal{K}_{h_y}(u-Y_i) \\
&+ v(X_i)D_{p,y}(u,X_i)K_{h_y}(u-Y_i) + v(X_i)D_{f,0}(u,X_i) - \frac{\partial v(X_i)D_{f,x}(u,X_i)}{\partial x_1}\Big) \\
&+ o_p\Big(\frac{1}{\sqrt{n}}\Big).
\end{aligned}$$

Finally,

$$\sup_{y\in\mathcal{K}}|\hat\lambda(u)-\lambda(u)| = \mathcal{O}_p\Bigg(\sqrt{\frac{\log(n)}{nh_y}}\Bigg)$$

follows as in the proof of Lemma 4.2.12 below. □

4.6.2 Proof of Theorem 4.2.2

The main idea of the proof is to find an expression

$$\int_{y_1}^{y}\Big(\frac{1}{\hat\lambda(u)}-\frac{1}{\lambda(u)}\Big)\,du = \frac{1}{n}\sum_{i=1}^n(\eta_i(y)-E[\eta_i(y)]) + o_p\Big(\frac{1}{\sqrt{n}}\Big) \tag{4.35}$$

(η_i will be defined later) for which some weak convergence results can be applied.
First, remark that since $[u_1,u_2]\subseteq(y_0,\infty)$ is compact, $u\mapsto\frac{1}{\lambda(u)}$ is bounded and bounded away from zero on $[u_1,u_2]$. Hence, one has

$$\int_{y_1}^{y}\Big(\frac{1}{\hat\lambda(u)}-\frac{1}{\lambda(u)}\Big)\,du = \int_{y_1}^{y}\frac{\lambda(u)-\hat\lambda(u)}{\hat\lambda(u)\lambda(u)}\,du$$

$$= \int_{y_1}^{y} \frac{\lambda(u) - \hat{\lambda}(u)}{\lambda(u)^2} \left(1 - \frac{\hat{\lambda}(u) - \lambda(u)}{\hat{\lambda}(u)}\right) du.$$

Possibly, extend $[u_1, u_2]$ such that y_1 is included (e.g. consider $[\min(y_1, u_1), \max(y_1, u_2)]$). Due to assumption (**B3**), Lemma 4.2.1 leads to

$$\begin{aligned} Z_n(u) &= \sqrt{n}\left(\int_{y_1}^{y} \frac{\lambda(u) - \hat{\lambda}(u)}{\lambda(u)^2}\, du + \mathcal{O}_p\Big(\sup_{u \in [u_1, u_2]} |\hat{\lambda}(u) - \lambda(u)|^2\Big)\right) \\ &\overset{(4.22)}{=} \sqrt{n}\left(\int_{y_1}^{y} \frac{\lambda(u) - \hat{\lambda}(u)}{\lambda(u)^2}\, du\right) + o_p(1). \end{aligned}$$

Moreover, Lemma 4.2.1 yields

$$\begin{aligned} \hat{\lambda}(u) - \lambda(u) = \frac{1}{n}\sum_{i=1}^{n} \Bigg(& v(X_i) D_{p,0}(u, X_i) \mathcal{K}_{h_y}(u - Y_i) - \frac{\partial v(X_i) D_{p,x}(u, X_i)}{\partial x_1} \mathcal{K}_{h_y}(u - Y_i) \\ & + v(X_i) D_{p,y}(u, X_i) K_{h_y}(u - Y_i) + v(X_i) D_{f,0}(u, X_i) - \frac{\partial v(X_i) D_{f,x}(u, X_i)}{\partial x_1}\Bigg) \\ & + o_p\left(\frac{1}{\sqrt{n}}\right) \end{aligned}$$

uniformly in $u \in [u_1, u_2]$, that is

$$\begin{aligned} \int_{y_1}^{y} \frac{\lambda(u) - \hat{\lambda}(u)}{\lambda(u)^2}\, du = \frac{1}{n}\sum_{i=1}^{n} \int_{y_1}^{y} \frac{-1}{\lambda(u)^2} \Bigg(& v(X_i) D_{p,0}(u, X_i) \mathcal{K}_{h_y}(u - Y_i) \\ & - \frac{\partial v(X_i) D_{p,x}(u, X_i)}{\partial x_1} \mathcal{K}_{h_y}(u - Y_i) + v(X_i) D_{p,y}(u, X_i) K_{h_y}(u - Y_i) \\ & + v(X_i) D_{f,0}(u, X_i) - \frac{\partial v(X_i) D_{f,x}(u, X_i)}{\partial x_1}\Bigg) du + o_p\left(\frac{1}{\sqrt{n}}\right) \\ =: \frac{1}{n}\sum_{i=1}^{n} \tilde{\eta}_i(y) + o_p\left(\frac{1}{\sqrt{n}}\right). \end{aligned}$$

The following lemma is similar to Proposition 2 of Colling and Van Keilegom (2019). Since the reasoning in the proof differs from that of Colling and Van Keilegom (2019), the proof is given as well.

Lemma 4.6.1 *Let $[u_1, u_2] \subseteq (y_0, \infty)$ be compact, $l : \mathbb{R} \times \mathbb{R}^{d_X} \to \mathbb{R}, (u, x) \mapsto l(u, x)$, be bounded on compact sets and let l have a compact support with respect to the x-component, which will be denoted by $\mathrm{supp}_x(l)$ in the following. Then, under the conditions of Theorem 4.2.2 one has*

$$\frac{1}{n}\sum_{i=1}^{n} \int_{y_1}^{y} l(u, X_i)\big(\mathcal{K}_{h_y}(u - Y_i) - I_{\{Y_i \le u\}}\big)\, du = o_p\left(\frac{1}{\sqrt{n}}\right)$$

uniformly in $y \in [u_1, u_2]$.

Proof: Define $q(x, z|y, a) = \int_{y_1}^{y} l(u, x) I_{\{u \ge z + a\}}\, du$. Then, q is bounded. One has

$$\frac{1}{n}\sum_{i=1}^{n} \int_{y_1}^{y} l(u, X_i)\big(\mathcal{K}_{h_y}(u - Y_i) - I_{\{Y_i \le u\}}\big)\, du$$

$$= \frac{1}{n}\sum_{i=1}^{n}\int_{y_1}^{y} l(u,X_i)\int \big(I_{\{Y_i\le u-th_y\}} - I_{\{Y_i\le u\}}\big)K(t)\,dt\,du$$
$$= \frac{1}{n}\sum_{i=1}^{n}\int \big(q(X_i,Y_i|y,th_y) - q(X_i,Y_i|y,0)\big)K(t)\,dt.$$

Let $\delta>0$ be such that $[u_1-\delta,u_2+\delta]\subseteq(y_0,\infty)$. Further, for all $y,\tilde{y}\in[u_1,u_2]$, $a,\tilde{a}\in[-\delta,\delta]$ and probability measures Q one has

$$\sqrt{E_Q\big[\big(q(X_1,Y_1|y,a)-q(X_1,Y_1|\tilde{y},\tilde{a})\big)^2\big]}$$
$$= \sqrt{E_Q\bigg[\bigg(\int_{y_1}^{\infty} l(u,X_1)\big(I_{\{y\ge u\ge Y_1+a\}} - I_{\{\tilde{y}\ge u\ge Y_1+\tilde{a}\}}\big)du\bigg)^2\bigg]}$$
$$\le \sup_{u\in[u_1,u_2],x\in\mathrm{supp}_x(l)} |l(u,x)|(|y-\tilde{y}|+|a-\tilde{a}|).$$

Hence, the covering numbers of the class

$$\mathcal{F} = \big\{(x,z)\mapsto q(x,z|y,a) : y\in[u_1,u_2], a\in[-\delta,\delta]\big\}$$

can be bounded by $\mathcal{N}(\varepsilon,\mathcal{F},L_2(Q))\le\frac{C}{\varepsilon^2}$ for an appropriate constant $C>0$ (independent of Q), so that by Theorem 2.5.2 of Van der Vaart and Wellner (1996) $\mathcal{F}$ is Donsker. Since

$$\sup_{y\in[u_1,u_2],t\in\mathrm{supp}(K)} E\big[\big(q(X_i,Y_i|y,th_y)-q(X_i,Y_i|y,0)\big)^2\big] = \mathcal{O}(h_y^2) = o(1)$$

Corollary 2.3.12 of Van der Vaart and Wellner (1996) leads to

$$\sup_{y\in[u_1,u_2],t\in\mathrm{supp}(K)} \bigg|\frac{1}{n}\sum_{i=1}^{n}\big(q(X_i,Y_i|y,th_y)-q(X_i,Y_i|y,0)$$
$$- E\big[q(X_i,Y_i|y,th_y)-q(X_i,Y_i|y,0)\big]\big)\bigg|$$
$$= o_p\bigg(\frac{1}{\sqrt{n}}\bigg).$$

The integrated expectation in turn can be bounded via a Taylor expansion

$$\sup_{y\in[u_1,u_2]}\bigg|E\bigg[\int \big(q(X,Y|y,th_y)-q(X,Y|y,0)\big)K(t)\,dt\bigg]\bigg|$$
$$= \sup_{y\in[u_1,u_2]}\bigg|\int E\bigg[\int_{y_1}^{y} l(u,X)(F_{Y|X}(u-th_y|X)-F_{Y|X}(u|X))\,du\bigg]K(t)\,dt\bigg|$$
$$= \mathcal{O}\big(h_y^m\big).$$

Now, the assertion is implied by assumption (**B3**). □

Define

$$\eta_i(y) := \int_{y_1}^{y}\frac{-1}{\lambda(u)^2}\bigg(v(X_i)D_{p,0}(u,X_i) - \frac{\partial v(X_i)D_{p,x}(u,X_i)}{\partial x_1}\bigg)I_{\{u\ge Y_i\}}\,du$$

$$- \frac{v(X_i)D_{p,y}(Y_i, X_i)}{\lambda(Y_i)^2}\big(I_{\{Y_i \le y\}} - I_{\{Y_i \le y_1\}}\big)$$

$$+ \int_{y_1}^{y} \frac{-1}{\lambda(u)^2}\left(v(X_i)D_{f,0}(u, X_i) - \frac{\partial v(X_i)D_{f,x}(u, X_i)}{\partial x_1}\right) du. \tag{4.36}$$

Then, Lemma 4.6.1 leads to

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} \tilde{\eta}_i(y) = \frac{1}{n}\sum_{i=1}^{n} \int_{y_1}^{y} \frac{-1}{\lambda(u)^2}\bigg(& v(X_i)D_{p,0}(u, X_i)I_{\{u \ge Y_i\}} - \frac{\partial v(X_i)D_{p,x}(u, X_i)}{\partial x_1} I_{\{u \ge Y_i\}} \\
&+ v(X_i)D_{p,y}(u, X_i)K_{h_y}(u - Y_i) + v(X_i)D_{f,0}(u, X_i) - \frac{\partial v(X_i)D_{f,x}(u, X_i)}{\partial x_1}\bigg) du \\
&+ o_p\left(\frac{1}{\sqrt{n}}\right) \\
= \frac{1}{n}\sum_{i=1}^{n} \eta_i(y) + o_p\left(\frac{1}{\sqrt{n}}\right),&
\end{aligned}$$

where

$$\begin{aligned}
&\int_{y_1}^{y} \frac{-1}{\lambda(u)^2} v(X_i)D_{p,y}(u, X_i)K_{h_y}(u - Y_i)\, du \\
&= -\int \frac{1}{\lambda(Y_i + h_y u)^2} v(X_i)D_{p,y}(Y_i + h_y u)K(u)\big(I_{\{Y_i \le y - h_y u\}} - I_{\{Y_i \le y_1 - h_y u\}}\big)\, du \\
&= -\frac{v(X_i)D_{p,y}(Y_i, X_i)}{\lambda(Y_i)^2}\big(I_{\{Y_i \le y\}} - I_{\{Y_i \le y_1\}}\big) + o_p\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

can be shown similarly to Lemma 4.6.1. If one is able to prove $E[\eta_i(y)] = o_p\left(\frac{1}{\sqrt{n}}\right)$ uniformly in $y \in [u_1, u_2]$ this would prove equation (4.35). Indeed, one has

$$\begin{aligned}
&E[\eta_i(y)] \\
&= \int_{y_1}^{y} \frac{-1}{\lambda(u)^2} \int \bigg(v(x)D_{p,0}(u, x)\int_{-\infty}^{u} f_{Y,X}(z, x)\, dz - \frac{\partial v(x)D_{p,x}(u, x)}{\partial x_1}\int_{-\infty}^{u} f_{Y,X}(z, x)\, dz \\
&\quad + v(x)D_{p,y}(u, x)f_{Y,X}(u, x) + v(x)D_{f,0}(u, x)f(x) - \frac{\partial v(x)D_{f,x}(u, x)}{\partial x_1} f(x)\bigg)\, dx\, du \\
&= \int_{y_1}^{y} \frac{-1}{\lambda(u)^2} \int \bigg(v(x)D_{p,0}(u, x)p(u, x) - \frac{\partial v(x)D_{p,x}(u, x)}{\partial x_1} p(u, x) \\
&\quad + v(x)D_{p,y}(u, x)p_y(u, x) + v(x)D_{f,0}(u, x)f(x) - \frac{\partial v(x)D_{f,x}(u, x)}{\partial x_1} f(x)\bigg)\, dx\, du \\
&= \int_{y_1}^{y} \frac{-1}{\lambda(u)^2} \int v(x)\bigg(D_{p,0}(u, x)p(u, x) + D_{p,x}(u, x)p_x(u, x) \\
&\quad + D_{p,y}(u, x)p_y(u, x) + D_{f,0}(u, x)f(x) + D_{f,x}(u, x)f_x(x)\bigg)\, dx\, du \\
&= 0.
\end{aligned} \tag{4.37}$$

So far, the asymptotic representation

$$\int_{y_1}^{y}\left(\frac{1}{\hat{\lambda}(u)}-\frac{1}{\lambda(u)}\right)du=\frac{1}{n}\sum_{i=1}^{n}(\eta_i(y)-E[\eta_i(y)])+o_p\left(\frac{1}{\sqrt{n}}\right)$$

was proven. It remains to show weak convergence of the corresponding process to an appropriate Gaussian process. For this purpose, define

$$\begin{aligned}\eta_{z,x}^a(y)&:=\int_{y_1}^{y}\frac{-1}{\lambda(u)^2}\left(v(x)D_{p,0}(u,x)-\frac{\partial v(x)D_{p,x}(u,x)}{\partial x_1}\right)I_{\{u\geq z\}}\,du\\&\qquad+\int_{y_1}^{y}\frac{-1}{\lambda(u)^2}\left(v(x)D_{f,0}(u,x)-\frac{\partial v(x)D_{f,x}(u,x)}{\partial x_1}\right)du,\\ \eta_{z,x}^b(y)&:=-\left(\frac{v(x)D_{p,y}(z,x)}{\lambda(z)^2}\right)_+\big(I_{\{z\leq y\}}-I_{\{z\leq y_1\}}\big),\\ \eta_{z,x}^c(y)&:=\left(\frac{v(x)D_{p,y}(z,x)}{\lambda(z)^2}\right)_-\big(I_{\{z\leq y\}}-I_{\{z\leq y_1\}}\big),\end{aligned}$$

where for some value $a\in\mathbb{R}$ the terms $(a)_+$ and $(a)_-$ denote the positive and negative part of a, respectively. Hence,

$$\eta_i(y)=\eta_{Y_i,X_i}^a(y)+\eta_{Y_i,X_i}^b(y)+\eta_{Y_i,X_i}^c(y).$$

It can be easily seen that $\eta_{z,x}^a(y),\eta_{z,x}^b(y)$ and $\eta_{z,x}^c(y)$ are bounded by some constant $\tilde{C}>0$ uniformly in $y,\tilde{y}\in[u_1,u_2]$. In the following, it will be proven, that the function classes

$$\mathcal{F}^j:=\big\{(z,x)\mapsto\eta_{z,x}^j(y),y\in[u_1,u_2]\big\},\quad j\in\{a,b,c\},$$

are Donsker. Example 2.10.7 of Van der Vaart and Wellner (1996) then implies that the class $\mathcal{F}=\{(z,x)\mapsto\eta_{z,x}(y),y\in[u_1,u_2]\}$ is Donsker as well. While the Donsker property of $\mathcal{F}^b$ and $\mathcal{F}^c$ can be shown by standard arguments as for indicator functions, one has

$$\begin{aligned}|\eta_{z,x}^a(y)-\eta_{z,x}^a(\tilde{y})|&=\bigg|\int_{\tilde{y}}^{y}\frac{-1}{\lambda(u)^2}\left(v(x)D_{p,0}(u,x)+\frac{\partial v(x)D_{p,x}(u,x)}{\partial x_1}\right)I_{\{u\geq z\}}\,du\\&\qquad+\int_{\tilde{y}}^{y}\frac{-1}{\lambda(u)^2}\left(v(x)D_{f,0}(u,x)+\frac{\partial v(x)D_{f,x}(u,x)}{\partial x_1}\right)du\bigg|\\&\leq C|y-\tilde{y}|\end{aligned}$$

for all $y,\tilde{y}\in[u_1,u_2]$ and an appropriate constant $C>0$, so that

$$\sqrt{E[(\eta_{Z_1,X_1}^a(y)-\eta_{Z_1,X_1}^a(\tilde{y}))^2]}\leq C|\tilde{y}-y|.$$

Let $\xi>0$. Then, ξ-brackets $[l,u]$ for the function class $\mathcal{F}^a$ can be defined as

$$l(z,x)=\eta_{z,x}^a(y_j^*)-\frac{\sqrt{\xi}}{C}\quad\text{and}\quad u(z,x)=\eta_{z,x}^a(y_j^*)+\frac{\sqrt{\xi}}{C},\quad j=1,...,K,$$

for some $K\in\mathbb{N}$ and appropriate values $y_1^*,...,y_K^*\in[u_1,u_2]$. Consequently, the bracketing number can be deduced from that of $[u_1,u_2]$ and for some constant C the bracketing integral

$$\int_0^\infty\sqrt{\log(\mathcal{N}_{[]}(\varepsilon,\mathcal{F},L_2(P^{Y,X})))}\,d\varepsilon=C\int_0^\infty\sqrt{\log\left(\max\left(\frac{1}{\varepsilon^2},1\right)\right)}\,d\varepsilon<\infty$$

is finite. Theorem 2.5.6 of Van der Vaart and Wellner (1996) ensures that $\mathcal{F}^a$ is Donsker, as long as the finite dimensional distributions converge, but this in turn (as for $\mathcal{F}^b, \mathcal{F}^c$ and $\mathcal{F}$) is implied by the multivariate Central Limit Theorem. It was already shown that $E[\eta_i(y)] = 0$ for $y \in [u_1, u_2]$. Let $\tilde{y}, y \in [u_1, u_2]$. After some rather technical computations for the indicator functions, the covariance function can be written as

$$
\begin{aligned}
&\kappa_Z(y, \tilde{y}) \\
&= E[\eta_1(y)\eta_1(\tilde{y})] \\
&= E\Bigg[\Bigg(\int_{y_1}^{y} \frac{-1}{\lambda(u)^2}\bigg(v(X_1)D_{p,0}(u, X_1) - \frac{\partial v(X_1)D_{p,x}(u, X_1)}{\partial x_1}\bigg) I_{\{u \geq Y_1\}}\, du \\
&\quad - \frac{v(X_1)D_{p,y}(Y_1, X_1)}{\lambda(Y_1)^2}\big(I_{\{Y_1 \leq y\}} - I_{\{Y_1 \leq y_1\}}\big) \\
&\quad + \int_{y_1}^{y} \frac{-1}{\lambda(u)^2}\bigg(v(X_1)D_{f,0}(u, X_1) - \frac{\partial v(X_1)D_{f,x}(u, X_1)}{\partial x_1}\bigg)\, du\Bigg) \\
&\quad \Bigg(\int_{y_1}^{\tilde{y}} \frac{-1}{\lambda(t)^2}\bigg(v(X_1)D_{p,0}(t, X_1) - \frac{\partial v(X_1)D_{p,x}(t, X_1)}{\partial x_1}\bigg) I_{\{t \geq Y_1\}}\, dt \\
&\quad - \frac{v(X_1)D_{p,y}(Y_1, X_1)}{\lambda(Y_1)^2}\big(I_{\{Y_1 \leq \tilde{y}\}} - I_{\{Y_1 \leq y_1\}}\big) \\
&\quad + \int_{y_1}^{\tilde{y}} \frac{-1}{\lambda(t)^2}\bigg(v(X_1)D_{f,0}(t, X_1) - \frac{\partial v(X_1)D_{f,x}(t, X_1)}{\partial x_1}\bigg)\, dt\Bigg)\Bigg] \\
&= \int_{y_1}^{y}\int_{y_1}^{\tilde{y}}\int \frac{1}{\lambda(u)^2\lambda(t)^2}\bigg(v(X_1)D_{p,0}(u, X_1) - \frac{\partial v(X_1)D_{p,x}(u, X_1)}{\partial x_1}\bigg) \\
&\quad \bigg(v(X_1)D_{p,0}(t, X_1) - \frac{\partial v(X_1)D_{p,x}(t, X_1)}{\partial x_1}\bigg) p(u \wedge t, x)\, dx\, dt\, du \\
&\quad + \int_{y_1 \wedge (y \vee \tilde{y})}^{y_1 \vee (y \wedge \tilde{y})} \int \bigg(\frac{v(x)D_{p,y}(z, x)}{\lambda(z)^2}\bigg)^2 f_{Y,X}(z, x)\, dx\, dz \\
&\quad + \int_{y_1}^{y}\int_{y_1}^{\tilde{y}}\int \frac{1}{\lambda(u)^2\lambda(t)^2}\bigg(v(x)D_{f,0}(u, x) - \frac{\partial v(x)D_{f,x}(u, x)}{\partial x_1}\bigg) \\
&\quad \bigg(v(x)D_{f,0}(t, x) - \frac{\partial v(x)D_{f,x}(t, x)}{\partial x_1}\bigg) f(x)\, dx\, dt\, du \\
&\quad + (1 - 2I_{\{y_1 \geq \tilde{y}\}}) \int_{y_1}^{y}\int_{u \wedge \tilde{y} \wedge y_1}^{u \wedge (\tilde{y} \vee y_1)}\int \frac{1}{\lambda(u)^2\lambda(z)^2}\bigg(v(x)D_{p,0}(u, x) - \frac{\partial v(x)D_{p,x}(u, x)}{\partial x_1}\bigg) \\
&\quad v(x)D_{p,y}(z, x) f_{Y,X}(z, x)\, dx\, dz\, du \\
&\quad + (1 - 2I_{\{y_1 \geq y\}}) \int_{y_1}^{\tilde{y}}\int_{u \wedge y \wedge y_1}^{u \wedge (y \vee y_1)}\int \frac{1}{\lambda(u)^2\lambda(z)^2}\bigg(v(x)D_{p,0}(u, x) - \frac{\partial v(x)D_{p,x}(u, x)}{\partial x_1}\bigg) \\
&\quad v(x)D_{p,y}(z, x) f_{Y,X}(z, x)\, dx\, dz\, du
\end{aligned}
$$

$$+\int_{y_1}^{y}\int_{y_1}^{\tilde{y}}\frac{1}{\lambda(u)^2\lambda(t)^2}\left(v(x)D_{p,0}(u,x)-\frac{\partial v(x)D_{p,x}(u,x)}{\partial x_1}\right)$$

$$\left(v(x)D_{f,0}(t,x)-\frac{\partial v(x)D_{f,x}(t,x)}{\partial x_1}\right)p(u,x)\,dx\,dt\,du$$

$$+\int_{y_1}^{\tilde{y}}\int_{y_1}^{y}\frac{1}{\lambda(u)^2\lambda(t)^2}\left(v(x)D_{p,0}(u,x)-\frac{\partial v(x)D_{p,x}(u,x)}{\partial x_1}\right)$$

$$\left(v(x)D_{f,0}(t,x)-\frac{\partial v(x)D_{f,x}(t,x)}{\partial x_1}\right)p(u,x)\,dx\,dt\,du$$

$$+(1-2I_{\{y_1\leq y\}})\int_{y_1}^{\tilde{y}}\int_{y\wedge y_1}^{y\vee y_1}\frac{1}{\lambda(z)^2\lambda(t)^2}v(x)D_{p,y}(z,x)$$

$$\left(v(x)D_{f,0}(t,x)-\frac{\partial v(x)D_{f,x}(t,x)}{\partial x_1}\right)f_{Y,X}(z,x)\,dx\,dz\,dt$$

$$+(1-2I_{\{y_1\leq \tilde{y}\}})\int_{y_1}^{y}\int_{\tilde{y}\wedge y_1}^{\tilde{y}\vee y_1}\frac{1}{\lambda(z)^2\lambda(t)^2}v(x)D_{p,y}(z,x)$$

$$\left(v(x)D_{f,0}(t,x)-\frac{\partial v(x)D_{f,x}(t,x)}{\partial x_1}\right)f_{Y,X}(z,x)\,dx\,dz\,dt.$$

Finally, the weak convergence

$$(Z_n(y))_{y\in[u_1,u_2]}\rightsquigarrow(Z(y))_{y\in[u_1,u_2]}$$

was proven, where Z is a centred Gaussian process with covariance function κ_Z. □

4.6.3 Proof of Theorem 4.2.4

As already mentioned, the estimation procedure described in Section 4.1.2 is related to the Mean-Square-Distance-From-Independence approach of Linton et al. (2008). There, the results of Chen et al. (2003) were used to prove asymptotic normality. Although calculations can not be carried over directly to the approach here, the following proof uses quite similar modifications of the results of Chen et al. (2003) as Linton et al. (2008). Let $\mathfrak{h}, f_{m_\tau}$ and f_{m_β} be some appropriate functional parameters. With the notations of Theorem 2 of Chen et al. (2003) one would have (view h as a functional parameter as well)

$$\begin{aligned}\theta&=c,\\ h&=(\mathfrak{h},f_{m_\tau},f_{m_\beta})^t=s,\\ M(\theta,h)(x,e)&=G_{MD}(c,s)(x,e).\end{aligned}$$

In this proof, use the notations of Linton et al. (2008) and write

$$\begin{aligned}s&=(\mathfrak{h},f_{m_\tau},f_{m_\beta})^t,\\ G_{MD}(c,s)(x,e)&=P(X\leq x,\tilde{\varepsilon}_c(\mathfrak{h},f_{m_\tau},f_{m_\beta})\leq e|X\in M_X)\\ &\quad-P(X\leq x|X\in M_X)P(\tilde{\varepsilon}_c(\mathfrak{h},f_{m_\tau},f_{m_\beta})\leq e|X\in M_X),\end{aligned}$$

$$G_{nMD}(c,s)(x,e) = \hat{P}(X \le x, \tilde{\varepsilon}_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta}) \le e | X \in M_X)$$
$$- \hat{P}(X \le x | X \in M_X)\hat{P}(\tilde{\varepsilon}_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta}) \le e | X \in M_X)$$

and

$$A(c,s) = ||G_{MD}(c,s)||_2$$

instead. Recall the definition of $\hat{B}$:

$$\hat{B} = \arg\min_{c \in [B_1, B_2]} \hat{A}(c, \hat{s}) = \arg\min_{c \in [B_1, B_2]} ||G_{nMD}(c, \hat{s})||_2.$$

Here, $||.||_2$ denotes the $\mathcal{L}^2$-norm on $M_X \times [e_a, e_b]$. Define

$$k_c(s,x,e) = \frac{h_1\big(\mathfrak{h}^{-1}\big((\mathfrak{h}^c(f_{m_\tau}(x)) + e(\mathfrak{h}^c(f_{m_\beta}(x)) - \mathfrak{h}^c(f_{m_\tau}(x))))^{\frac{1}{c}}\big)\big)^B - g(x)}{\sigma(x)}$$

with h_1 as in (4.8), so that $h = h_1^B$ and due to the model equation (3.1), it holds that

$$\begin{aligned}
&G_{MD}(c,s)(x,e)\\
&= P(X \le x, \tilde{\varepsilon}_c(s) \le e | X \in M_X) - P(X \le x | X \in M_X) P(\tilde{\varepsilon}_c(s) \le e | X \in M_X)\\
&= P\left(X \le x, \frac{\mathfrak{h}^c(Y) - \mathfrak{h}^c(f_{m_\tau}(X))}{\mathfrak{h}^c(f_{m_\beta}(X)) - \mathfrak{h}^c(f_{m_\tau}(X))} \le e | X \in M_X\right)\\
&\quad - P(X \le x | X \in M_X) P\left(\frac{\mathfrak{h}^c(Y) - \mathfrak{h}^c(f_{m_\tau}(X))}{\mathfrak{h}^c(f_{m_\beta}(X)) - \mathfrak{h}^c(f_{m_\tau}(X))} \le e | X \in M_X\right)\\
&= P(X \le x, \varepsilon \le k_c(s, X, e) | X \in M_X)\\
&\quad - P(X \le x | X \in M_X) P(\varepsilon \le k_c(s, X, e) | X \in M_X).
\end{aligned}$$

The function classes $\mathcal{H}$ and $\tilde{\mathcal{H}}$ were defined in (4.23) and (4.24). For any $\delta > 0$, B_δ was defined in (**C3**) as a δ-neighbourhood of B in $[B_1, B_2]$ and $\tilde{\mathcal{H}}_\delta = \{s \in \tilde{\mathcal{H}} : ||s - s_0||_{\mathcal{H}} < \delta\}$. Sometimes, the indices will be omitted if it is clear from the context, which norm is used. To proceed as in the proof of Theorem 2 of Chen et al. (2003) or more precisely as in the proof of a slightly modified version in the paper of Linton et al. (2008), there are several conditions that have to be proven, namely (**C1**)–(**C3**) and (**C5**). The conditions (**C4**) and (**C6**) have been assumed in the statement, see Lemma 4.2.3 for a discussion on validity of (**C4**) and (**C6**).
Each of the following lemmas is dedicated to one of these assumptions. Throughout the rest of the proof and especially in each of the following lemmas, (**A1**)–(**A7**) will be assumed. On the following pages, these auxiliary lemmas are proven. The actual proof, in which all of these statements are connected to finally prove the original assertion, can be found on page 163. The next Lemma uses similar techniques as Corollary 3.2.3 of Van der Vaart and Wellner (1996).

Lemma 4.6.2 *With A as in (4.13), it holds that $A(B, s_0) = 0$ and $\hat{B} - B = o_p(1)$, that is, (**C1**) is valid.*

Proof: The first part was already shown in Section 4.1.2. For the second part, consider the function classes

$$\mathcal{F} = \{(X, \varepsilon) \mapsto I_{\{X \in M_X\}} I_{\{\varepsilon \leq k_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta}, X, e)\}} : s \in \mathcal{H}, c \in [B_1, B_2], e \in [e_a, e_b]\}$$

and

$$\tilde{\mathcal{F}} = \{(X, \varepsilon) \mapsto I_{\{X \in M_X\}} I_{\{X \leq x\}} I_{\{\varepsilon \leq k_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta}, X, e)\}} : s \in \mathcal{H}, c \in [B_1, B_2], x \in M_X, e \in [e_a, e_b]\}.$$

It will be shown in the proof of Lemma 4.6.6 that the classes $\mathcal{F}$ and $\tilde{\mathcal{F}}$ are Donsker with respect to $\mathcal{L}^2\big(P^{(X,\varepsilon)}\big)$. Hence,

$$\begin{aligned}
\hat{P}(X \leq x, \tilde{\varepsilon}_c(s_0) \leq e | X \in M_X) &= \hat{P}(X \leq x, \varepsilon \leq k_c(s_0, X, e) | X \in M_X) \\
&= \frac{\frac{1}{n} \sum_{i=1}^n I_{\{X_i \leq x, \varepsilon_i \leq k_c(s_0, X_i, e)\}} I_{\{X_i \in M_X\}}}{\frac{1}{n} \sum_{i=1}^n I_{\{X_i \in M_X\}}} \\
&= P(X \leq x, \varepsilon \leq k_c(s_0, X, e) | X \in M_X) + \mathcal{O}_p\Big(\frac{1}{\sqrt{n}}\Big).
\end{aligned}$$

Assumption (**C4**) yields

$$\hat{f}_{m_\tau}(x) - F^{-1}_{Y|X}(\tau|x) = o_p\big(n^{-\frac{1}{4}}\big), \quad \hat{f}_{m_\beta}(x) - F^{-1}_{Y|X}(\beta|x) = o_p\big(n^{-\frac{1}{4}}\big)$$

and

$$\bar{h}_1(y) - h_1(y) = o_p\big(n^{-\frac{1}{4}}\big)$$

uniformly in $y \in [z_a, z_b]$ and $x \in M_X$. Consequently, it holds that

$$\sup_{c \in [B_1, B_2], x \in M_X, e \in [e_a, e_b]} |k_c(\hat{s}, x, e) - k_c(s_0, x, e)| = o_p(\delta_n),$$

where the sequence $(\delta_n)_{n \in \mathbb{N}}$ can be obtained from Lemma 1.5.1. Assumption (**C4**) ensures $\hat{s} \in \tilde{\mathcal{H}}$ with probability converging to one, so that Corollary 2.3.12 of Van der Vaart and Wellner (1996) leads to

$$\begin{aligned}
&\sup_{c \in [B_1, B_2], x \in M, e \in [e_a, e_b]} \big|\hat{P}(X \leq x, \tilde{\varepsilon}_c(\hat{s}) \leq e | X \in M_X) - P(X \leq x, \tilde{\varepsilon}_c(s_0) \leq e | X \in M_X)\big| \\
&= \sup_{c \in [B_1, B_2], x \in M, e \in [e_a, e_b]} \big|\hat{P}(X \leq x, \varepsilon \leq k_c(\hat{s}, X, e) | X \in M_X) \\
&\qquad - P(X \leq x, \varepsilon \leq k_c(s_0, X, e) | X \in M_X)\big| \\
&= o_p(1).
\end{aligned}$$

Analogous calculations can be done for $\hat{P}(X \leq x | X \in M_X)$ and $\hat{P}(\tilde{\varepsilon}_c(s_0) \leq e | X \in M_X)$. Therefore,

$$\begin{aligned}
\hat{A}(c, \hat{s}) = \bigg(\int_M \int_{[e_a, e_b]} \big(&\hat{P}(X \leq x, \tilde{\varepsilon}_c(\hat{s}) \leq e | X \in M_X) \\
&- \hat{P}(X \leq x | X \in M_X) \hat{P}(\tilde{\varepsilon}_c(\hat{s}) \leq e | X \in M_X)\big)^2 \, de \, dx \bigg)^{\frac{1}{2}}
\end{aligned}$$

$$= \Big(\int_M \int_{[e_a,e_b]} \big(P(X \le x, \tilde{\varepsilon}_c(s_0) \le e | X \in M_X)$$

$$- P(X \le x | X \in M_X) P(\tilde{\varepsilon}_c(s_0) \le e | X \in M_X)\big)^2 \, de \, dx \Big)^{\frac{1}{2}} + o_p(1)$$

$$= A(c, s_0) + o_p(1)$$

uniformly in $c \in [B_1, B_2]$. Since the map $c \mapsto A(c, s_0)$ is continuous and $c = B$ is the unique minimizer, it holds that

$$\inf_{c \in [B_1,B_2], |c-B|>\delta} A(c, s_0) > 0$$

for all $\delta > 0$ and thus, $\hat{B} = \underset{c \in [B_1,B_2]}{\arg\min}\, \hat{A}(c, \hat{s}) = \underset{c \in [B_1,B_2]}{\arg\min}\, A(c, s_0) + o_p(1) = B + o_p(1)$.

Lemma 4.6.3 *The ordinary derivative $\Gamma_1(c, s_0)(x, e)$ of $G_{MD}(c, s_0)(x, e)$ (with respect to c) exists for all $(x, e) \in M_X \times [e_a, e_b]$ in a neighbourhood of B and is continuous at $c = B$. $\Gamma_1(B, s_0)(x, e)$ is different from zero on a set with positive $\lambda_{M_X \times [e_a,e_b]}$-measure. Consequently, (**C2**) holds true.*

Proof: The proof can be divided into three steps namely the proof of the continuous differentiability of $c \mapsto k_c(s_0, x, e)$, the proof of continuous differentiability of $c \mapsto P(X \le x, \varepsilon \le k_c(s_0, X, e) | X \in M_X)$ and $c \mapsto G_{MD}(c, s_0)$ (each for all $(x, e) \in M \times [e_a, e_b]$) and finally the proof of $\Gamma_1(B, s_0)(x, e) \ne 0$.
First, for all $(x, e) \in M \times [e_a, e_b]$

$$\frac{\partial}{\partial c} k_c(s_0, x, e)$$

$$= \frac{\partial}{\partial c} \frac{\big(h_c(F^{-1}_{Y|X}(\tau|x)) + e(h_c(F^{-1}_{Y|X}(\beta|x)) - h_c(F^{-1}_{Y|X}(\tau|x)))\big)^{\frac{B}{c}} - g(x)}{\sigma(x)}$$

$$= \frac{1}{\sigma(x)} \Bigg[- \frac{B}{c^2} \Big(h_1\big(F^{-1}_{Y|X}(\tau|x)\big)^c + e\Big(h_1\big(F^{-1}_{Y|X}(\beta|x)\big)^c - h_1\big(F^{-1}_{Y|X}(\tau|x)\big)^c\Big)\Big)^{\frac{B}{c}}$$

$$\log \Big(h_1\big(F^{-1}_{Y|X}(\tau|x)\big)^c + e\Big(h_1\big(F^{-1}_{Y|X}(\beta|x)\big)^c - h_1\big(F^{-1}_{Y|X}(\tau|x)\big)^c\Big)\Big)$$

$$+ \frac{B}{c} \Big(h_1\big(F^{-1}_{Y|X}(\tau|x)\big)^c + e\Big(h_1\big(F^{-1}_{Y|X}(\beta|x)\big)^c - h_1\big(F^{-1}_{Y|X}(\tau|x)\big)^c\Big)\Big)^{\frac{B}{c}-1}$$

$$\Big(\log\Big(h_1\big(F^{-1}_{Y|X}(\tau|x)\big)\Big) h_1\big(F^{-1}_{Y|X}(\tau|x)\big)^c + e\Big(\log\Big(h_1\big(F^{-1}_{Y|X}(\beta|x)\big)\Big) h_1\big(F^{-1}_{Y|X}(\beta|x)\big)^c$$

$$- \log\Big(h_1\big(F^{-1}_{Y|X}(\tau|x)\big)\Big) h_1\big(F^{-1}_{Y|X}(\tau|x)\big)^c\Big)\Big)\Bigg]. \tag{4.38}$$

Due to $0 < h_c(F^{-1}_{Y|X}(\tau|x)), h_c(F^{-1}_{Y|X}(\beta|x))$ as well as

$$0 < h_c(z_a) \le h_c(F^{-1}_{Y|X}(\tau|x)) + e(h_c(F^{-1}_{Y|X}(\beta|x)) - h_c(F^{-1}_{Y|X}(\tau|x))) \le h_c(z_b)$$

for all $x \in M_X, e \in [e_a, e_b]$ the function $(c, x, e) \mapsto \frac{\partial}{\partial c} k_c(s_0, x, e)$ is well defined, continuous and thus bounded on $[B_1, B_2] \times M_X \times [e_a, e_b]$. Additionally, for each $c \in [B_1, B_2]$ the points

$(x,e) \in M_X \times [e_a, e_b]$ with $\frac{\partial}{\partial c} k_c(s_0, x, e) = 0$ form a null set with respect to $\lambda_{M_X \times [e_a,e_b]}$. Second, $P(X \leq x, \varepsilon \leq k_c(s_0, X, e) | X \in M_X)$ can be written as

$$\begin{aligned} P(X \leq x, \varepsilon \leq k_c(s_0, X, e) | X \in M_X) &= \frac{P(X \leq x, \varepsilon \leq k_c(s_0, X, e), X \in M_X)}{P(X \in M)} \\ &= \frac{1}{P(X \in M)} \int_{M_X \cap (-\infty, x]} F_\varepsilon(k_c(s_0, v, e)) f_X(v)\, dv. \end{aligned}$$

Analogously,

$$P(\varepsilon \leq k_c(s_0, X, e) | X \in M_X) = \frac{1}{P(X \in M)} \int_{M_X} F_\varepsilon(k_c(s_0, v, e)) f_X(v)\, dv.$$

The Dominated Convergence Theorem leads to

$$\begin{aligned} &\frac{\partial}{\partial c} P(X \leq x, \varepsilon \leq k_c(s_0, X, e) | X \in M_X) \\ &= \frac{1}{P(X \in M_X)} \int_{M_X \cap (-\infty, x]} \frac{\partial}{\partial c} F_\varepsilon(k_c(s_0, v, e)) f_X(v)\, dv \\ &= \frac{1}{P(X \in M_X)} \int_{M_X \cap (-\infty, x]} f_\varepsilon(k_c(s_0, v, e)) \frac{\partial}{\partial c} k_c(s_0, v, e) f_X(v)\, dv, \end{aligned}$$

where the supremum of the integrand, which is continuous and evaluated on a compact set, can be taken as a majorant. Consequently

$$\begin{aligned} \frac{\partial}{\partial c} G_{MD}(c, s_0)(x, e) = \frac{1}{P(X \in M_X)} \int_{M_X} & f_\varepsilon(k_c(s_0, v, e)) \\ & \frac{\partial}{\partial c} k_c(s_0, v, e)(I_{(-\infty, x]}(v) - P(X \leq x | X \in M_X)) f_X(v)\, dv. \end{aligned}$$

Hence, $\frac{\partial}{\partial c} G_{MD}(B, s_0)(x, e) = 0$ for all $(x,e) \in M_X \times [e_a, e_b]$ is equivalent to

$$\begin{aligned} &\int_{M_X} f_\varepsilon(k_B(s_0, v, e)) \frac{\partial}{\partial c} k_c(s_0, v, e)\Big|_{c=B} I_{(-\infty, x]}(v) f_X(v)\, dv \\ &= P(X \leq x | X \in M_X) \int_{M_X} f_\varepsilon(k_B(s_0, v, e)) \frac{\partial}{\partial c} k_c(s_0, v, e)\Big|_{c=B} f_X(v)\, dv \\ &= \int_{M_X} \frac{\int_{M_X} f_\varepsilon(k_B(s_0, v, e)) \frac{\partial}{\partial c} k_c(s_0, v, e)\big|_{c=B} f_X(v)\, dv}{P(X \in M_X)} I_{(-\infty, x]}(w) f_X(w)\, dw \end{aligned}$$

for almost all $(x,e) \in M_X \times [e_a, e_b]$ with respect to $\lambda_{M_X \times [e_a,e_b]}$.
Therefore, $v \mapsto f_\varepsilon(k_B(s_0, v, e)) \frac{\partial}{\partial c} k_c(s_0, v, e)\big|_{c=B}$ would be constant on M_X for almost all $e \in [e_a, e_b]$. Due to

$$k_B(s_0, v, e) = F_\varepsilon^{-1}(\tau) + e\big(F_\varepsilon^{-1}(\beta) - F_\varepsilon^{-1}(\tau)\big)$$

and (plug $c = B$ into (4.38))

$$\begin{aligned} \frac{\partial}{\partial c} k_c(s_0, v, e)\Big|_{c=B} = -\frac{1}{B} \bigg(& \frac{g(v)}{\sigma(v)} + F_\varepsilon^{-1}(\tau) + e\big(F_\varepsilon^{-1}(\beta) - F_\varepsilon^{-1}(\tau)\big) \bigg) \\ & \log\Big(g(v) + \sigma(v)\big(F_\varepsilon^{-1}(\tau) + e\big(F_\varepsilon^{-1}(\beta) - F_\varepsilon^{-1}(\tau)\big)\big)\Big) \end{aligned}$$

$$+\frac{1}{B}\Big(\log\big(g(v)+\sigma(v)F_\varepsilon^{-1}(\tau)\big)\big(g(v)+\sigma(v)F_\varepsilon^{-1}(\tau)\big)$$
$$+e\Big(\log\big(g(v)+\sigma(v)F_\varepsilon^{-1}(\beta)\big)\big(g(v)+\sigma(v)F_\varepsilon^{-1}(\beta)\big)$$
$$-\log\big(g(v)+\sigma(v)F_\varepsilon^{-1}(\tau)\big)\big(g(v)+\sigma(v)F_\varepsilon^{-1}(\tau)\big)\Big)\Big)$$

the map $v\mapsto f_\varepsilon(k_B(s_0,v,e))\frac{\partial}{\partial c}k_c(s_0,v,e)\big|_{c=B}$ depends on v at least for some $e\in[e_a,e_b]$, that is $\frac{\partial}{\partial c}G_{MD}(B,s_0)(x,e)\neq 0$ for a set with $\lambda_{M_X\times[e_a,e_b]}$-measure greater than zero. $\square$

Lemma 4.6.4 *There exists a $\delta>0$ such that for all $c\in B_\delta,(x,e)\in M_X\times[e_a,e_b]$ the directional derivative $\Gamma_2(c,s_0)(x,e)[s-s_0]$ of $G_{MD}(c,s_0)(x,e)$ with respect to s exists in all directions $[s-s_0]$. Moreover, consider a positive sequence $\delta_n\to 0$ and $(c,s)\in B_{\delta_n}\times\tilde{\mathcal{H}}_{\delta_n}$. Then,*

(i) for an appropriate constant $C\geq 0$ one has

$$\begin{aligned}&||G_{MD}(c,s)-G_{MD}(c,s_0)-\Gamma_2(c,s_0)[s-s_0]||_2\\&\leq C\big(||\mathfrak{h}-h_1||^{\frac{3}{2}}_{[z_a,z_b]}+||f_{m_\tau}-F^{-1}_{Y|X}(\tau|\cdot)||^2_{M_X}+||f_{m_\beta}-F^{-1}_{Y|X}(\beta|\cdot)||^2_{M_X}\big).\end{aligned}$$

(ii) one has $||\Gamma_2(c,s_0)[\hat{s}-s_0]-\Gamma_2(B,s_0)[\hat{s}-s_0]||=o_p(|c-B|)+o_p\big(n^{-\frac{1}{2}}\big)$.

Therefore, (**C3**) is valid.

Proof: First, existence of the directional derivatives is shown, before conditions (i) and (ii) are proven.

Directional derivative with respect to h: Define for some fixed $c,\mathfrak{h},x,e$

$$\begin{aligned}f_{h,t}&:=h_1+t(\mathfrak{h}-h_1),\\\psi(t,z)&:=f_{h,t}^{-1}(z),\\z_c(f_{h,t},f_{m_\tau},f_{m_\beta},x,e)&:=\big(f_{h,t}^c(f_{m_\tau}(x))+e(f_{h,t}^c(f_{m_\beta}(x))-f_{h,t}^c(f_{m_\tau}(x)))\big)^{\frac{1}{c}}.\end{aligned}$$

Mostly, the components x,e will be omitted and $z_c(t)$ will be written as an abbreviation for $z_c(f_{h,t},f_{m_\tau},f_{m_\beta},x,e)$. Further, all derivatives with respect to t are marked with a "$\cdot$", those with respect to y are marked with a "$\prime$". Then, one has

$$h_1'(z)=-\frac{h_1(z)}{\lambda(z)},$$

$$\psi'(t,z)=\frac{1}{f_{h,t}'(f_{h,t}^{-1}(z_c(t)))}\overset{t=0}{=}-\frac{\lambda(h_1^{-1}(z_c(0)))}{z_c(0)}$$

as well as

$$\begin{aligned}&\frac{\partial}{\partial t}\psi(t,f_{h,t}(h_1^{-1}(z_c(t))))\\&=\dot{\psi}(t,f_{h,t}(h_1^{-1}(z_c(t))))\end{aligned}$$

$$
\begin{aligned}
&+\psi'(t, f_{h,t}(h_1^{-1}(z_c(t))))\Big(\dot{f}_{h,t}(h_1^{-1}(z_c(t))) + f'_{h,t}(h_1^{-1}(z_c(t)))\frac{\frac{\partial}{\partial t}z_c(t)}{h_1'(h_1^{-1}(z_c(t)))}\Big)\\
&\overset{t\to 0}{\longrightarrow} \dot{\psi}(0, z_c(0)) + \psi'(0, z_c(0))\Big((\mathfrak{h} - h_1)(h_1^{-1}(z_c(0))) + \frac{\partial}{\partial t}z_c(t)\big|_{t=0}\Big)\\
&= \dot{\psi}(0, z_c(0)) - \frac{\lambda(h_1^{-1}(z_c(0)))}{z_c(0)}\Big(\mathfrak{h}(h_1^{-1}(z_c(0))) - z_c(0) + \frac{\partial}{\partial t}z_c(t)\Big|_{t=0}\Big).
\end{aligned}
$$

Due to

$$
\frac{\partial}{\partial t}\psi(t, f_{h,t}(h_1^{-1}(z_c(t)))) = \frac{\partial}{\partial t}h_1^{-1}(z_c(t))) = \frac{\frac{\partial}{\partial t}z_c(t)}{h_1'(h_1^{-1}(z_c(t)))} \overset{t\to 0}{\longrightarrow} -\frac{\lambda(h_1^{-1}(z_c(0)))\frac{\partial}{\partial t}z_c(t)\big|_{t=0}}{z_c(0)},
$$

it holds that

$$
\dot{\psi}(0, z_c(0)) = \frac{\lambda(h_1^{-1}(z_c(0)))}{z_c(0)}(\mathfrak{h}(h_1^{-1}(z_c(0))) - z_c(0)),
$$

so that

$$
\begin{aligned}
\frac{\partial}{\partial t}\psi(t, z_c(t)) &= \dot{\psi}(t, z_c(t)) + \psi'(t, z_c(t))\frac{\partial}{\partial t}z_c(t)\\
&\overset{t\to 0}{\longrightarrow} \dot{\psi}(0, z_c(0)) + \psi'(0, z_c(0))\frac{\partial}{\partial t}z_c(t)\Big|_{t=0}\\
&= \frac{\lambda(h_1^{-1}(z_c(0)))}{z_c(0)}\Big(\mathfrak{h}(h_1^{-1}(z_c(0))) - z_c(0) - \frac{\partial}{\partial t}z_c(t)\Big|_{t=0}\Big).
\end{aligned}
$$

Additionally,

$$
\begin{aligned}
\frac{\partial}{\partial t}z_c(t)\Big|_{t=0} &= \frac{\partial}{\partial t}\big(f_{h,t}^c(f_{m_\tau}(x)) + e(f_{h,t}^c(f_{m_\beta}(x)) - f_{h,t}^c(f_{m_\tau}(x)))\big)^{\frac{1}{c}}\Big|_{t=0}\\
&= \frac{1}{c}\big(h_1^c(f_{m_\tau}(x)) + e(h_1^c(f_{m_\beta}(x)) - h_1^c(f_{m_\tau}(x)))\big)^{\frac{1}{c}-1}\\
&\quad\big(ch_1^{c-1}(f_{m_\tau}(x))(\mathfrak{h}(f_{m_\tau}(x)) - h_1(f_{m_\tau}(x)))\\
&\quad + e(ch_1^{c-1}(f_{m_\beta}(x))(\mathfrak{h}(f_{m_\beta}(x)) - h_1(f_{m_\beta}(x))\\
&\qquad - ch_1^{c-1}(f_{m_\tau}(x))(\mathfrak{h}(f_{m_\tau}(x)) - h_1(f_{m_\tau}(x))))\big).
\end{aligned}
$$

This in turn results in (for the special case $f_{m_\tau} = F_{Y|X}^{-1}(\tau|\cdot), f_{m_\beta} = F_{Y|X}^{-1}(\beta|\cdot)$)

$$
\begin{aligned}
D_h k_c(s_0, x, e)[\mathfrak{h} - h_1] &= \frac{\partial}{\partial t}k_c(f_{h,t}, F_{Y|X}^{-1}(\tau|\cdot), F_{Y|X}^{-1}(\beta|\cdot), x, e)\Big|_{t=0}\\
&= \frac{\partial}{\partial t}\frac{h_1(\psi(t, z_c(t)))^B - g(x)}{\sigma(x)}\Big|_{t=0}\\
&= \frac{Bh_1(\psi(t, z_c(t)))^{B-1}h_1'(\psi(t, z_c(t)))\frac{\partial}{\partial t}\psi(t, z_c(t))}{\sigma(x)}\Big|_{t=0}\\
&= \frac{Bz_c(0)^{B-1}\big(z_c(0) - \mathfrak{h}(h_1^{-1}(z_c(0))) + \frac{\partial}{\partial t}z_c(t)\big|_{t=0}\big)}{\sigma(x)} \qquad (4.39)
\end{aligned}
$$

and by applying the Dominated Convergence Theorem

$$
D_h G_{MD}(c, s_0)(x, e)[\mathfrak{h} - h_1]
$$

$$= \frac{1}{P(X \in M)} \frac{\partial}{\partial t} \Big(\int_{M_X} F_\varepsilon(k_c(f_{h,t}, F^{-1}_{Y|X}(\tau|\cdot), F^{-1}_{Y|X}(\beta|\cdot), w, e)) I_{\{w \le x\}} f_X(w)\, dw$$

$$- P(X \le x | X \in M) \int_{M_X} F_\varepsilon(k_c(f_{h,t} F^{-1}_{Y|X}(\tau|\cdot), F^{-1}_{Y|X}(\beta|\cdot), w, e)) f_X(w)\, dw \Big) \Big|_{t=0}$$

$$= \frac{1}{P(X \in M)} \Big(\int_{M_X} f_\varepsilon(k_c(s_0, w, e)) \big(I_{\{w \le x\}} - P(X \le x | X \in M_X) \big)$$

$$D_h k_c(s_0, w, e)[\mathfrak{h} - h_1] f_X(w)\, dw \Big).$$

Directional derivative with respect to $\mathbf{f_{m_\tau}}$ and $\mathbf{f_{m_\beta}}$: For $\mathfrak{h} = h_1$, k_c simplifies to

$$k_c(h_1, f_{m_\tau}, f_{m_\beta}, x, e) = \frac{z_c(h_1, f_{m_\tau}, f_{m_\beta})^B - g(x)}{\sigma(x)}.$$

Hence, with

$$f_{m_\tau,t} = F^{-1}_{Y|X}(\tau|\cdot) + t\big(f_{m_\tau} - F^{-1}_{Y|X}(\tau|\cdot)\big) \quad \text{and} \quad f_{m_\beta,t} = F^{-1}_{Y|X}(\beta|\cdot) + t\big(f_{m_\beta} - F^{-1}_{Y|X}(\beta|\cdot)\big)$$

one has

$$D_{f_{m_\tau}} k_c(h_1, F^{-1}_{Y|X}(\tau|\cdot), F^{-1}_{Y|X}(\beta|\cdot), x, e)\big[f_{m_\tau} - F^{-1}_{Y|X}(\tau|\cdot)\big]$$

$$= \frac{\partial}{\partial t} k_c(h_1, f_{m_\tau,t}, F^{-1}_{Y|X}(\beta|\cdot), x, e)\Big|_{t=0}$$

$$= \frac{\partial}{\partial t} \frac{z_c(h_1, f_{m_\tau,t}, F^{-1}_{Y|X}(\beta|\cdot))^B - g(x)}{\sigma(x)}\Bigg|_{t=0}$$

$$= \frac{\partial}{\partial t} \frac{\big(h_c(f_{m_\tau,t}(x)) + e(h_c(F^{-1}_{Y|X}(\beta|x)) - h_c(f_{m_\tau,t}(x)))\big)^{\frac{B}{c}} - g(x)}{\sigma(x)}\Bigg|_{t=0}$$

$$= -\frac{B(...)^{\frac{B}{c}-1}(1-e) h_c(F^{-1}_{Y|X}(\tau|x))(f_{m_\tau}(x) - F^{-1}_{Y|X}(\tau|x))}{\sigma(x)\lambda(F^{-1}_{Y|X}(\tau|x))}$$

as well as

$$D_{f_\beta} k_c(s_0, x, e)\big[f_\beta - F^{-1}_{Y|X}(\beta|\cdot)\big]$$

$$= \frac{\partial}{\partial t} k_c(h_1, F^{-1}_{Y|X}(\tau|\cdot), f_{m_\beta,t}, x, e)\Big|_{t=0}$$

$$= \frac{\partial}{\partial t} \frac{z_c(h_1, F^{-1}_{Y|X}(\tau|\cdot), f_{m_\beta,t})^B - g(x)}{\sigma(x)}\Bigg|_{t=0}$$

$$= \frac{\partial}{\partial t} \frac{\big(h_c(F^{-1}_{Y|X}(\tau|x)) + e(h_c(f_{m_\beta,t}(x)) - h_c(F^{-1}_{Y|X}(\tau|x)))\big)^{\frac{B}{c}} - g(x)}{\sigma(x)}\Bigg|_{t=0}$$

$$= -\frac{B(...)^{\frac{B}{c}-1} e h_c(F^{-1}_{Y|X}(\beta|x))(f_{m_\beta}(x) - F^{-1}_{Y|X}(\beta|x))}{\sigma(x)\lambda(F^{-1}_{Y|X}(\beta|x))}.$$

Directional derivative with respect to s: This results from the previous parts of the proof as follows: Define $s_t = (f_{h,t}, f_{m_\tau,t}, f_{m_\beta,t})$ and $t \mapsto \tilde{z}_c(t) := z_c(s_t, x, e)$. Then,

$$k_c(s_t, x, e) = \frac{\partial}{\partial t} k_c(s_t, x, e)\Big|_{t=0} = \frac{\partial}{\partial t} \frac{h_1(\psi(t, \tilde{z}_c(t)))^B - g(x)}{\sigma(x)}\Big|_{t=0}$$

only depends on $f_{m_\tau,t}$ and $f_{m_\beta,t}$ via $\tilde{z}_c(t)$, respectively. Due to $\tilde{z}_c(0) = z_c(s_0)$ ($= z_c(0)$ with the notation from before), one can proceed as for the derivative with respect to h to obtain

$$D_s k_c(s_0, x, e)[s - s_0] = \frac{B\tilde{z}_c(0)^{B-1}(\tilde{z}_c(0) - \mathfrak{h}(h_1^{-1}(\tilde{z}_c(0))) + \frac{\partial}{\partial t}\tilde{z}_c(t)\big|_{t=0})}{\sigma(x)}.$$

At the same time,

$$\begin{aligned}
\frac{\partial}{\partial t}\tilde{z}_c(t)\big|_{t=0} &= \frac{\partial}{\partial t} z_c(f_{h,t}, f_{m_\tau,t}, f_{m_\beta,t})\Big|_{t=0} \\
&= \begin{pmatrix} D_h z_c(s_0) & , & D_{f_{m_\tau}} z_c(s_0) & , & D_{f_{m_\beta}} z_c(s_0) \end{pmatrix} \begin{pmatrix} \mathfrak{h} - h_1 \\ f_{m_\tau} - F_{Y|X}^{-1}(\tau|x) \\ f_{m_\beta} - F_{Y|X}^{-1}(\beta|x) \end{pmatrix},
\end{aligned}$$

which in total leads to

$$\begin{aligned}
D_s k_c(s_0, x, e)[s - s_0] &= D_h k_c(s_0, x, e)[\mathfrak{h} - h_1] + D_{f_{m_\tau}} k_c(s_0, x, e)\big[f_{m_\tau} - F_{Y|X}^{-1}(\tau|\cdot)\big] \\
&\quad + D_{f_{m_\beta}} k_c(s_0, x, e)\big[f_{m_\beta} - F_{Y|X}^{-1}(\beta|\cdot)\big]
\end{aligned}$$

and (after applying the Dominated Convergence Theorem)

$$\begin{aligned}
&D_s G_{MD}(c, s_0)(x, e)[s - s_0] \\
&= \frac{1}{P(X \in M)} \frac{\partial}{\partial t}\Big(\int_{M_X} F_\varepsilon(k_c(f_{h,t}, f_{m_\tau,t}, f_{m_\beta,t}, w, e)) I_{\{w \leq x\}} f_X(w)\, dw \\
&\quad - P(X \leq x | X \in M) \int_{M_X} F_\varepsilon(k_c(f_{h,t}, f_{m_\tau,t}, f_{m_\beta,t}, w, e)) f_X(w)\, dw\Big)\Big|_{t=0} \\
&= \frac{1}{P(X \in M)}\Big(\int_{M_X} f_\varepsilon(k_c(s_0, w, e))\big(I_{\{w \leq x\}} - P(X \leq x | X \in M)\big)\big(D_h k_c(s_0, w, e)[\mathfrak{h} - h_1] \\
&\quad + D_{f_{m_\tau}} k_c(s_0, w, e)\big[f_{m_\tau} - F_{Y|X}^{-1}(\tau|\cdot)\big] + D_{f_{m_\beta}} k_c(s_0, w, e)\big[f_{m_\beta} - F_{Y|X}^{-1}(\beta|\cdot)\big]\big) f_X(w)\, dw\Big) \\
&= D_h G_{MD}(c, s_0)(x, e)[\mathfrak{h} - h_1] + D_{f_{m_\tau}} G_{MD}(c, s_0)(x, e)\big[f_{m_\tau} - F_{Y|X}^{-1}(\tau|\cdot)\big] \\
&\quad + D_{f_{m_\beta}} G_{MD}(c, s_0)(x, e)\big[f_{m_\beta} - F_{Y|X}^{-1}(\beta|\cdot)\big]. \qquad (4.40)
\end{aligned}$$

Proof of (i): First, an auxiliary lemma is proven.

Lemma 4.6.5 *Let $\delta_n \searrow 0, s = (\mathfrak{h}, f_{m_\tau}, f_{m_\beta}) \in \tilde{\mathcal{H}}_{\delta_n}$ and $0 < \eta < \frac{h_1(z_b) - h_1(z_a)}{2}$. Then,*

$$\sup_{t \in [h_1(z_a) + \eta, h_1(z_b) - \eta]} |\mathfrak{h}^{-1}(t) - h_1^{-1}(t)| = \mathcal{O}\big(||\mathfrak{h} - h_1||_{[z_a, z_b]}\big)$$

and

$$||\mathfrak{h}' - h_1'||_{[z_a, z_b]} = \mathcal{O}\Big(\sqrt{||\mathfrak{h} - h_1||_{[z_a, z_b]}}\Big) = \sqrt{\delta_n}.$$

Proof: Let $t \in [h_1(z_a)+\eta, h_1(z_b)-\eta]$, so that $t \in (\mathfrak{h}(z_a), \mathfrak{h}(z_b))$ for n sufficiently large. For an appropriate $\tilde{t} \in (h_1(z_a), h_1(z_b))$, one has

$$\begin{aligned}
\mathfrak{h}^{-1}(t) - h_1^{-1}(t) &= h_1^{-1}\big(h_1(\mathfrak{h}^{-1}(t))\big) - h_1^{-1}(t) \\
&= \frac{1}{h_1'(h_1^{-1}(\tilde{t}))}(h_1(\mathfrak{h}^{-1}(t)) - t) \\
&= \frac{1}{h_1'(h_1^{-1}(\tilde{t}))}(h_1(\mathfrak{h}^{-1}(t)) - \mathfrak{h}(\mathfrak{h}^{-1}(t))) \\
&= \mathcal{O}(||\mathfrak{h} - h_1||).
\end{aligned}$$

To handle $||\mathfrak{h}' - h_1'||_{[z_a,z_b]}$, notice that the second derivatives of all $\mathfrak{h}$ with $(\mathfrak{h}, f_{m_\tau}, f_{m_\beta}) \in \tilde{\mathcal{H}}_{\delta_n} \subseteq \tilde{\mathcal{H}}$ are bounded. Hence, for all $z \in (z_a, z_b)$ and some appropriate $\tilde{z}_1, \tilde{z}_2 \in (z_a, z_b)$

$$\mathfrak{h}\big(z + \sqrt{\delta_n}\big) - h_1\big(z + \sqrt{\delta_n}\big) - \mathfrak{h}(z) + h_1(z) = \mathcal{O}(\delta_n)$$

by definition of $\tilde{\mathcal{H}}_{\delta_n}$ as well as

$$\begin{aligned}
&\mathfrak{h}\big(z + \sqrt{\delta_n}\big) - h_1\big(z + \sqrt{\delta_n}\big) - \mathfrak{h}(z) + h_1(z) \\
&= (\mathfrak{h}'(z) - h_1'(z))\sqrt{\delta_n} + \frac{\mathfrak{h}''(\tilde{z}_1) - h_1''(\tilde{z}_2)}{2}\delta_n \\
&= (\mathfrak{h}'(z) - h_1'(z))\sqrt{\delta_n} + \mathcal{O}(\delta_n)
\end{aligned}$$

uniformly in $z \in (z_a, z_b)$. Therefore, one has $||\mathfrak{h}' - h_1'||_{[z_a,z_b]} = \mathcal{O}\big(\sqrt{\delta_n}\big)$. The same argument with $\tilde{\delta}_n = ||\mathfrak{h} - h_1||_{[z_a,z_b]}$ leads to $||\mathfrak{h}' - h_1'||_{[z_a,z_b]} = \mathcal{O}\Big(\sqrt{||\mathfrak{h} - h_1||_{[z_a,z_b]}}\Big)$. □

Let $\delta_n \searrow 0$ and $(c, s) \in B_{\delta_n} \times \tilde{\mathcal{H}}_{\delta_n}$. To apply the lemma from above, split the norm into three parts (see (4.40))

$$\begin{aligned}
&||G_{MD}(c, s) - G_{MD}(c, s_0) - \Gamma_2(c, s_0)[s - s_0]||_2 \\
&= ||G_{MD}(c, \mathfrak{h}, f_{m_\tau}, f_{m_\beta}) - G_{MD}(c, h_1, f_{m_\tau}, f_{m_\beta}) - D_h G_{MD}(c, s_0)[\mathfrak{h} - h_1] \\
&\quad + G_{MD}(c, h_1, f_{m_\tau}, f_{m_\beta}) - G_{MD}(c, h_1, F_{Y|X}^{-1}(\tau|\cdot), f_{m_\beta}) - D_{f_{m_\tau}} G_{MD}(c, s_0)\big[f_{m_\tau} - F_{Y|X}^{-1}(\tau|\cdot)\big] \\
&\quad + G_{MD}(c, h_1, F_{Y|X}^{-1}(\tau|\cdot), f_{m_\beta}) - G_{MD}(c, h_1, F_{Y|X}^{-1}(\tau|\cdot), F_{Y|X}^{-1}(\beta|\cdot)) \\
&\quad - D_{f_{m_\beta}} G_{MD}(c, s_0)\big[f_{m_\beta} - F_{Y|X}^{-1}(\beta|\cdot)\big]||_2 \\
&\leq ||G_{MD}(c, \mathfrak{h}, f_{m_\tau}, f_{m_\beta}) - G_{MD}(c, h_1, f_{m_\tau}, f_{m_\beta}) - D_h G_{MD}(c, s_0)[\mathfrak{h} - h_1]||_2 \\
&\quad + ||G_{MD}(c, h_1, f_{m_\tau}, f_{m_\beta}) - G_{MD}(c, h_1, F_{Y|X}^{-1}(\tau|\cdot), f_{m_\beta}) - D_{f_{m_\tau}} G_{MD}(c, s_0)\big[f_{m_\tau} - F_{Y|X}^{-1}(\tau|\cdot)\big]||_2 \\
&\quad + ||G_{MD}(c, h_1, F_{Y|X}^{-1}(\tau|\cdot), f_{m_\beta}) - G_{MD}(c, s_0) - D_{f_{m_\beta}} G_{MD}(c, s_0)\big[f_{m_\beta} - F_{Y|X}^{-1}(\beta|\cdot)\big]||_2.
\end{aligned}$$

Notice for the first summand that due to

$$\begin{aligned}
&||G_{MD}(c, \mathfrak{h}, f_{m_\tau}, f_{m_\beta}) - G_{MD}(c, h_1, f_{m_\tau}, f_{m_\beta}) - D_h G_{MD}(c, s_0)[\mathfrak{h} - h_1]||_2 \\
&= \bigg(\int_M \int_{[e_a,e_b]} \big(P(X \leq x, \varepsilon \leq k_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta}, X, e)|X \in M)
\end{aligned}$$

$$
\begin{aligned}
&- P(X \le x | X \in M) P(\varepsilon \le k_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta}, X, e) | X \in M) \\
&- P(X \le x, \varepsilon \le k_c(h_1, f_{m_\tau}, f_{m_\beta}, X, e) | X \in M) \\
&+ P(X \le x | X \in M) P(\varepsilon \le k_c(h_1, f_{m_\tau}, f_{m_\beta}, X, e) | X \in M) \\
&- D_h G_{MD}(h_1, F^{-1}_{Y|X}(\tau|\cdot), F^{-1}_{Y|X}(\beta|\cdot), X, e)(x, e)[\mathfrak{h} - h_1]\big)^2 \, de \, dx \Big)^{\frac{1}{2}} \\
= &\frac{1}{P(X \in M_X)} \int_{M_X} \big(I_{\{v \le x\}} - P(X \le x | X \in M)\big)\big(F_\varepsilon(k_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta}, v, e)) \\
&- F_\varepsilon(k_c(h_1, f_{m_\tau}, f_{m_\beta}, v, e)) - f_\varepsilon(k_c(h_1, f_{m_\tau}, f_{m_\beta}, v, e)) \\
&\quad D_h k_c(h_1, F^{-1}_{Y|X}(\tau|\cdot), F^{-1}_{Y|X}(\beta|\cdot), v, e)[\mathfrak{h} - h_1]\big) f_X(v) \, dv
\end{aligned}
$$

and

$$
\begin{aligned}
&F_\varepsilon(k_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta}, v, e)) - F_\varepsilon(k_c(h_1, f_{m_\tau}, f_{m_\beta}, v, e)) \\
&- f_\varepsilon(k_c(h_1, f_{m_\tau}, f_{m_\beta}, v, e)) D_h k_c(h_1, F^{-1}_{Y|X}(\tau|\cdot), F^{-1}_{Y|X}(\beta|\cdot), v, e)[\mathfrak{h} - h_1] \\
= \quad &f_\varepsilon(k_c(h_1, f_{m_\tau}, f_{m_\beta}, v, e))(k_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta}, v, e) - k_c(h_1, f_{m_\tau}, f_{m_\beta}, v, e)) \\
&+ f'_\varepsilon(\tilde{k})(k_c(h_1, f_{m_\tau}, f_{m_\beta}, v, e) - k_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta}, v, e))^2 \\
&- f_\varepsilon(k_c(h_1, f_{m_\tau}, f_{m_\beta}, v, e)) D_h k_c(h_1, F^{-1}_{Y|X}(\tau|\cdot), F^{-1}_{Y|X}(\beta|\cdot), v, e)[\mathfrak{h} - h_1]
\end{aligned}
$$

for some $\tilde{k}$ between $k_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta}, v, e)$ and $k_c(h_1, f_{m_\tau}, f_{m_\beta}, v, e)$ it suffices to prove

$$
\big|k_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta}, v, e) - k_c(h_1, f_{m_\tau}, f_{m_\beta}, v, e) - D_h k_c(s_0, v, e)[\mathfrak{h} - h_1]\big| \le C||\mathfrak{h} - h_1||^{\frac{3}{2}}_{[z_a, z_b]}
$$

for an appropriate $C > 0$ and uniformly in $c \in B_\delta, v \in M_X, e \in [e_a, e_b], f_{m_\tau}, f_{m_\beta}$, such that $(\mathfrak{h}, f_{m_\tau}, f_{m_\beta}) \in \tilde{\mathcal{H}}$. Analogous calculations for $D_{f_{m_\tau}}$ and $D_{f_{m_\beta}}$ yield the sufficient conditions

$$
\big|k_c(h_1, f_{m_\tau}, f_{m_\beta}, v, e) - k_c(h_1, F^{-1}_{Y|X}(\tau|\cdot), f_{m_\beta}, v, e) \tag{4.41}
$$

$$
- D_{f_{m_\tau}} k_c(h_1, F^{-1}_{Y|X}(\tau|\cdot), f_{m_\beta}, v, e)\big[f_{m_\tau} - F^{-1}_{Y|X}(\tau|\cdot)\big]\big| \tag{4.42}
$$

$$
\le C||f_{m_\tau} - F^{-1}_{Y|X}(\tau|\cdot)||^2_{M_X} \tag{4.43}
$$

(uniformly in $c \in B_\delta, v \in M_X, e \in [e_a, e_b]$ and f_{m_β}, such that $(h_1, f_{m_\tau}, f_{m_\beta}) \in \tilde{\mathcal{H}}$) and

$$
\begin{aligned}
&\big|k_c(h_1, F^{-1}_{Y|X}(\tau|\cdot), f_{m_\beta}, v, e) - k_c(s_0, v, e) - D_{f_{m_\beta}} k_c(s_0, v, e)\big[f_{m_\beta} - F^{-1}_{Y|X}(\beta|\cdot)\big]\big| \\
&\le C||f_{m_\beta} - F^{-1}_{Y|X}(\beta|\cdot)||^2_{M_X}
\end{aligned} \tag{4.44}
$$

uniformly in $c \in B_\delta, v \in M_X, e \in [e_a, e_b]$ to handle the second and third summand, respectively.
(4.39) leads to

$$
\begin{aligned}
&k_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta}, v, e) - k_c(h_1, f_{m_\tau}, f_{m_\beta}, v, e) - D_h k_c(s_0, v, e)[\mathfrak{h} - h_1] \\
&= \frac{1}{\sigma(v)} \bigg(h_1(\mathfrak{h}^{-1}(z_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta})))^B - h_1(h_1^{-1}(z_c(h_1, f_{m_\tau}, f_{m_\beta})))^B - B z_c(h_1, f_{m_\tau}, f_{m_\beta})^{B-1}
\end{aligned}
$$

$$\left(\frac{\partial}{\partial t} z_c(f_{h,t}, f_{m_\tau}, f_{m_\beta})\Big|_{t=0} + h_1(h_1^{-1}(z_c(h_1, f_{m_\tau}, f_{m_\beta}))) - \mathfrak{h}(h_1^{-1}(z_c(h_1, f_{m_\tau}, f_{m_\beta})))\right)\Bigg)$$

$$= \frac{1}{\sigma(v)}\Bigg(h_1(\mathfrak{h}^{-1}(z_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta})))^B - h_1(h_1^{-1}(z_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta})))^B + z_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta})^B$$

$$- z_c(h_1, f_{m_\tau}, f_{m_\beta})^B - B(z_c(h_1, f_{m_\tau}, f_{m_\beta}))^{B-1}\Bigg(\frac{\partial}{\partial t} z_c(f_{h,t}, f_{m_\tau}, f_{m_\beta})\Big|_{t=0}$$

$$+ h_1(h_1^{-1}(z_c(h_1, f_{m_\tau}, f_{m_\beta}))) - \mathfrak{h}(h_1^{-1}(z_c(h_1, f_{m_\tau}, f_{m_\beta})))\Bigg)\Bigg)$$

$$= \frac{1}{\sigma(v)}\Big(h_1(\mathfrak{h}^{-1}(z_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta})))^B - h_1(h_1^{-1}(z_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta})))^B$$

$$- Bz_c(h_1, f_{m_\tau}, f_{m_\beta})^{B-1}(h_1(h_1^{-1}(z_c(h_1, f_{m_\tau}, f_{m_\beta}))) - \mathfrak{h}(h_1^{-1}(z_c(h_1, f_{m_\tau}, f_{m_\beta}))))\Big)$$

$$+ \mathcal{O}\big(||\mathfrak{h} - h_1||^2_{[z_a, z_b]}\big),$$

because

$$z_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta})^B - z_c(0)^B - Bz_c(0)^{B-1}\frac{\partial}{\partial t} z_c(t)\Big|_{t=0} = \frac{\frac{\partial^2}{\partial t^2} z_c(f_{h,t}, f_{m_\tau}, f_{m_\beta})\big|_{t=\tilde{t}}}{2}$$

$$= \mathcal{O}\big(||\mathfrak{h} - h_1||^2_{[z_a, z_b]}\big)$$

for an appropriate $\tilde{t}$ in $(0, 1)$. Apply Lemma 4.6.5 to obtain

$$h_1(\mathfrak{h}^{-1}(z_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta})))^B - h_1(h_1^{-1}(z_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta})))^B$$

$$- Bz_c(h_1, f_{m_\tau}, f_{m_\beta})^{B-1}(h_1(h_1^{-1}(z_c(h_1, f_{m_\tau}, f_{m_\beta}))) - \mathfrak{h}(h_1^{-1}(z_c(h_1, f_{m_\tau}, f_{m_\beta}))))$$

$$= Bh_1(h_1^{-1}(z_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta})))^{B-1}(h_1(\mathfrak{h}^{-1}(z_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta}))) - h_1(h_1^{-1}(z_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta}))))$$

$$- Bz_c(h_1, f_{m_\tau}, f_{m_\beta})^{B-1}(h_1(h_1^{-1}(z_c(h_1, f_{m_\tau}, f_{m_\beta}))) - \mathfrak{h}(h_1^{-1}(z_c(h_1, f_{m_\tau}, f_{m_\beta}))))$$

$$+ \mathcal{O}(||\mathfrak{h} - h_1||^2_{[z_a, z_b]})$$

$$= Bz_c(h_1, f_{m_\tau}, f_{m_\beta})^{B-1}\Big(h_1(\mathfrak{h}^{-1}(z_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta}))) - h_1(h_1^{-1}(z_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta})))$$

$$- h_1(h_1^{-1}(z_c(h_1, f_{m_\tau}, f_{m_\beta}))) + \mathfrak{h}(h_1^{-1}(z_c(h_1, f_{m_\tau}, f_{m_\beta})))\Big) + \mathcal{O}(||\mathfrak{h} - h_1||^2_{[z_a, z_b]})$$

as well as

$$h_1(\mathfrak{h}^{-1}(z_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta}))) - h_1(h_1^{-1}(z_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta}))) - h_1(h_1^{-1}(z_c(h_1, f_{m_\tau}, f_{m_\beta})))$$

$$+ \mathfrak{h}(h_1^{-1}(z_c(h_1, f_{m_\tau}, f_{m_\beta})))$$

$$= h_1(\mathfrak{h}^{-1}(z_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta}))) - h_1(h_1^{-1}(z_c(h_1, f_{m_\tau}, f_{m_\beta})))$$

$$+ \mathfrak{h}(h_1^{-1}(z_c(h_1, f_{m_\tau}, f_{m_\beta}))) - \mathfrak{h}(\mathfrak{h}^{-1}(z_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta})))$$

$$= h_1'(h_1^{-1}(z_c(0)))(\mathfrak{h}^{-1}(z_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta})) - h_1^{-1}(z_c(h_1, f_{m_\tau}, f_{m_\beta})))$$

$$+ \mathfrak{h}'(h_1^{-1}(z_c(0)))(h_1^{-1}(z_c(h_1, f_{m_\tau}, f_{m_\beta})) - \mathfrak{h}^{-1}(z_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta}))) + \mathcal{O}(||\mathfrak{h} - h_1||^2_{[z_a, z_b]})$$

$$
\begin{aligned}
&= \big(h_1'(h_1^{-1}(z_c(0))) - \mathfrak{h}'(h_1^{-1}(z_c(0)))\big)\big(\mathfrak{h}^{-1}(z_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta})) - h_1^{-1}(z_c(h_1, f_{m_\tau}, f_{m_\beta}))\big) \\
&\quad + \mathcal{O}(||\mathfrak{h} - h_1||^2_{[z_a,z_b]}) \\
&= \mathcal{O}(||\mathfrak{h} - h_1||^{\frac{3}{2}}_{[z_a,z_b]}).
\end{aligned}
$$

It remains to treat the second and the third summand. Recall that it is sufficient to prove the equations (4.43) and (4.44), that is

$$
\begin{aligned}
&\big|k_c(h_1, f_{m_\tau}, f_{m_\beta}, v, e) - k_c(h_1, F^{-1}_{Y|X}(\tau|\cdot), f_{m_\beta}, v, e) \\
&- D_{f_{m_\tau}} k_c(h_1, F^{-1}_{Y|X}(\tau|\cdot), f_{m_\beta}, v, e)\big[f_{m_\tau} - F^{-1}_{Y|X}(\tau|\cdot)\big]\big| \\
&\quad \le C||f_{m_\tau} - F^{-1}_{Y|X}(\tau|\cdot)||^2_{M_X}
\end{aligned}
$$

(uniformly in $c \in B_\delta, v \in M_X, e \in [e_a, e_b]$ and f_{m_β}, such that $(h_1, f_{m_\tau}, f_{m_\beta}) \in \tilde{\mathcal{H}}$) and

$$
\begin{aligned}
&\big|k_c(h_1, F^{-1}_{Y|X}(\tau|\cdot), f_{m_\beta}, v, e) - k_c(s_0, v, e) - D_{f_{m_\beta}} k_c(s_0, v, e)\big[f_{m_\beta} - F^{-1}_{Y|X}(\beta|\cdot)\big]\big| \\
&\le C||f_{m_\beta} - F^{-1}_{Y|X}(\beta|\cdot)||^2_{M_X}
\end{aligned}
$$

uniformly in $c \in B_\delta, v \in M_X, e \in [e_a, e_b]$ for some appropriate $C > 0$. For that purpose, notice that

$$
\begin{aligned}
&z_c(h_1, f_{m_\tau}, f_{m_\beta}) - z_c(h_1, F^{-1}_{Y|X}(\tau|\cdot), f_{m_\beta}) \\
&= \big(h_1(f_{m_\tau}(v))^c + e(h_1(f_{m_\beta}(v))^c - h_1(f_{m_\tau}(v))^c)\big)^{\frac{1}{c}} \\
&\quad - \big(h_1(F^{-1}_{Y|X}(\tau|v))^c + e(h_1(f_{m_\beta}(v))^c - h_1(F^{-1}_{Y|X}(\tau|v))^c)\big)^{\frac{1}{c}} \\
&= \frac{1}{c}\big(h_1(F^{-1}_{Y|X}(\tau|v))^c + e(h_1(f_{m_\beta}(v))^c - h_1(F^{-1}_{Y|X}(\tau|v))^c)\big)^{\frac{1}{c}-1} \\
&\qquad c h_1(F^{-1}_{Y|X}(\tau|v))^{c-1}(1-e)h_1'(F^{-1}_{Y|X}(\tau|v))(f_{m_\tau}(v) - F^{-1}_{Y|X}(\tau|v)) \\
&\quad + \mathcal{O}(||f_{m_\tau} - F^{-1}_{Y|X}(\tau|\cdot)||^2_{M_X}) \\
&= -\frac{(1-e)z_c(h_1, F^{-1}_{Y|X}(\tau|\cdot), f_{m_\beta})^{1-c} h_1(F^{-1}_{Y|X}(\tau|v))^c (f_{m_\tau}(v) - F^{-1}_{Y|X}(\tau|v))}{\lambda(F^{-1}_{Y|X}(\tau|v))} \\
&\quad + \mathcal{O}(||f_{m_\tau} - F^{-1}_{Y|X}(\tau|\cdot)||^2_{M_X}).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&k_c(h_1, f_{m_\tau}, f_{m_\beta}, v, e) - k_c(h_1, F^{-1}_{Y|X}(\tau|\cdot), f_{m_\beta}, v, e) - D_{f_{m_\tau}} k_c(s_0, v, e)\big[f_{m_\tau} - F^{-1}_{Y|X}(\tau|\cdot)\big] \\
&= \frac{1}{\sigma(v)}\Bigg(z_c(h_1, f_{m_\tau}, f_{m_\beta})^B - z_c(h_1, F^{-1}_{Y|X}(\tau|\cdot), f_{m_\beta})^B \\
&\quad - \frac{B z_c(h_1, F^{-1}_{Y|X}(\tau|\cdot), f_{m_\beta})^{B-c}(1-e)h_1(F^{-1}_{Y|X}(\tau|v))^c (f_{m_\tau}(v) - F^{-1}_{Y|X}(\tau|v))}{\lambda(F^{-1}_{Y|X}(\tau|v))}\Bigg)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{\sigma(v)}\Bigg(Bz_c(h_1, F^{-1}_{Y|X}(\tau|\cdot), f_{m_\beta})^{B-1}(z_c(h_1, f_{m_\tau}, f_{m_\beta}) - z_c(h_1, F^{-1}_{Y|X}(\tau|\cdot), f_{m_\beta}))\\
&\quad - \frac{Bz_c(h_1, F^{-1}_{Y|X}(\tau|\cdot), f_{m_\beta})^{B-c}(1-e)h_1(F^{-1}_{Y|X}(\tau|v))^c(f_{m_\tau}(v) - F^{-1}_{Y|X}(\tau|v))}{\lambda(F^{-1}_{Y|X}(\tau|v))}\Bigg)\\
&\quad + \mathcal{O}(||f_{m_\tau} - F^{-1}_{Y|X}(\tau|\cdot)||^2_{M_X})\\
&= \mathcal{O}(||f_{m_\tau} - F^{-1}_{Y|X}(\tau|\cdot)||^2_{M_X}).
\end{aligned}
$$

Analogously,

$$
\begin{aligned}
&k_c(h_1, F^{-1}_{Y|X}(\tau|\cdot), f_{m_\beta}, v, e) - k_c(s_0, v, e) - D_{f_{m_\tau}} k_c(s_0, v, e)\big[f_{m_\beta} - F^{-1}_{Y|X}(\beta|\cdot)\big]\\
&= \frac{1}{\sigma(v)}\Bigg(z_c(h_1, F^{-1}_{Y|X}(\tau|\cdot), f_{m_\beta})^B - z_c(s_0)^B\\
&\quad - \frac{Bz_c(s_0)^{B-c}eh_1(F^{-1}_{Y|X}(\beta|v))^c(f_{m_\beta}(v) - F^{-1}_{Y|X}(\beta|v))}{\lambda(F^{-1}_{Y|X}(\beta|v))}\Bigg)\\
&= \frac{1}{\sigma(v)}\Bigg(Bz_c(s_0)^{B-1}(z_c(h_1, F^{-1}_{Y|X}(\tau|\cdot), f_{m_\beta}) - z_c(s_0))\\
&\quad - \frac{Bz_c(h_1, F^{-1}_{Y|X}(\tau|\cdot), f_{m_\beta})^{B-c}eh_1(F^{-1}_{Y|X}(\beta|v))^c(f_{m_\beta}(v) - F^{-1}_{Y|X}(\beta|v))}{\lambda(F^{-1}_{Y|X}(\beta|v))}\Bigg)\\
&\quad + \mathcal{O}(||f_{m_\beta} - F^{-1}_{Y|X}(\beta|\cdot)||^2_{M_X})\\
&= \mathcal{O}(||f_{m_\beta} - F^{-1}_{Y|X}(\beta|\cdot)||^2_{M_X}).
\end{aligned}
$$

Hence, (i) is proven.

Proof of (ii): Remember (**C4**) and let $c \in B_{\delta_n}$. As before, one has

$$
\begin{aligned}
&||D_s G_{MD}(c, s_0)(x, e)[\hat{s} - s_0] - D_s G_{MD}(B, s_0)(x, e)[\hat{s} - s_0]||\\
&= ||D_h G_{MD}(c, s_0)(x, e)[\bar{h}_1 - h_1] - D_h G_{MD}(B, s_0)(x, e)[\bar{h}_1 - h_1]\\
&\quad + D_{f_{m_\tau}} G_{MD}(c, s_0)(x, e)[\hat{f}_{m_\tau} - F^{-1}_{Y|X}(\tau|\cdot)]\\
&\quad - D_{f_{m_\tau}} G_{MD}(B, s_0)(x, e)[\hat{f}_{m_\tau} - F^{-1}_{Y|X}(\tau|\cdot)]\\
&\quad + D_{f_{m_\beta}} G_{MD}(c, s_0)(x, e)[\hat{f}_{m_\beta} - F^{-1}_{Y|X}(\beta|\cdot)]\\
&\quad - D_{f_{m_\beta}} G_{MD}(B, s_0)(x, e)[\hat{f}_{m_\beta} - F^{-1}_{Y|X}(\beta|\cdot)]||\\
&\leq ||D_h G_{MD}(c, s_0)(x, e)[\bar{h}_1 - h_1] - D_h G_{MD}(B, s_0)(x, e)[\bar{h}_1 - h_1]||\\
&\quad + ||D_{f_{m_\tau}} G_{MD}(c, s_0)(x, e)[\hat{f}_{m_\tau} - F^{-1}_{Y|X}(\tau|\cdot)]\\
&\quad - D_{f_{m_\tau}} G_{MD}(B, s_0)(x, e)[\hat{f}_{m_\tau} - F^{-1}_{Y|X}(\tau|\cdot)]||\\
&\quad + ||D_{f_{m_\beta}} G_{MD}(c, s_0)(x, e)[\hat{f}_{m_\beta} - F^{-1}_{Y|X}(\beta|\cdot)]
\end{aligned}
$$

$$- D_{f_{m_\beta}} G_{MD}(B, s_0)(x, e)[\hat{f}_{m_\beta} - F^{-1}_{Y|X}(\beta|\cdot)]||$$

so that it is again sufficient to prove the condition for each of the summands. To treat the first summand, let $\mathfrak{h} \in \tilde{\mathcal{H}}, c \in [B_1, B_2]$ and recall the definitions of $f_{h,t}$ and $z_c(t)$ from page 147 as well as

$$|z_c(0) - z_B(0)| = \mathcal{O}(|c - B|) \quad \text{uniformly in } (x, e)$$

and

$$\begin{aligned}
&D_h G_{MD}(c, s_0)(x, e)[\mathfrak{h} - h_1] \\
&= \frac{1}{P(X \in M)} \frac{\partial}{\partial t} \Big(\int_{M_X} F_\varepsilon(k_c(f_{h,t}, F^{-1}_{Y|X}(\tau|\cdot), F^{-1}_{Y|X}(\beta|\cdot), w, e)) I_{\{w \le x\}} f_X(w)\, dw \\
&\quad - P(X \le x | X \in M) \int_{M_X} F_\varepsilon(k_c(f_{h,t}, F^{-1}_{Y|X}(\tau|\cdot), F^{-1}_{Y|X}(\beta|\cdot), w, e)) f_X(w)\, dw \Big) \Big|_{t=0} \\
&= \frac{1}{P(X \in M)} \Big(\int_{M_X} f_\varepsilon(k_c(s_0, w, e)) \big(I_{\{w \le x\}} - P(X \le x | X \in M_X) \big) \\
&\quad D_h k_c(s_0, w, e)[\mathfrak{h} - h_1] f_X(w)\, dw \Big).
\end{aligned}$$

At the beginning of the proof of this lemma, it was shown in (4.39) that

$$\begin{aligned}
D_h k_c(s_0, x, e)[\mathfrak{h} - h_1] &= \frac{\partial}{\partial t} k_c(f_{h,t}, F^{-1}_{Y|X}(\tau|\cdot), F^{-1}_{Y|X}(\beta|\cdot), x, e) \Big|_{t=0} \\
&= \frac{\partial}{\partial t} \frac{h_1(\psi(t, z_c(t)))^B - g(x)}{\sigma(x)} \Big|_{t=0} \\
&= \frac{B h_1(\psi(t, z_c(t)))^{B-1} h_1'(\psi(t, z_c(t))) \frac{\partial}{\partial t} \psi(t, z_c(t))}{\sigma(x)} \Big|_{t=0} \\
&= \frac{B z_c(0)^{B-1} \big(z_c(0) - \mathfrak{h}(h_1^{-1}(z_c(0))) + \frac{\partial}{\partial t} z_c(t)\big|_{t=0} \big)}{\sigma(x)}
\end{aligned}$$

(for $f_{m_\tau} = F^{-1}_{Y|X}(\tau|\cdot), f_{m_\beta} = F^{-1}_{Y|X}(\beta|\cdot)$ in $z_c(t)$), where

$$\begin{aligned}
\frac{\partial}{\partial t} z_c(t) \Big|_{t=0} &= \frac{\partial}{\partial t} \big(f^c_{h,t}(F^{-1}_{Y|X}(\tau|x)) + e(f^c_{h,t}(F^{-1}_{Y|X}(\beta|x)) - f^c_{h,t}(F^{-1}_{Y|X}(\tau|x))) \big)^{\frac{1}{c}} \Big|_{t=0} \\
&= \frac{1}{c} \big(h_1^c(F^{-1}_{Y|X}(\tau|x)) + e(h_1^c(F^{-1}_{Y|X}(\beta|x)) - h_1^c(F^{-1}_{Y|X}(\tau|x))) \big)^{\frac{1}{c} - 1} \\
&\quad \big(c h_1^{c-1}(F^{-1}_{Y|X}(\tau|x)) (\mathfrak{h}(F^{-1}_{Y|X}(\tau|x)) - h_1(F^{-1}_{Y|X}(\tau|x))) \\
&\quad + e(c h_1^{c-1}(F^{-1}_{Y|X}(\beta|x)) (\mathfrak{h}(F^{-1}_{Y|X}(\beta|x)) - h_1(F^{-1}_{Y|X}(\beta|x)) \\
&\quad - c h_1^{c-1}(F^{-1}_{Y|X}(\tau|x)) (\mathfrak{h}(F^{-1}_{Y|X}(\tau|x)) - h_1(F^{-1}_{Y|X}(\tau|x))) \big).
\end{aligned}$$

Hence,

$$\sup_{x \in M_X, e \in [e_a, e_b]} \left| \frac{\partial}{\partial t} z_c(t) \Big|_{t=0} \right| = \mathcal{O}(||\mathfrak{h} - h_1||_{[z_a, z_b]})$$

and

$$\sup_{x \in M_X, e \in [e_a, e_b]} \left| \frac{\partial}{\partial t} z_c(t) \Big|_{t=0} - \frac{\partial}{\partial t} z_B(t) \Big|_{t=0} \right| = \mathcal{O}\big(||\mathfrak{h} - h_1||_{[z_a, z_b]} |c - B| \big),$$

so that

$$\begin{aligned}
&D_hG_{MD}(c,s_0)(x,e)[\mathfrak{h}-h_1]-D_hG_{MD}(B,s_0)(x,e)[\mathfrak{h}-h_1]\\
&=\int_{M_X}\big(I_{\{w\leq x\}}-P(X\leq x|X\in M_X)\big)\Big(\varphi(c,w,e)\big(z_c(0)-\mathfrak{h}(h_1^{-1}(z_c(0)))\big)\\
&\quad-\varphi(B,w,e)\big(z_B(0)-\mathfrak{h}(h_1^{-1}(z_B(0)))\big)\Big)\,dw+o\big(|c-B|\big)
\end{aligned}$$

for some continuously differentiable function $\varphi:[B_1,B_2]\times M_X\times[e_a,e_b]\to\mathbb{R}$. Due to

$$\begin{aligned}
&\varphi(c,w,e)\big(h_1(h_1^{-1}(z_c(0)))-\mathfrak{h}(h_1^{-1}(z_c(0)))\big)-\varphi(B,w,e)\big(h_1(h_1^{-1}(z_B(0)))-\mathfrak{h}(h_1^{-1}(z_B(0)))\big)\\
&=(\varphi(c,w,e)-\varphi(B,w,e))\big(h_1(h_1^{-1}(z_c(0)))-\mathfrak{h}(h_1^{-1}(z_c(0)))\big)\\
&\quad+\varphi(B,w,e)\big(h_1(h_1^{-1}(z_c(0)))-h_1(h_1^{-1}(z_B(0)))+\mathfrak{h}(h_1^{-1}(z_B(0)))-\mathfrak{h}(h_1^{-1}(z_c(0)))\big)\\
&=\varphi(B,w,e)\big(h_1'(h_1^{-1}(z_B(0)))(h_1^{-1}(z_c(0))-h_1^{-1}(z_B(0)))\\
&\quad-\mathfrak{h}'(h_1^{-1}(z_B(0)))(h_1^{-1}(z_c(0))-h_1^{-1}(z_B(0)))\big)+o(|c-B|)\\
&=\mathcal{O}(||\mathfrak{h}'-h_1'||_{[z_a,z_b]}|c-B|)+o(|c-B|)\\
&=o(|c-B|),
\end{aligned}$$

it holds that

$$||D_hG_{MD}(c,s_0)(x,e)[\bar{h}_1-h_1]-D_hG_{MD}(B,s_0)(x,e)[\bar{h}_1-h_1]||=o_p(|c-B|).$$

The second summand can be written as

$$\begin{aligned}
&||D_{f_{m_\tau}}G_{MD}(c,s_0)(x,e)[\hat{F}^{-1}_{Y|X}(\tau|\cdot)-F^{-1}_{Y|X}(\tau|\cdot)]\\
&-D_{f_{m_\tau}}G_{MD}(B,s_0)(x,e)[\hat{F}^{-1}_{Y|X}(\tau|\cdot)-F^{-1}_{Y|X}(\tau|\cdot)]\,||\\
&=\Bigg|\Bigg|\frac{1}{P(X\in M_X)}\int_{M_X}\big(I_{\{w\leq\cdot\}}-P(X\leq\cdot|X\in M_X)\big)\\
&\quad\big(f_\varepsilon(k_c(s_0,w,.))D_{f_{m_\tau}}k_c(s_0,w,.)[\hat{F}^{-1}_{Y|X}(\tau|\cdot)-F^{-1}_{Y|X}(\tau|\cdot)]\\
&\quad-f_\varepsilon(k_B(s_0,w,.))D_{f_{m_\tau}}k_B(s_0,w,.)[\hat{F}^{-1}_{Y|X}(\tau|\cdot)-F^{-1}_{Y|X}(\tau|\cdot)]\big)f_X(w)\,dw\Bigg|\Bigg|.
\end{aligned}$$

Thus, it is sufficient to prove

$$\begin{aligned}
&f_\varepsilon(k_c(s_0,w,e))D_{f_{m_\tau}}k_c(s_0,w,e)[\hat{F}^{-1}_{Y|X}(\tau|\cdot)-F^{-1}_{Y|X}(\tau|\cdot)]\\
&-f_\varepsilon(k_B(s_0,w,e))D_{f_{m_\tau}}k_B(s_0,w,e)[\hat{F}^{-1}_{Y|X}(\tau|\cdot)-F^{-1}_{Y|X}(\tau|\cdot)]\\
&=\big(f_\varepsilon(k_c(s_0,w,e))-f_\varepsilon(k_B(s_0,w,e))\big)D_{f_{m_\tau}}k_c(s_0,w,e)[\hat{F}^{-1}_{Y|X}(\tau|\cdot)-F^{-1}_{Y|X}(\tau|\cdot)]\\
&\quad+f_\varepsilon(k_B(s_0,w,e))\big(D_{f_{m_\tau}}k_c(s_0,w,e)[\hat{F}^{-1}_{Y|X}(\tau|\cdot)-F^{-1}_{Y|X}(\tau|\cdot)]\\
&\quad-D_{f_{m_\tau}}k_B(s_0,w,e)[\hat{F}^{-1}_{Y|X}(\tau|\cdot)-F^{-1}_{Y|X}(\tau|\cdot)]\big)
\end{aligned}$$

$$= o_p(|c-B|) + \mathcal{O}_p\big(n^{-\frac{1}{2}}\big)$$

uniformly in $(w,e) \in M_X \times [e_a, e_b]$. By condition (**C4**), $\hat{f}_{m_\tau}(x) - F_{Y|X}^{-1}(\tau|x) = o_p(1)$ uniformly in $x \in M_X$, so that for an appropriate $\tilde{c}$ between c and B

$$\begin{aligned}
&\big(f_\varepsilon(k_c(s_0,w,e)) - f_\varepsilon(k_B(s_0,w,e))\big) D_{f_{m_\tau}} k_c(s_0,w,e)[\hat{f}_{m_\tau} - F_{Y|X}^{-1}(\tau|\cdot)] \\
&= f_\varepsilon'(k_{\tilde{c}}(s_0,w,e)) \frac{\partial}{\partial c} k_c(s_0,w,e)\Big|_{c=\tilde{c}} (c-B) o_p(1) \\
&= o_p(|c-B|).
\end{aligned}$$

On the other hand, the remaining term can be rewritten via

$$\begin{aligned}
&D_{f_{m_\tau}} k_c(s_0,w,e)[[\hat{f}_{m_\tau} - F_{Y|X}^{-1}(\tau|\cdot)] - D_{f_{m_\tau}} k_B(s_0,w,.)[\hat{f}_{m_\tau} - F_{Y|X}^{-1}(\tau|\cdot)] \\
&= \frac{B(e-1)(\hat{f}_{m_\tau}(w) - F_{Y|X}^{-1}(\tau|w))}{\sigma(w)\lambda(F_{Y|X}^{-1}(\tau|w))} (\psi(B,w,e) - \psi(c,w,e)),
\end{aligned}$$

where

$$\psi(c,w,e) = \big(h_c(F_{Y|X}^{-1}(\tau|w)) + e\big(h_c(F_{Y|X}^{-1}(\beta|w)) - h_c(F_{Y|X}^{-1}(\tau|w))\big)\big)^{\frac{B}{c}-1} h_c(F_{Y|X}^{-1}(\tau|x)).$$

Due to

$$\begin{aligned}
&\frac{\partial}{\partial c}\psi(c,w,e) \\
&= \frac{\partial}{\partial c}\big(h_c(F_{Y|X}^{-1}(\tau|w)) + e\big(h_c(F_{Y|X}^{-1}(\beta|w)) - h_c(F_{Y|X}^{-1}(\tau|w))\big)\big)^{\frac{B}{c}-1} h_c(F_{Y|X}^{-1}(\tau|x)) \\
&= -\frac{B}{c^2}\log\big(h_c(F_{Y|X}^{-1}(\tau|w)) + e\big(h_c(F_{Y|X}^{-1}(\beta|w)) - h_c(F_{Y|X}^{-1}(\tau|w))\big)\big)\psi(c,w,e) \\
&\quad + \left(\frac{B}{c}-1\right) h_c(F_{Y|X}^{-1}(\tau|x))\Big(h_c(F_{Y|X}^{-1}(\tau|w)) + e\big(h_c(F_{Y|X}^{-1}(\beta|w)) - h_c(F_{Y|X}^{-1}(\tau|w))\big)\Big)^{\frac{B}{c}-2} \\
&\qquad \Bigg(\log\big(h_1(F_{Y|X}^{-1}(\tau|w))\big)h_c(F_{Y|X}^{-1}(\tau|w)) + e\Big(\log\big(h_1(F_{Y|X}^{-1}(\beta|w))\big)h_c(F_{Y|X}^{-1}(\beta|w)) \\
&\qquad - \log\big(h_1(F_{Y|X}^{-1}(\tau|w))\big)h_c(F_{Y|X}^{-1}(\tau|w))\Big)\Bigg) + \psi(c,w,e)\log\big(h_1(F_{Y|X}^{-1}(\beta|w))\big),
\end{aligned}$$

the derivative of ψ with respect to c is uniformly bounded in $(w,e) \in M_X \times [e_a, e_b]$. Hence,

$$\begin{aligned}
&D_{f_{m_\tau}} k_c(s_0,w,e)[\hat{f}_{m_\tau} - F_{Y|X}^{-1}(\tau|\cdot)] - D_{f_{m_\tau}} k_B(s_0,w,e)[\hat{f}_{m_\tau} - F_{Y|X}^{-1}(\tau|\cdot)] \\
&= o_p(|c-B|)
\end{aligned}$$

uniformly in $(w,e) \in M_X \times [e_a, e_b]$. The same reasoning can be applied for

$$D_{f_{m_\beta}} G_{MD}(c,s_0)(x,e)[\hat{f}_{m_\beta} - F_{Y|X}^{-1}(\beta|\cdot)],$$

which completes the proof of Lemma 4.6.4.

Lemma 4.6.6 *It holds that*

$$\sup_{||c-B||\leq\delta_n,||s-s_0||\leq\delta_n} ||G_{nMD}(c,s)-G_{MD}(c,s)-G_{nMD}(B,s_0)||_2 = o_p(n^{-\frac{1}{2}}),$$

*that is, **(C5)** is valid.*

Proof: In a moment, it will be shown that the process

$$G_n(c,s,x,e) = G_{nMD}(c,s)(x,e) - G_{MD}(c,s)(x,e),$$

as a process in $c\in[B_1,B_2], s\in\tilde{\mathcal{H}}, x\in M_X, e\in[e_a,e_b]$ is Donsker. Then, Corollary 2.3.12 of Van der Vaart and Wellner (1996) yields

$$\sup_{\substack{c,\tilde{c}\in[B_1,B_2],s,\tilde{s}\in\tilde{\mathcal{H}},x\in M_X,e\in[e_a,e_b]\\ ||s-\tilde{s}||_{\mathcal{H}}<\delta_n,|c-\tilde{c}|<\delta_n}} \sqrt{n}|G_n(c,s,x,e)-G_n(\tilde{c},\tilde{s},x,e)| = o_p(1).$$

Due to $G_{MD}(B,s_0)(x,e)=0$ for all $x\in M_X, e\in[e_a,e_b]$ the assertion then follows from the compactness of M_X and $[e_a,e_b]$.
First, define the function class

$$\mathcal{F} = \{(X,\varepsilon)\mapsto I_{\{X\in M_X\}}I_{\{\varepsilon\leq k_c(\mathfrak{h},f_{m_\tau},f_{m_\beta},X,e)\}} : s\in\tilde{\mathcal{H}}, c\in[B_1,B_2], e\in[e_a,e_b]\}.$$

Due to the definition of $\tilde{\mathcal{H}}$ in (4.24) and the compactness of $[B_1,B_2], M_X, [e_a,e_b]$, there exists a compact set $\mathcal{K}$ such that

$$k_c(s,x,e)\in\mathcal{K}, \quad \text{for all } s\in\tilde{\mathcal{H}}, x\in M_X, e\in[e_a,e_b].$$

Consider $s,\tilde{s}\in\tilde{\mathcal{H}}, c,\tilde{c}\in[B_1,B_2], e,\tilde{e}\in[e_a,e_b]$. For some s^*, c^* and e^* between s and $\tilde{s}$, c and $\tilde{c}$ and e and $\tilde{e}$, respectively, as well as some $C>0$ the L^2-distance can be bounded by

$$\begin{aligned}
&||I_{\{\cdot\in M_X\}}I_{\{\cdot\leq k_c(s,\cdot,e)\}} - I_{\{\cdot\in M_X\}}I_{\{\cdot\leq k_{\tilde{c}}(\tilde{s},\cdot,\tilde{e})\}}||_2\\
&= E\big[I_{\{X\in M_X\}}\big(I_{\{\varepsilon\leq k_c(s,X,e)\}} - I_{\{\varepsilon\leq k_{\tilde{c}}(\tilde{s},X,\tilde{e})\}}\big)^2\big]^{\frac{1}{2}}\\
&= \bigg(\int_{M_X}|F_\varepsilon(k_c(s,w,e)) - F_\varepsilon(k_{\tilde{c}}(\tilde{s},w,\tilde{e}))|f_X(w)\,dw\bigg)^{\frac{1}{2}}\\
&\leq \sup_{e\in\mathcal{K}}|f_\varepsilon(e)|\bigg(\int_{M_X}|f_X(w)|\,dw\bigg)^{\frac{1}{2}} \sup_{w\in M_X}\big|k_c(s,w,e)-k_{\tilde{c}}(\tilde{s},w,\tilde{e})\big|^{\frac{1}{2}}\\
&\leq \sup_{e\in\mathcal{K}}|f_\varepsilon(e)|\bigg(\int_{M_X}|f_X(w)|\,dw\bigg)^{\frac{1}{2}} \sup_{w\in M_X}\Big|D_hk_c(s^*,w,e^*)[\tilde{\mathfrak{h}}-\mathfrak{h}]\\
&\qquad + D_{f_{m_\tau}}k_c(s^*,w,e^*)[\tilde{f}_{m_\tau}-f_{m_\tau}] + D_{f_{m_\beta}}k_c(s^*,w,e^*)[\tilde{f}_{m_\beta}-f_{m_\beta}]\\
&\qquad + D_ek_c(s^*,w,e^*)[\tilde{e}-e]| + D_ck_{c^*}(s^*,w,e^*)[\tilde{c}-c]\Big|^{\frac{1}{2}}.
\end{aligned}$$

Similar to the the proof of (4.39), one can show with

$$s^* = (\mathfrak{h}^*, f^*_{m_\tau}, f^*_{m_\beta}),$$

$$f_{h,t} = \mathfrak{h}^* + t(\tilde{\mathfrak{h}} - \mathfrak{h}),$$

$$\tilde{z}_c(t) = \big(f_{h,t}^c(f_{m_\tau}^*(x)) + e(f_{h,t}^c(f_{m_\beta}^*(x)) - f_{h,t}^c(f_{m_\tau}^*(x)))\big)^{\frac{1}{c}}$$

that

$$D_h k_c(s^*, w, e)[\tilde{\mathfrak{h}} - \mathfrak{h}]$$

$$= \frac{Bh_1\big((h^*)^{-1}(\tilde{z}_c(0))\big)^{B-1} h_1'\big((h^*)^{-1}(z_c(0))\big)\Big(\frac{\partial}{\partial t}\tilde{z}_c(t)\big|_{t=0} - \big(\tilde{\mathfrak{h}} - \mathfrak{h}\big)\big((h^*)^{-1}(z_c(0))\big)\Big)}{(h^*)'\big(h_1^{-1}(z_c(0))\big)\sigma(w)}$$

and that

$$\sup_{c\in[B_1,B_2], w\in M_X, e\in[e_a,e_b]} |D_h k_c(s^*, w, e)[\tilde{\mathfrak{h}} - \mathfrak{h}]| \leq \tilde{C}||\tilde{\mathfrak{h}} - \mathfrak{h}||_{[z_a,z_b]}$$

for an appropriate constant $\tilde{C} > 0$ and all $s, \tilde{s} \in \tilde{\mathcal{H}}$ and s^* between s and $\tilde{s}$. A similar reasoning for $D_{f_{m_\tau}} k_{c^*}(s^*, w, e^*)[\tilde{f}_{m_\tau} - f_{m_\tau}], ..., D_{c^*} k_{c^*}(s^*, w, e^*)[\tilde{c} - c]$ leads to

$$\sup_{w\in M_X} \big|k_c(s, w, e) - k_{\tilde{c}}(\tilde{s}, w, \tilde{e})\big|$$

$$\leq \bar{C}\Big(||\tilde{\mathfrak{h}} - \mathfrak{h}||_{[z_a,z_b]} + ||\tilde{f}_{m_\tau} - f_{m_\tau}||_{M_X} + ||\tilde{f}_{m_\beta} - f_{m_\beta}||_{M_X} + |\tilde{e} - e| + |\tilde{c} - c|\Big) \quad (4.45)$$

for some appropriate constant $\bar{C} > 0$, which is independent of c^*, s^*, e^*. This will be used in the following to define brackets for $\mathcal{F}$. Let $\xi, \eta > 0$ and consider ξ-brackets for $c \in [B_1, B_2], e \in [e_a, e_b]$ and $\mathfrak{h}, f_{m_\tau}, f_{m_\beta}$ such that $s \in \tilde{\mathcal{H}}$. Construct η-brackets for $\mathcal{F}$ as follows. Let

$$\xi = \xi(\eta) = \frac{\eta^2}{10\bar{C}\sup_{e\in\mathcal{K}} f_\varepsilon(e)^2 \int |f_X(w)|\, dw}$$

with $\bar{C}$ from (4.45). For each combination of the ξ-brackets take representatives $\bar{\mathfrak{h}}, \bar{f}_{m_\tau}, \bar{f}_{m_\beta}$, $\bar{c}, \bar{e}$ within these brackets and define

$$l(X, \varepsilon) = I\{X \in M_X\} I\left\{\varepsilon \leq k_{\bar{c}}(\bar{\mathfrak{h}}, \bar{f}_{m_\tau}, \bar{f}_{m_\beta}, X, \bar{e}) - \frac{\eta^2}{2\sup_{e\in\mathcal{K}} f_\varepsilon(e)^2 \int |f_X(w)|\, dw}\right\}$$

and

$$u(X, \varepsilon) = I\{X \in M_X\} I\left\{\varepsilon \leq k_{\bar{c}}(\bar{\mathfrak{h}}, \bar{f}_{m_\tau}, \bar{f}_{m_\beta}, X, \bar{e}) + \frac{\eta^2}{2\sup_{e\in\mathcal{K}} f_\varepsilon(e)^2 \int |f_X(w)|\, dw}\right\}.$$

Then, $||u - l||_2 \leq \eta$ by the same reasoning as above and equation (4.45) ensures that each combination of the $\xi(\eta)$-brackets for $\mathfrak{h}, f_{m_\tau}, f_{m_\beta}$ such that $s \in \tilde{\mathcal{H}}$ and $c \in [B_1, B_2], e \in [e_a, e_b]$ is covered by its corresponding $[l, u]$-bracket.
Since $\tilde{\mathcal{H}} \subseteq C_{R_h}^{\gamma_h}([z_a, z_b]) \times C_{R_{f_{m_\tau}}}^{\gamma_{f_{m_\tau}}}(M_X) \times C_{R_{f_{m_\tau}}}^{\gamma_{f_{m_\beta}}}(M_X)$ one has for all $\eta > 0$

$$\mathcal{N}_{[]}(\eta, \mathcal{F}, L^2(P)) \leq \mathcal{N}_{[]}\big(\xi(\eta), C_{R_h}^{\gamma_h}([z_a, z_b]), ||.||_{[z_a,z_b]}\big)\mathcal{N}_{[]}\big(\xi(\eta), C_{R_{f_{m_\tau}}}^{\gamma_{f_{m_\tau}}}(M_X), ||.||_{M_X}\big)$$

$$\mathcal{N}_{[]}\big(\xi(\eta), C_{R_{f_{m_\beta}}}^{\gamma_{f_{m_\beta}}}(M_X), ||.||_{M_X}\big)\mathcal{N}_{[]}\big(\xi(\eta), [e_a, e_b], |.|\big)$$

$$\mathcal{N}_{[]}\big(\xi(\eta), [B_1, B_2], |.|\big).$$

According to Theorem 2.7.1 of Van der Vaart and Wellner (1996), one has

$$\log\left(\mathcal{N}_{[]}\left(\xi, C_{R_h}^{\gamma_h}([z_a, z_b]), ||.||_{[z_a,z_b]}\right)\right) \leq C_h \xi^{-\frac{1}{\gamma_h}},$$

$$\log\left(\mathcal{N}_{[]}\left(\xi, C_{R_{f_{m_\tau}}}^{\gamma_{f_{m_\tau}}}(M_X), ||.||_{M_X}\right)\right) \leq C_{f_{m_\tau}} \xi^{-\frac{d_X}{\gamma_{f_{m_\tau}}}},$$

$$\log\left(\mathcal{N}_{[]}\left(\xi, C_{R_{f_{m_\beta}}}^{\gamma_{f_{m_\beta}}}(M_X), ||.||_{M_X}\right)\right) \leq C_{f_{m_\beta}} \xi^{-\frac{d_X}{\gamma_{f_{m_\beta}}}}$$

for some appropriate constants $C_h, C_{f_{m_\tau}}, C_{f_{m_\beta}} > 0$. Note that

$$\gamma_h > 1, \quad \gamma_{f_{m_\tau}} > d_X \quad \text{and} \quad \gamma_{f_{m_\beta}} > d_X$$

by definition of $\tilde{\mathcal{H}}$ in (4.24). Hence, for some $C > 0$

$$\begin{aligned}
&\int_0^1 \sqrt{\log\left(\mathcal{N}_{[]}(\eta, \mathcal{F}, L^2(P))\right)}\, d\eta \\
&\leq \int_0^1 \sqrt{\log\left(\mathcal{N}_{[]}\left(\xi(\eta), C_{R_h}^{\gamma_h}([z_a, z_b]), ||.||_{[z_a,z_b]}\right)\right)}\, d\eta \\
&\quad + \int_0^1 \sqrt{\log\left(\mathcal{N}_{[]}\left(\xi(\eta), C_{R_{f_{m_\tau}}}^{\gamma_{f_{m_\tau}}}(M_X), ||.||_{M_X}\right)\right)}\, d\eta \\
&\quad + \int_0^1 \sqrt{\log\left(\mathcal{N}_{[]}\left(\xi(\eta), C_{R_{f_{m_\beta}}}^{\gamma_{f_{m_\beta}}}(M_X), ||.||_{M_X}\right)\right)}\, d\eta \\
&\quad + \int_0^1 \sqrt{\log\left(\mathcal{N}_{[]}\left(\xi(\eta), [e_a, e_b], |.|\right)\right)}\, d\eta \\
&\quad + \int_0^1 \sqrt{\log\left(\mathcal{N}_{[]}\left(\xi(\eta), [B_1, B_2], |.|\right)\right)}\, d\eta \\
&\leq C \int_0^1 \left(\left(\frac{1}{\eta}\right)^{\frac{1}{\gamma_h}} + \left(\frac{1}{\eta}\right)^{\frac{d_X}{\gamma_{f_{m_\tau}}}} + \left(\frac{1}{\eta}\right)^{\frac{d_X}{\gamma_{f_{m_\beta}}}} + \log\left(\frac{1}{\eta^2}\right) + \log\left(\frac{1}{\eta^2}\right)\right) d\eta \\
&< \infty,
\end{aligned}$$

so that the function class $\mathcal{F}$ is Donsker. Of course, the function class $\{X \mapsto I_{\{X \leq x\}} : x \in M_X\}$ is Donsker and by the same reasoning as before it can be shown that the class

$$\tilde{\mathcal{F}} = \{(X, \varepsilon) \mapsto I_{\{X \leq x\}} I_{\{X \in M_X\}} I_{\{\varepsilon \leq k_c(s,X,e)\}} : s \in \tilde{\mathcal{H}}, c \in [B_1, B_2], e \in [e_a, e_b], x \in M_X\}$$

is Donsker as well.
Finally,

$$\begin{aligned}
&G_{nMD}(c, s)(x, e) - G_{MD}(c, s)(x, e) \\
&= \hat{P}(X \leq x, \varepsilon \leq k_c(s, X, e) | X \in M_X) - \hat{P}(X \leq x | X \in M_X) \hat{P}(\varepsilon \leq k_c(s, X, e) | X \in M_X) \\
&\quad - P(X \leq x, \varepsilon \leq k_c(s, X, e) | X \in M_X) + P(X \leq x | X \in M_X) P(\varepsilon \leq k_c(s, X, e) | X \in M_X) \\
&= \hat{P}(X \leq x, \varepsilon \leq k_c(s, X, e) | X \in M_X) - P(X \leq x, \varepsilon \leq k_c(s, X, e) | X \in M_X)
\end{aligned}$$

$$
\begin{aligned}
&- \hat{P}(\varepsilon \le k_c(s,X,e)|X \in M_X)(\hat{P}(X \le x|X \in M_X) - P(X \le x|X \in M_X)) \\
&- P(X \le x|X \in M_X)(\hat{P}(\varepsilon \le k_c(s,X,e)|X \in M_X) - P(\varepsilon \le k_c(s,X,e)|X \in M_X)) \\
= &\frac{\frac{1}{n}\sum_{i=1}^n I_{\{X_i \le x\}} I_{\{X_i \in M_X\}} I_{\{\varepsilon_i \le k_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta}, X_i, e)\}}}{\frac{1}{n}\sum_{i=1}^n I_{\{X_i \in M_X\}}} - P(X \le x, \varepsilon \le k_c(s,X,e)|X \in M_X) \\
&- P(\varepsilon \le k_c(s,X,e)|X \in M_X)(\hat{P}(X \le x|X \in M_X) - P(X \le x|X \in M_X)) \\
&- P(X \le x|X \in M_X)(\hat{P}(\varepsilon \le k_c(s,X,e)|X \in M_X) - P(\varepsilon \le k_c(s,X,e)|X \in M_X)) + o_p\Big(\frac{1}{\sqrt{n}}\Big) \\
= &\left(\frac{1}{\frac{1}{n}\sum_{i=1}^n I_{\{X_i \in M_X\}}} - \frac{1}{P(X \in M_X)}\right)\frac{1}{n}\sum_{i=1}^n I_{\{X_i \le x\}} I_{\{X_i \in M_X\}} I_{\{\varepsilon_i \le k_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta}, X_i, e)\}} \\
&+ \frac{1}{P(X \in M_X)}\left(\frac{1}{n}\sum_{i=1}^n I_{\{X_i \le x\}} I_{\{X_i \in M_X\}} I_{\{\varepsilon_i \le k_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta}, X_i, e)\}}\right. \\
&\quad \left. - P(X \in M_X, X \le x, \varepsilon \le k_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta}, X, e))\right) \\
&- P(\varepsilon \le k_c(s,X,e)|X \in M_X)(\hat{P}(X \le x|X \in M_X) - P(X \le x|X \in M_X)) \\
&- P(X \le x|X \in M_X)(\hat{P}(\varepsilon \le k_c(s,X,e)|X \in M_X) - P(\varepsilon \le k_c(s,X,e)|X \in M_X)) + o_p\Big(\frac{1}{\sqrt{n}}\Big) \\
= &-\frac{P(X \le x, X \in M_X, \varepsilon \le k_c(s,X,e))}{nP(X \in M_X)^2}\sum_{i=1}^n (I_{\{X_i \in M_X\}} - P(X \in M_X)) \\
&+ \frac{1}{P(X \in M_X)}\left(\frac{1}{n}\sum_{i=1}^n I_{\{X_i \le x\}} I_{\{X_i \in M_X\}} I_{\{\varepsilon_i \le k_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta}, X_i, e)\}}\right. \\
&\quad \left. - P(X \in M_X, X \le x, \varepsilon \le k_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta}, X, e))\right) \\
&- P(\varepsilon \le k_c(s,X,e)|X \in M_X)(\hat{P}(X \le x|X \in M_X) - P(X \le x|X \in M_X)) \\
&- P(X \le x|X \in M_X)(\hat{P}(\varepsilon \le k_c(s,X,e)|X \in M_X) - P(\varepsilon \le k_c(s,X,e)|X \in M_X)) + o_p\Big(\frac{1}{\sqrt{n}}\Big) \\
= &-\frac{P(X \le x, \varepsilon \le k_c(s,X,e)|X \in M_X) - 2P(X \le x|X \in M_X)P(\varepsilon \le k_c(s,X,e)|X \in M_X)}{P(X \in M_X)} \\
&\left(\frac{1}{n}\sum_{i=1}^n (I_{\{X_i \in M_X\}} - P(X \in M_X))\right) \\
&+ \frac{1}{P(X \in M_X)}\frac{1}{n}\sum_{i=1}^n \Big(I_{\{X_i \le x\}} I_{\{X_i \in M_X\}} I_{\{\varepsilon_i \le k_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta}, X_i, e)\}} \\
&\quad - P(X \in M_X, X \le x, \varepsilon \le k_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta}, X, e)) \\
&\quad - I_{\{X_i \le x\}} I_{\{X_i \in M_X\}} + P(X \in M_X, X \le x) - I_{\{X_i \in M_X\}} I_{\{\varepsilon_i \le k_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta}, X_i, e)\}} \\
&\quad + P(X \in M_X, \varepsilon \le k_c(\mathfrak{h}, f_{m_\tau}, f_{m_\beta}, X, e))\Big) + o_p\Big(\frac{1}{\sqrt{n}}\Big),
\end{aligned}
$$

so that because of Corollary 2.3.1 of Van der Vaart and Wellner (1996),

$$\sup_{\substack{c,\tilde{c}\in[B_1,B_2],s,\tilde{s}\in\tilde{\mathcal{H}},x\in M_X,e\in[e_a,e_b]\\ ||s-\tilde{s}||_{\mathcal{H}}<\delta_n,|c-\tilde{c}|<\delta_n}} \sqrt{n}|G_n(c,s,x,e)-G_n(\tilde{c},\tilde{s},x,e)| = o_p(1).$$

□

Putting things together:

Similar to Linton et al. (2008), define

$$L_n(x,e) = G_{nMD}(B,s_0)(x,e) - G_{MD}(B,s_0)(x,e)$$

as well as

$$\mathcal{L}_n(c)(x,e) = L_n(x,e) + \Gamma_1(B,s_0)(x,e)(c-B) + \Gamma_2(B,s_0)(x,e)[\hat{s}-s_0].$$

In the proof of Lemma 4.6.6, it was shown that

$$||L_n||_2 = ||G_{nMD}(B,s_0)||_2 = \mathcal{O}_p\left(\frac{1}{\sqrt{n}}\right). \tag{4.46}$$

Then, one has for all sequences $\delta_n \searrow 0$

$$\begin{aligned}
&||G_{nMD}(c,\hat{s}) - \mathcal{L}_n(c)||_2\\
&= ||G_{nMD}(c,\hat{s}) - G_{nMD}(B,s_0) + G_{MD}(B,s_0) - \Gamma_1(B,s_0)(c-B) - \Gamma_2(B,s_0)[\hat{s}-s_0]||_2\\
&\overset{\textbf{(C1)}}{\leq} ||G_{nMD}(c,\hat{s}) - G_{MD}(c,\hat{s}) - G_{nMD}(B,s_0)||_2\\
&\quad + ||G_{MD}(c,\hat{s}) - \Gamma_1(B,s_0)(c-B) - \Gamma_2(B,s_0)[\hat{s}-s_0]||_2\\
&\overset{\textbf{(C5)}}{=} ||G_{MD}(c,\hat{s}) - \Gamma_1(B,s_0)(c-B) - \Gamma_2(B,s_0)[\hat{s}-s_0]||_2 + o_p\left(\frac{1}{\sqrt{n}}\right)\\
&\leq ||G_{MD}(c,\hat{s}) - G_{MD}(c,s_0) - \Gamma_2(B,s_0)[\hat{s}-s_0]||_2 + ||G_{MD}(c,s_0) - \Gamma_1(B,s_0)(c-B)||_2\\
&\quad + o_p\left(\frac{1}{\sqrt{n}}\right)\\
&\overset{\textbf{(C3)}}{\leq} ||G_{MD}(c,\hat{s}) - G_{MD}(c,s_0) - \Gamma_2(c,s_0)[\hat{s}-s_0]||_2 + ||G_{MD}(c,s_0) - \Gamma_1(B,s_0)(c-B)||_2\\
&\quad + o_p(|c-B|) + o_p\left(\frac{1}{\sqrt{n}}\right)\\
&\overset{\textbf{(C3)}}{\leq} ||G_{MD}(c,s_0) - G_{MD}(B,s_0) - \Gamma_1(B,s_0)(c-B)||_2 + o_p(|c-B|) + o_p\left(\frac{1}{\sqrt{n}}\right)\\
&\overset{\textbf{(C2)}}{=} o_p(|c-B|) + o_p\left(\frac{1}{\sqrt{n}}\right)
\end{aligned} \tag{4.47}$$

uniformly in $c \in B_{\delta_n}$. Denote the minimizer of $c \mapsto ||\mathcal{L}_n(c)||_2$ by $\bar{B}$. Then, $\bar{B}$ can be calculated explicitly by solving

$$\frac{\partial}{\partial c}||\mathcal{L}_n(c)||_2^2$$

$$
\begin{aligned}
&= \frac{\partial}{\partial c} ||L_n + \Gamma_1(B, s_0)(c - B) + \Gamma_2(B, s_0)[\hat{s} - s_0]||_2^2 \\
&= \frac{\partial}{\partial c} \bigg(||L_n||_2^2 + ||\Gamma_1(B, s_0)||_2^2 (c - B)^2 + ||\Gamma_2(B, s_0)[\hat{s} - s_0]||_2^2 \\
&\quad + 2 \int_{M_X} \int_{[e_a, e_b]} \Gamma_1(B, s_0)(x, e)(L_n(x, e) + \Gamma_2(B, s_0)(x, e)[\hat{s} - s_0]) \, de \, dx (c - B) \\
&\quad + 2 \int_{M_X} \int_{[e_a, e_b]} L_n(x, e) \Gamma_2(B, s_0)(x, e)[\hat{s} - s_0] \, de \, dx \bigg) \\
&= 2 ||\Gamma_1(B, s_0)||_2^2 (c - B) \\
&\quad + 2 \int_{M_X} \int_{[e_a, e_b]} \Gamma_1(B, s_0)(x, e)(L_n(x, e) + \Gamma_2(B, s_0)(x, e)[\hat{s} - s_0]) \, de \, dx \\
&= 0.
\end{aligned}
$$

Therefore, one has

$$
\begin{aligned}
\bar{B} &= B - \frac{\int_{M_X} \int_{[e_a, e_b]} \Gamma_1(B, s_0)(x, e)(L_n(x, e) + \Gamma_2(B, s_0)(x, e)[\hat{s} - s_0]) \, de \, dx}{||\Gamma_1(B, s_0)||_2^2} \\
&\overset{(\mathbf{C6})}{=} B + \mathcal{O}_p\Big(\frac{1}{\sqrt{n}}\Big).
\end{aligned} \tag{4.48}
$$

Proof of (i) in Theorem 4.2.4: One has

$$
||\mathcal{L}_n(\bar{B})||_2 \leq ||\mathcal{L}_n(B)||_2 \leq ||L_n(x, e)||_2 + ||\Gamma_2(B, s_0)[\hat{s} - s_0]||_2 \overset{(4.26)+(4.46)}{=} \mathcal{O}_p\Big(\frac{1}{\sqrt{n}}\Big),
$$

$$
||G_{nMD}(\hat{B}, \hat{s})||_2 \leq ||G_{nMD}(B, \hat{s})||_2 \overset{(4.47)}{\leq} ||\mathcal{L}_n(B)||_2 + o_p\Big(\frac{1}{\sqrt{n}}\Big) = \mathcal{O}_p\Big(\frac{1}{\sqrt{n}}\Big)
$$

and for all $c \in B_{\delta_n}$ and by a Taylor expansion with some B^* between c and $\bar{B}$

$$
\begin{aligned}
||\mathcal{L}_n(c)||_2^2 &= ||\mathcal{L}_n(\bar{B})||_2^2 + \frac{\partial}{\partial c} ||\mathcal{L}_n(c)||_2^2 \Big|_{c = \bar{B}} (c - \bar{B}) + \frac{\frac{\partial^2}{\partial c^2} ||\mathcal{L}_n(c)||_2^2 \big|_{c = B^*}}{2} (c - \bar{B})^2 \\
&= ||\mathcal{L}_n(\bar{B})||_2^2 + ||\Gamma_1(B, s_0)||_2^2 (c - \bar{B})^2.
\end{aligned} \tag{4.49}
$$

These assertions in turn can be used to obtain

$$
\begin{aligned}
&||G_{nMD}(\hat{B}, \hat{s})||_2^2 \\
&\quad \leq ||G_{nMD}(\bar{B}, \hat{s})||_2^2 \\
&\quad \overset{(4.47)}{=} \big(||\mathcal{L}_n(\bar{B})||_2 + o_p(|\bar{B} - B|) + o_p\big(n^{-\frac{1}{2}}\big) \big)^2 \\
&\quad = ||\mathcal{L}_n(\bar{B})||_2^2 + ||\mathcal{L}_n(\bar{B})||_2 o_p\big(n^{-\frac{1}{2}}\big) + o_p(n^{-1}) \\
&\quad \overset{(4.49)}{=} ||\mathcal{L}_n(\hat{B})||_2^2 - ||\Gamma_1(B, s_0)||_2^2 (\hat{B} - \bar{B})^2 + o_p(n^{-1}) \\
&\quad \overset{(4.47)}{=} \big(||G_{nMD}(\hat{B}, \hat{s})||_2 + o_p(|\hat{B} - B|) + o_p\big(n^{-\frac{1}{2}}\big) \big)^2 - ||\Gamma_1(B, s_0)||_2^2 (\hat{B} - \bar{B})^2 + o_p(n^{-1})
\end{aligned}
$$

$$\overset{(4.48)}{=} \big(||G_{nMD}(\hat{B},\hat{s})||_2 + o_p(|\hat{B}-\bar{B}|) + o_p\big(n^{-\frac{1}{2}}\big)\big)^2 - ||\Gamma_1(B,s_0)||_2^2(\hat{B}-\bar{B})^2 + o_p(n^{-1})$$
$$= ||G_{nMD}(\hat{B},\hat{s})||_2^2 - ||\Gamma_1(B,s_0)||_2^2(\hat{B}-\bar{B})^2 + o_p\big(n^{-\frac{1}{2}}|\hat{B}-\bar{B}|\big) + o_p(|\hat{B}-\bar{B}|^2)$$
$$+ o_p(n^{-1}).$$

Thus,

$$||\Gamma_1(B,s_0)||_2^2(\hat{B}-\bar{B})^2 = o_p\big(n^{-\frac{1}{2}}|\hat{B}-\bar{B}|\big) + o_p(|\hat{B}-\bar{B}|^2) + o_p(n^{-1})$$

and consequently $\hat{B}-\bar{B} = o_p(n^{-\frac{1}{2}})$. Finally, (**C6**) yields

$$\sqrt{n}(\hat{B}-B)$$
$$= \sqrt{n}(\bar{B}-B) + o_p(1)$$
$$= -\frac{\sqrt{n}\int_{M_X}\int_{[e_a,e_b]}\Gamma_1(B,s_0)(x,e)(L_n(x,e) + \Gamma_2(B,s_0)(x,e)[\hat{s}-s_0])\,de\,dx}{||\Gamma_1(B,s_0)||_2^2} + o_p(1)$$
$$= -\frac{\sqrt{n}\int_{M_X}\int_{[e_a,e_b]}\Gamma_1(B,s_0)(x,e)(G_{nMD}(x,e) + \Gamma_2(B,s_0)(x,e)[\hat{s}-s_0])\,de\,dx}{||\Gamma_1(B,s_0)||_2^2} + o_p(1)$$
$$\xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{\sigma_A^2}{||\Gamma_1(B,s_0)||_2^4}\right). \tag{4.50}$$

Proof of (ii) in Theorem 4.2.4: The reasoning is similar to before, although due to

$$||\mathcal{L}_n(\bar{B})||_2 \le ||\mathcal{L}_n(B)||_2 \le ||L_n(x,e)||_2 + ||\Gamma_2(B,s_0)[\hat{s}-s_0]||_2 = \mathcal{O}_p(a_n)$$

and

$$||G_{nMD}(\hat{B},\hat{s})||_2 \le ||G_{nMD}(B,\hat{s})||_2 \le ||\mathcal{L}_n(B)||_2 + o_p\left(\frac{1}{\sqrt{n}}\right) = \mathcal{O}_p(a_n)$$

the orders of the negligible terms change:

$$||G_{nMD}(\hat{B},\hat{s})||_2^2$$
$$\le ||G_{nMD}(\bar{B},\hat{s})||_2^2$$
$$= \big(||\mathcal{L}_n(\bar{B})||_2 + o_p(|\bar{B}-B|) + o_p\big(n^{-\frac{1}{2}}\big)\big)^2$$
$$= ||\mathcal{L}_n(\bar{B})||_2^2 + ||\mathcal{L}_n(\bar{B})||_2 o_p\big(n^{-\frac{1}{2}}\big) + o_p(n^{-1})$$
$$= ||\mathcal{L}_n(\hat{B})||_2^2 - ||\Gamma_1(B,s_0)||_2^2(\hat{B}-\bar{B})^2 + o_p\big(a_n n^{-\frac{1}{2}}\big)$$
$$= \big(||G_{nMD}(\hat{B},\hat{s})||_2 + o_p(|\hat{B}-B|) + o_p\big(n^{-\frac{1}{2}}\big)\big)^2 - ||\Gamma_1(B,s_0)||_2^2(\hat{B}-\bar{B})^2 + o_p\big(a_n n^{-\frac{1}{2}}\big)$$
$$= ||G_{nMD}(\hat{B},\hat{s})||_2^2 - ||\Gamma_1(B,s_0)||_2^2(\hat{B}-\bar{B})^2 + o_p\big(a_n|\hat{B}-\bar{B}|\big) + o_p(|\hat{B}-\bar{B}|^2)$$
$$+ o_p\big(a_n n^{-\frac{1}{2}}\big).$$

Therefore, $\hat{B}-\bar{B} = o_p(a_n)$ and

$$\hat{B}-B = \hat{B}-\bar{B}+\bar{B}-B \overset{(4.48)}{=} o_p(a_n).$$

4.6.4 Proof of Lemma 4.2.3

First, validity of assumptions (**C4**) and (**C6'**) is proven in the following two Lemmas. Note that (**C6**) follows from (**C6'**) by the Central Limit Theorem.

Lemma 4.6.7 *One has $\hat{s} \in \tilde{\mathcal{H}}$ with probability converging to one, $||\hat{h}_1 - h_1||^{\frac{3}{2}}_{[z_a, z_b]} = o_p(n^{-\frac{1}{2}})$ and $||\hat{F}^{-1}_{Y|X}(\tau|\cdot) - F^{-1}_{Y|X}(\tau|\cdot)||_{M_X}, ||\hat{F}^{-1}_{Y|X}(\beta|\cdot) - F^{-1}_{Y|X}(\beta|\cdot)||_{M_X} = o_p(n^{-\frac{1}{4}})$.*

Proof: Recall the definition of $\tilde{\mathcal{H}} \subseteq \mathcal{H}$ (remember $\gamma_h = \gamma_{f_{m_\tau}} = \gamma_{f_{m_\beta}} = 2$):

$$\tilde{\mathcal{H}} = \left\{ s \in \mathcal{H} : h \in C^2_{R_h}([z_a, z_b]), f_{m_\tau} \in C^2_{R_{f_{m_\tau}}}(M_X), f_{m_\beta} \in C^2_{R_{f_{m_\beta}}}(M_X) \right\}.$$

The convergence rates directly follow from Lemma 4.2.15 and Theorem 4.2.2.
To prove $\hat{s} \in \tilde{\mathcal{H}}$ with probability converging to one it suffices to show uniform convergence of the functions $\hat{h}_1, \hat{F}^{-1}_{Y|X}(\tau|\cdot)$, $\hat{F}^{-1}_{Y|X}(\beta|\cdot)$ and their derivatives up to order two to $h_1, F^{-1}_{Y|X}(\tau|\cdot), F^{-1}_{Y|X}(\beta|\cdot)$ and the corresponding derivatives, respectively. Without loss of generality, when proving $\hat{F}^{-1}_{Y|X}(\tau|\cdot) \in C^{\gamma_{f_{m_\tau}}}_{R_{f_{m_\tau}}}(M_X)$ and $\hat{F}^{-1}_{Y|X}(\beta|\cdot) \in C^{\gamma_{f_{m_\beta}}}_{R_{f_{m_\beta}}}(M_X)$ only derivatives with respect to x_1 are considered since other derivatives can be treated analogously. For $\hat{h}_1$ this follows from Corollary 4.2.14, since

$$\frac{\partial}{\partial y}\hat{h}_1(y) = \frac{\partial}{\partial y}\exp\bigg(- \int_{y_1}^{y} \frac{1}{\hat{\lambda}(u)}\, du \bigg) = -\frac{\hat{h}_1(y)}{\hat{\lambda}(y)}$$

and

$$\frac{\partial^2}{\partial y^2}\hat{h}_1(y) = -\frac{\hat{\lambda}(y)\frac{\partial}{\partial y}\hat{h}_1(y) - \hat{h}_1(y)\frac{\partial}{\partial y}\hat{\lambda}(y)}{\hat{\lambda}(y)^2} = \frac{\hat{h}_1(y) + \hat{h}_1(y)\frac{\partial}{\partial y}\hat{\lambda}(y)}{\hat{\lambda}(y)^2}.$$

As will be seen in the following, the assertion for $\hat{F}^{-1}_{Y|X}(\tau|\cdot)$ and $\hat{F}^{-1}_{Y|X}(\beta|\cdot)$ follows from the corresponding assertion for $\hat{F}_{Y|X}(y|\cdot)$ and hence can be deduced from Corollary 4.2.14 and Lemma 4.2.15 as well. One has

$$\tau = \hat{F}_{Y|X}(\hat{F}^{-1}_{Y|X}(\tau|x)|x),$$

so that

$$0 = \frac{\partial}{\partial x_1}\hat{F}_{Y|X}(\hat{F}^{-1}_{Y|X}(\tau|x)|x) = \hat{F}_y(\hat{F}^{-1}_{Y|X}(\tau|x)|x)\frac{\partial}{\partial x_1}\hat{F}^{-1}_{Y|X}(\tau|x) + \hat{F}_x(\hat{F}^{-1}_{Y|X}(\tau|x)|x),$$

where F_y and F_x denote the derivative of $(y, x) \mapsto \hat{F}_{Y|X}(y|x)$ with respect to y and x, respectively. Note that $d_X = 1$ was assumed. Therefore,

$$\frac{\partial}{\partial x_1}\hat{F}^{-1}_{Y|X}(\tau|x) = -\frac{\hat{F}_x(\hat{F}^{-1}_{Y|X}(\tau|x)|x)}{\hat{F}_y(\hat{F}^{-1}_{Y|X}(\tau|x)|x)}.$$

Corollary 4.2.14 and Lemma 4.2.15 lead to

$$\sup_{x \in M_X}\left|\frac{\partial}{\partial x_1}\hat{F}^{-1}_{Y|X}(\tau|x) - \frac{\partial}{\partial x_1}F^{-1}_{Y|X}(\tau|x)\right| = o_p(1).$$

The second derivative can be written as

$$\frac{\partial^2}{\partial x_1^2}\hat{F}^{-1}_{Y|X}(\tau|x)$$

$$= -\frac{\hat{F}_y(\hat{F}_{Y|X}^{-1}(\tau|x)|x)\frac{\partial}{\partial x_1}\hat{F}_x(\hat{F}_{Y|X}^{-1}(\tau|x)|x) - \hat{F}_x(\hat{F}_{Y|X}^{-1}(\tau|x)|x)\frac{\partial}{\partial x_1}\hat{F}_y(\hat{F}_{Y|X}^{-1}(\tau|x)|x)}{\hat{F}_y(\hat{F}_{Y|X}^{-1}(\tau|x)|x)^2}.$$

Similar to before, let F_{yy}, F_{xy} and F_{xx} denote the partial derivatives of $(y,x) \mapsto \hat{F}_{Y|X}(y|x)$ of order two. Then, it holds that

$$\frac{\partial}{\partial x_1}\hat{F}_y(\hat{F}_{Y|X}^{-1}(\tau|x)|x) = \hat{F}_{xy}(\hat{F}_{Y|X}^{-1}(\tau|x)|x) + \hat{F}_{yy}(\hat{F}_{Y|X}^{-1}(\tau|x)|x)\frac{\partial}{\partial x_1}\hat{F}_{Y|X}^{-1}(\tau|x)$$

as well as

$$\frac{\partial}{\partial x_1}\hat{F}_x(\hat{F}_{Y|X}^{-1}(\tau|x)|x) = \hat{F}_{xx}(\hat{F}_{Y|X}^{-1}(\tau|x)|x) + \hat{F}_{xy}(\hat{F}_{Y|X}^{-1}(\tau|x)|x)\frac{\partial}{\partial x_1}\hat{F}_{Y|X}^{-1}(\tau|x),$$

so that again

$$\sup_{x\in M_X}\left|\frac{\partial^2}{\partial x_1^2}\hat{F}_{Y|X}^{-1}(\tau|x) - \frac{\partial^2}{\partial x_1^2}F_{Y|X}^{-1}(\tau|x)\right| = o_p(1)$$

is implied by Corollary 4.2.14. □

Lemma 4.6.8 *For some $\sigma_A^2 > 0$, one has*

$$\sqrt{n}\int_{M_X}\int_{[e_a,e_b]}\Gamma_1(B,s_0)(x,e)\big(G_{nMD}(B,s_0)(x,e) + \Gamma_2(B,s_0)(x,e)[\hat{s}-s_0]\big)\,de\,dx$$
$$\xrightarrow{\mathcal{D}} \mathcal{N}(0,\sigma_A^2).$$

Proof: The concept of the proof is quite simple. First, the left hand side is rewritten such that

$$\sqrt{n}\int_{M_X}\int_{[e_a,e_b]}\Gamma_1(B,s_0)(x,e)\big(G_{nMD}(B,s_0)(x,e) + \Gamma_2(B,s_0)(x,e)[\hat{s}-s_0]\big)\,de\,dx$$
$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\psi(Y_i,X_i) - E[\psi(Y,X)]) + o_p(1) \tag{4.51}$$

for some appropriate function $\psi : \mathbb{R}^{d_X+1} \to \mathbb{R}$. Afterwards, the usual Central Limit Theorem can be applied to obtain the desired convergence.
For this purpose, note that in equation (4.40) it was shown that

$$\begin{aligned}\Gamma_2(B,s_0)(x,e)[\hat{s}-s_0] &= D_sG_{MD}(B,s_0)(x,e)[\hat{s}-s_0]\\ &= D_hG_{MD}(B,s_0)(x,e)[\hat{h}-h_1]\\ &\quad + D_{f_{m_\tau}}G_{MD}(B,s_0)(x,e)[\hat{F}_{Y|X}^{-1}(\tau|\cdot) - F_{Y|X}^{-1}(\tau|\cdot)]\\ &\quad + D_{f_{m_\beta}}G_{MD}(B,s_0)(x,e)[F_{Y|X}^{-1}(\beta|\cdot) - F_{Y|X}^{-1}(\beta|\cdot)].\end{aligned}$$

Hence, there are actually four terms that have to be fitted to Expression (4.51).

Rewriting $\mathbf{G_{nMD}(B,s_0)}$: For $c = B$ and $s = s_0 = (h_1, F_{Y|X}^{-1}(\tau|\cdot), F_{Y|X}^{-1}(\beta|\cdot))$ one has

$$k_B(s_0,x,e) = \frac{g(x) + \sigma(x)F_\varepsilon^{-1}(\tau) + e\sigma(x)(F_\varepsilon^{-1}(\beta) - F_\varepsilon^{-1}(\tau)) - g(x)}{\sigma(x)}$$

$$= F_\varepsilon^{-1}(\tau) + e(F_\varepsilon^{-1}(\beta) - F_\varepsilon^{-1}(\tau)),$$

which is independent of x. Using the expression derived in the proof of Lemma 4.6.6, this leads to

$$\begin{aligned}
&G_{nMD}(B, s_0)(x, e)\\
&= -\frac{P(X \le x, \varepsilon \le k_B(s_0, X, e)|X \in M_X) - 2P(X \le x|X \in M_X)P(\varepsilon \le k_B(s_0, X, e)|X \in M_X)}{P(X \in M_X)}\\
&\quad \left(\frac{1}{n}\sum_{i=1}^n (I_{\{X_i \in M_X\}} - P(X \in M_X))\right)\\
&\quad + \frac{1}{P(X \in M_X)}\frac{1}{n}\sum_{i=1}^n \Big(I_{\{X_i \le x\}} I_{\{X_i \in M_X\}} I_{\{\varepsilon_i \le k_B(s_0, X_i, e)\}}\\
&\quad - P(X \in M_X, X \le x, \varepsilon \le k_B(s_0, X, e)) - I_{\{X_i \le x\}} I_{\{X_i \in M_X\}}\\
&\quad + P(X \in M_X, X \le x) - I_{\{X_i \in M_X\}} I_{\{\varepsilon_i \le k_B(s_0, X_i, e)\}} + P(X \in M_X, \varepsilon \le k_B(s_0, X, e))\Big)\\
&\quad + o_p\left(\frac{1}{\sqrt{n}}\right)\\
&= \frac{P(X \le x|X \in M_X)P(\varepsilon \le F_\varepsilon^{-1}(\tau) + e(F_\varepsilon^{-1}(\beta) - F_\varepsilon^{-1}(\tau)))}{P(X \in M_X)}\frac{1}{n}\sum_{i=1}^n (I_{\{X_i \in M_X\}} - P(X \in M_X))\\
&\quad + \frac{1}{P(X \in M_X)}\frac{1}{n}\sum_{i=1}^n \Big(I_{\{X_i \le x\}} I_{\{X_i \in M_X\}} I_{\{\varepsilon_i \le F_\varepsilon^{-1}(\tau) + e(F_\varepsilon^{-1}(\beta) - F_\varepsilon^{-1}(\tau))\}}\\
&\quad - P(X \in M_X, X \le x, \varepsilon \le F_\varepsilon^{-1}(\tau) + e(F_\varepsilon^{-1}(\beta) - F_\varepsilon^{-1}(\tau))) - I_{\{X_i \le x\}} I_{\{X_i \in M_X\}}\\
&\quad + P(X \in M_X, X \le x) - I_{\{X_i \in M_X\}} I_{\{\varepsilon_i \le F_\varepsilon^{-1}(\tau) + e(F_\varepsilon^{-1}(\beta) - F_\varepsilon^{-1}(\tau))\}}\\
&\quad + P(X \in M_X, \varepsilon \le F_\varepsilon^{-1}(\tau) + e(F_\varepsilon^{-1}(\beta) - F_\varepsilon^{-1}(\tau)))\Big) + o_p\left(\frac{1}{\sqrt{n}}\right),
\end{aligned}$$

that is,

$$\begin{aligned}
&\sqrt{n}\int_{M_X}\int_{[e_a, e_b]} \Gamma_1(B, s_0)(x, e) G_{nMD}(B, s_0)(x, e)\, de\, dx\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n (\psi_1(Y_i, X_i) - E[\psi_1(Y_i, X_i)]) + o_p(1)
\end{aligned}$$

for

$$\begin{aligned}
&\psi_1(Y, X)\\
&= \int_{M_X}\int_{[e_a, e_b]} \Bigg(I_{\{X \in M_X\}} \Gamma_1(B, s_0)(x, e)\frac{P(X \le x|X \in M_X)P(\varepsilon \le F_\varepsilon^{-1}(\tau) + e(F_\varepsilon^{-1}(\beta) - F_\varepsilon^{-1}(\tau)))}{P(X \in M_X)}\\
&\quad + \frac{1}{P(X \in M_X)}\int_{M_X}\int_{[e_a, e_b]} \Gamma_1(B, s_0)(x, e)\\
&\quad \Big(I_{\{X \le x\}} I_{\{X \in M_X\}} I_{\left\{\frac{h_B(Y) - g(X)}{\sigma(X)} \le F_\varepsilon^{-1}(\tau) + e(F_\varepsilon^{-1}(\beta) - F_\varepsilon^{-1}(\tau))\right\}} - I_{\{X \le x\}} I_{\{X \in M_X\}}\\
&\quad - I_{\{X \in M_X\}} I_{\left\{\frac{h_B(Y) - g(X)}{\sigma(X)} \le F_\varepsilon^{-1}(\tau) + e(F_\varepsilon^{-1}(\beta) - F_\varepsilon^{-1}(\tau))\right\}}\Big)\Bigg)\, de\, dx.
\end{aligned}$$

Rewriting $\mathbf{D_h G_{MD}(B, s_0)[\hat{h} - h_1]}$: Again, as was shown in the proof of Lemma 4.6.4 it holds that

$$
\begin{aligned}
&D_h G_{MD}(B,s_0)(x,e)[\hat{h}-h_1] \\
&= \frac{1}{P(X\in M)}\Big(\int_{M_X} f_\varepsilon(k_B(s_0,w,e))\big(I_{\{w\le x\}}-P(X\le x|X\in M_X)\big) \\
&\quad D_h k_B(s_0,w,e)[\hat{h}-h_1]f_X(w)\,dw\Big),
\end{aligned}
$$

where (for $f_{m_\tau}=F^{-1}_{Y|X}(\tau|\cdot), f_{m_\beta}=F^{-1}_{Y|X}(\beta|\cdot)$ in $z_c(t)$)

$$
D_h k_B(s_0,w,e)[\hat{h}-h_1] = \frac{B z_B(0)^{B-1}\big(z_B(0)-\hat{h}(h_1^{-1}(z_B(0)))+\frac{\partial}{\partial t}z_B(t)\big|_{t=0}\big)}{\sigma(w)}, \tag{4.52}
$$

$$
\begin{aligned}
z_B(0) &= \big(g(w)+\sigma(w)F_\varepsilon^{-1}(\tau)+e\sigma(w)(F_\varepsilon^{-1}(\beta)-F_\varepsilon^{-1}(\tau))\big)^{\frac{1}{B}}, \\
\frac{\partial}{\partial t}z_B(t)\big|_{t=0} &= B\big(g(w)+\sigma(w)F_\varepsilon^{-1}(\tau)+e\sigma(w)(F_\varepsilon^{-1}(\beta)-F_\varepsilon^{-1}(\tau))\big)^{\frac{1}{B}-1} \\
&\quad \Big(Bh_{B-1}(F^{-1}_{Y|X}(\tau|w))(\hat{h}(F^{-1}_{Y|X}(\tau|w))-h_1(F^{-1}_{Y|X}(\tau|w))) \\
&\qquad + e\big(Bh_{B-1}(F^{-1}_{Y|X}(\beta|w))(\hat{h}(F^{-1}_{Y|X}(\beta|w))-h_1(F^{-1}_{Y|X}(\beta|w))) \\
&\qquad - Bh_{B-1}(F^{-1}_{Y|X}(\tau|w))(\hat{h}(F^{-1}_{Y|X}(\tau|w))-h_1(F^{-1}_{Y|X}(\tau|w)))\big)\Big).
\end{aligned} \tag{4.53}
$$

Recall from the proof of Theorem 4.2.2 that (see (4.36))

$$
\begin{aligned}
\hat{h}_1(y)-h_1(y) &= \exp\Big(-\int_{y_1}^{y}\frac{1}{\hat{\lambda}(u)}\,du\Big)-\exp\Big(-\int_{y_1}^{y}\frac{1}{\lambda(u)}\,du\Big) \\
&= -h_1(y)\int_{y_1}^{y}\Big(\frac{1}{\hat{\lambda}(u)}-\frac{1}{\lambda(u)}\Big)\,du+o_p\Big(\frac{1}{\sqrt{n}}\Big) \\
&= \frac{1}{n}\sum_{i=1}^{n}-h_1(y)\eta_i(y)+o_p\Big(\frac{1}{\sqrt{n}}\Big),
\end{aligned}
$$

where

$$
\begin{aligned}
\eta_i(y) = &\int_{y_1}^{y}\frac{-1}{\lambda(u)^2}\Big(v(X_i)D_{p,0}(u,X_i)+\frac{\partial v(X_i)D_{p,x}(u,X_i)}{\partial x_1}\Big)I_{\{u\ge Y_i\}}\,du \\
&-\frac{v(X_i)D_{p,y}(Y_i,X_i)}{\lambda(Y_i)^2}\big(I_{\{Y_i\le y\}}-I_{\{Y_i\le y_1\}}\big) \\
&+\int_{y_1}^{y}\frac{-1}{\lambda(u)^2}\Big(v(X_i)D_{f,0}(u,X_i)+\frac{\partial v(X_i)D_{f,x}(u,X_i)}{\partial x_1}\Big)\,du
\end{aligned}
$$

and $E[\eta_1(y)]=0$ uniformly in $[z_a,z_b]$. Therefore,

$$
\begin{aligned}
&D_h k_B(s_0,w,e)[\hat{h}-h_1] \\
&= \frac{Bz_B(0)^{B-1}}{n\sigma(w)}\sum_{i=1}^{n}\Big(z_B(0)\eta_i(h_1^{-1}(z_B(0)))+B\big(g(w)+\sigma(w)F_\varepsilon^{-1}(\tau)
\end{aligned}
$$

$$
\begin{aligned}
&- e\sigma(w)(F_\varepsilon^{-1}(\beta) - F_\varepsilon^{-1}(\tau)))^{\frac{1}{B}-1}\Big(Bh_B(F_{Y|X}^{-1}(\tau|w))\eta_i(F_{Y|X}^{-1}(\tau|w)) \\
&\quad + e\big(Bh_B(F_{Y|X}^{-1}(\beta|w))\eta_i(F_{Y|X}^{-1}(\beta|w)) - Bh_B(F_{Y|X}^{-1}(\tau|w))\eta_i(F_{Y|X}^{-1}(\tau|w))\big)\Big) \\
&+ o_p\Big(\frac{1}{\sqrt{n}}\Big) \\
&= \frac{1}{n}\sum_{i=1}^n \tilde{\psi}_2(Y_i, X_i, w, e) + o_p\Big(\frac{1}{\sqrt{n}}\Big)
\end{aligned}
$$

for an appropriate function $\tilde{\psi}_2$, which is centred and uniformly bounded in $(Y, X, w, e) \in \mathbb{R}^{d_X+1} \times M_X \times [e_a, e_b]$. Thus,

$$
\begin{aligned}
&\sqrt{n}\int_{M_X}\int_{[e_a,e_b]} \Gamma_1(B, s_0)(x, e) D_h G_{MD}(B, s_0)(x, e)[\hat{h} - h_1]\, de\, dx \\
&= \frac{1}{nP(X \in M)}\sum_{i=1}^n \int_{M_X}\int_{[e_a,e_b]}\int_{M_X} \Gamma_1(B, s_0)(x, e) f_\varepsilon(k_B(s_0, w, e)) \\
&\quad \big(I_{\{w \le x\}} - P(X \le x | X \in M_X)\big)\tilde{\psi}_2(Y_i, X_i, w, e) f_X(w)\, dw\, de\, dx + o_p(1) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_2(Y_i, X_i) + o_p(1)
\end{aligned}
$$

for

$$
\begin{aligned}
\psi_2(Y_i, X_i) = &\frac{1}{P(X \in M)}\int_{M_X}\int_{[e_a,e_b]}\int_{M_X} \Gamma_1(B, s_0)(x, e) f_\varepsilon(k_B(s_0, w, e)) \\
&\big(I_{\{w \le x\}} - P(X \le x | X \in M_X)\big)\tilde{\psi}_2(Y_i, X_i, w, e) f_X(w)\, dw\, de\, dx.
\end{aligned}
$$

$\mathbf{D_{f_{m_\tau}} G_{MD}(B, s_0)}[\hat{\mathbf{F}}^{-1}_{\mathbf{Y|X}}(\tau|\cdot) - \mathbf{F}^{-1}_{\mathbf{Y|X}}(\tau|\cdot)]$ and $\mathbf{D_{f_{m_\tau}} G_{MD}(B, s_0)}[\hat{\mathbf{F}}^{-1}_{\mathbf{Y|X}}(\beta|\cdot) - \mathbf{F}^{-1}_{\mathbf{Y|X}}(\beta|\cdot)]$: In the proof of Lemma 4.6.6, it was shown that

$$
\begin{aligned}
&D_{f_{m_\tau}} k_B(s_0, x, e)[\hat{F}_{Y|X}^{-1}(\tau|\cdot) - F_{Y|X}^{-1}(\tau|\cdot)] \\
&\quad = -\frac{B(1-e)h_B(F_{Y|X}^{-1}(\tau|x))(\hat{F}_{Y|X}^{-1}(\tau|x) - F_{Y|X}^{-1}(\tau|x))}{\sigma(x)\lambda(F_{Y|X}^{-1}(\tau|x))}
\end{aligned}
$$

and

$$
D_{f_\beta} k_B(s_0, x, e)[\hat{F}_{Y|X}^{-1}(\beta|\cdot) - F_{Y|X}^{-1}(\beta|\cdot)] = -\frac{Beh_B(F_{Y|X}^{-1}(\beta|x))(\hat{F}_{Y|X}^{-1}(\beta|x) - F_{Y|X}^{-1}(\beta|x))}{\sigma(x)\lambda(F_{Y|X}^{-1}(\beta|x))}.
$$

Referring to equation (4.67) in the proof of Lemma (4.2.15) below, one has

$$
\begin{aligned}
&\hat{F}_{Y|X}^{-1}(\iota|x) - F_{Y|X}^{-1}(\iota|x) \\
&= \frac{1}{f_{Y|X}(F_{Y|X}^{-1}(\iota|x)|x) f_X(x)}\frac{1}{n}\sum_{i=1}^n \mathbf{K}_{h_x}(x - X_i)\Big(\mathcal{K}_{h_y}(F_{Y|X}^{-1}(\iota|x) - Y_i) \\
&\qquad - \frac{p(F_{Y|X}^{-1}(\iota|x), x)}{f_X(x)}\Big) + o_p\Big(\frac{1}{\sqrt{n}}\Big)
\end{aligned}
$$

for $\iota \in \{\tau, \beta\}$. Note that the order of the remaining term, that is obtained in the proof there, is actually $o_p(\hat{F}^{-1}_{Y|X}(\iota|x) - F^{-1}_{Y|X}(\iota|x))$, but this order can be extended to

$$\mathcal{O}_p((\hat{F}^{-1}_{Y|X}(\iota|x) - F^{-1}_{Y|X}(\iota|x))^2) = o_p(n^{-\frac{1}{2}})$$

similarly to the proof of Lemma 2.8.1 by using the Lagrange form of the remainder.
Due to Lemma 1.1.1, one has (compare (2.48))

$$E\bigg[\mathbf{K}_{h_x}(x - X_1)\bigg(\mathcal{K}_{h_y}(F^{-1}_{Y|X}(\iota|x) - Y_1) - \frac{p(F^{-1}_{Y|X}(\iota|x), x)}{f_X(x)}\bigg)\bigg] = o_p\bigg(\frac{1}{\sqrt{n}}\bigg)$$

uniformly in $x \in M_X$ and $\iota \in \{\tau, \beta\}$. So far, a representation

$$\begin{aligned}&\sqrt{n}\int_{M_X}\int_{[e_a,e_b]} \Gamma_1(B, s_0)(x, e)\big(G_{nMD}(B, s_0)(x, e) + \Gamma_2(B, s_0)(x, e)[\hat{s} - s_0]\big)\, de\, dx \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big(\psi_1(Y_i, X_i) + \psi_2(Y_i, X_i) + \psi_{3,n}(Y_i, X_i) + \psi_{4,n}(Y_i, X_i) \\ &\quad - E[\psi_1(Y, X) - \psi_2(Y, X) - \psi_{3,n}(Y, X) - \psi_{4,n}(Y, X)]\big) + o_p(1),\end{aligned}$$

could be found, where

$$\begin{aligned}\psi_{3,n}(Y_i, X_i) = -\int_{M_X}\int_{[e_a,e_b]} &\Gamma_1(B, s_0)(x, e)\frac{B(1-e)h_B(F^{-1}_{Y|X}(\tau|x))}{\sigma(x)\lambda(F^{-1}_{Y|X}(\tau|x))f_{Y|X}(F^{-1}_{Y|X}(\tau|x)|x)f_X(x)} \\ &\mathbf{K}_{h_x}(x - X_i)\bigg(\mathcal{K}_{h_y}(F^{-1}_{Y|X}(\tau|x) - Y_i) - \frac{p(F^{-1}_{Y|X}(\tau|x), x)}{f_X(x)}\bigg)\, de\, dx\end{aligned}$$

and

$$\begin{aligned}\psi_{4,n}(Y_i, X_i) = -\int_{M_X}\int_{[e_a,e_b]} &\Gamma_1(B, s_0)(x, e)\frac{Beh_B(F^{-1}_{Y|X}(\beta|x))}{\sigma(x)\lambda(F^{-1}_{Y|X}(\beta|x))f_{Y|X}(F^{-1}_{Y|X}(\beta|x)|x)f_X(x)} \\ &\mathbf{K}_{h_x}(x - X_i)\bigg(\mathcal{K}_{h_y}(F^{-1}_{Y|X}(\tau|x) - Y_i) - \frac{p(F^{-1}_{Y|X}(\beta|x), x)}{f_X(x)}\bigg)\, de\, dx\end{aligned}$$

depend on n. To fit this expression to equation (4.51) it suffices to replace $\psi_{3,n}$ and $\psi_{4,n}$ with some functions ψ_3 and ψ_4 (independent of n and with finite second moments), respectively, such that

$$E[(\psi_{3,n}(Y, X) - \psi_3(Y, X))^2] = o(1) \quad \text{and} \quad E[(\psi_{4,n}(Y, X) - \psi_4(Y, X))^2] = o(1),$$

since it was already shown that $E[\psi_1(Y, X)^2], E[\psi_2(Y, X)^2] < \infty$. For this purpose, define

$$\psi_\tau(x, e) = -\frac{B(1-e)h_B(F^{-1}_{Y|X}(\tau|x))}{\sigma(x)\lambda(F^{-1}_{Y|X}(\tau|x))f_{Y|X}(F^{-1}_{Y|X}(\tau|x)|x)f_X(x)},$$

$$\psi_\beta(x, e) = -\frac{Beh_B(F^{-1}_{Y|X}(\beta|x))}{\sigma(x)\lambda(F^{-1}_{Y|X}(\beta|x))f_{Y|X}(F^{-1}_{Y|X}(\beta|x)|x)f_X(x)},$$

$$\psi_3(Y, X) = I_{\{X \in M_X\}}\bigg(I_{\{Y \leq F^{-1}_{Y|X}(\tau|X)\}} - \frac{p(F^{-1}_{Y|X}(\tau|X), X)}{f_X(X)}\bigg)$$

$$\int_{[e_a,e_b]} \Gamma_1(B,s_0)(X,e)\psi_\tau(X,e)\,de,$$

$$\psi_4(Y,X) = I_{\{X\in M_X\}}\left(I_{\{Y\le F^{-1}_{Y|X}(\beta|X)\}} - \frac{p(F^{-1}_{Y|X}(\beta|X),X)}{f_X(X)}\right)$$

$$\int_{[e_a,e_b]} \Gamma_1(B,s_0)(X,e)\psi_\beta(X,e)\,de.$$

Then,

$$\begin{aligned}
&E[(\psi_{3,n}(Y,X)-\psi_3(Y,X))^2]\\
&= E\Bigg[\Bigg(\int_{M_X}\int_{[e_a,e_b]} \Gamma_1(B,s_0)(x,e)\psi_\tau(x,e)\mathbf{K}_{h_x}(X-x)\\
&\quad \left(\mathcal{K}_{h_y}(F^{-1}_{Y|X}(\tau|x)-Y) - \frac{p(F^{-1}_{Y|X}(\tau|x),x)}{f_X(x)}\right) de\,dx\\
&\quad - I_{\{X\in M_X\}}\left(I_{\{Y\le F^{-1}_{Y|X}(\tau|X)\}} - \frac{p(F^{-1}_{Y|X}(\tau|X),X)}{f_X(X)}\right)\int_{[e_a,e_b]} \Gamma_1(B,s_0)(X,e)\psi_\tau(X,e)\,de\Bigg)^2\Bigg]\\
&= \int\int\Bigg(\int_{M_X}\left(\mathcal{K}_{h_y}(F^{-1}_{Y|X}(\tau|x)-z) - \frac{p(F^{-1}_{Y|X}(\tau|x),x)}{f_X(x)}\right)\mathbf{K}_{h_x}(v-x)\\
&\quad \int_{[e_a,e_b]} \Gamma_1(B,s_0)(x,e)\psi_\tau(x,e)\,de\,dx - I_{\{v\in M_X\}}\left(I_{\{z\le F^{-1}_{Y|X}(\tau|v)\}} - \frac{p(F^{-1}_{Y|X}(\tau|v),v)}{f_X(v)}\right)\\
&\quad \int_{[e_a,e_b]} \Gamma_1(B,s_0)(v,e)\psi_\tau(v,e)\,de\Bigg)^2 f_{Y,X}(z,v)\,dz\,dv\\
&= \int\int\Bigg(\int I_{\{v-h_x x\in M_X\}}\Big(\mathcal{K}_{h_y}(F^{-1}_{Y|X}(\tau|v-h_x x)-z)\\
&\quad - \frac{p(F^{-1}_{Y|X}(\tau|v-h_x x),v-h_x x)}{f_X(v-h_x x)}\Big)\mathbf{K}(x)\int_{[e_a,e_b]} \Gamma_1(B,s_0)(v-h_x x,e)\psi_\tau(v-h_x x,e)\,de\,dx\\
&\quad - I_{\{v\in M_X\}}\left(I_{\{z\le F^{-1}_{Y|X}(\tau|v)\}} - \frac{p(F^{-1}_{Y|X}(\tau|v),v)}{f_X(v)}\right)\\
&\quad \int_{[e_a,e_b]} \Gamma_1(B,s_0)(v,e)\psi_\tau(v,e)\,de\Bigg)^2 f_{Y,X}(z,v)\,dz\,dv.
\end{aligned}$$

Since M_X is an interval, the boundary of M_X has Lebesgue-measure equal to zero. Note that $\mathcal{K}_{h_y}(F^{-1}_{Y|X}(\tau|v-h_x x)-z)\to I_{\{z\le F^{-1}_{Y|X}(\tau|v)\}}$ for Lebesgue-all $z\in\mathbb{R}, v,x\in\mathbb{R}^{d_X}$, so that due to the boundedness of M_X and the continuity of ψ_τ,Γ_1 and $F^{-1}_{Y|X}(\tau|\cdot)$, the dominated convergence theorem leads to

$$E[(\psi_{3,n}(Y,X)-\psi_3(Y,X))^2] \overset{n\to\infty}{\longrightarrow} 0.$$

The same reasoning can be applied for $\psi_{4,n}$. In total, this leads to

$$\sqrt{n}\int_{M_X}\int_{[e_a,e_b]}\Gamma_1(B,s_0)(x,e)\big(G_{nMD}(B,s_0)(x,e)+\Gamma_2(B,s_0)(x,e)[\hat{s}-s_0]\big)\,de\,dx$$

$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\psi(Y_i,X_i)-E[\psi(Y,X)])+o_p(1)$$

for

$$\psi(Y,X)=\psi_1(Y,X)+\psi_2(Y,X)+\psi_3(Y,X)+\psi_4(Y,X).$$

The Central Limit Theorem implies

$$\sqrt{n}\int_{M_X}\int_{[e_a,e_b]}\Gamma_1(B,s_0)(x,e)\big(G_{nMD}(B,s_0)(x,e)+\Gamma_2(B,s_0)(x,e)[\hat{s}-s_0]\big)\,de\,dx\overset{\mathcal{D}}{\to}\mathcal{N}(0,\sigma_A^2)$$

for $\sigma_A^2=\operatorname{Var}(\psi(Y,X))$. $\square$

To prove Lemma 4.2.3, it remains to prove

$$||\Gamma_2(B,s_0)[\hat{s}-s_0]||_2=\mathcal{O}_p\bigg(\frac{1}{\sqrt{nh_x^{d_X}}}\bigg).$$

In the following, the complexity of the dominating term in $||\Gamma_2(B,s_0)[\hat{s}-s_0]||_2$ will be reduced stepwise. First, apply (4.40) to obtain

$$\begin{aligned}
&||\Gamma_2(B,s_0)[\hat{s}-s_0]||_2\\
&\le \big|\big|D_h G_{MD}(c,s_0)[\hat{h}_1-h_1]\big|\big|_2+\big|\big|D_{f_{m_\tau}}G_{MD}(B,s_0)\big[\hat{F}^{-1}_{Y|X}(\tau|\cdot)-F^{-1}_{Y|X}(\tau|\cdot)\big]\big|\big|_2\\
&\quad+\big|\big|D_{f_{m_\beta}}G_{MD}(B,s_0)\big[\hat{F}^{-1}_{Y|X}(\beta|\cdot)-F^{-1}_{Y|X}(\beta|\cdot)\big]\big|\big|_2\\
&=\big|\big|D_{f_{m_\tau}}G_{MD}(B,s_0)\big[\hat{F}^{-1}_{Y|X}(\tau|\cdot)-F^{-1}_{Y|X}(\tau|\cdot)\big]\big|\big|_2\\
&\quad+\big|\big|D_{f_{m_\beta}}G_{MD}(B,s_0)\big[\hat{F}^{-1}_{Y|X}(\beta|\cdot)-F^{-1}_{Y|X}(\beta|\cdot)\big]\big|\big|_2+\mathcal{O}_p\Big(\frac{1}{\sqrt{n}}\Big),
\end{aligned}$$

where the last equation follows from (4.52), (4.53) and Lemma 4.6.7. Both of the terms

$$\big|\big|D_{f_{m_\tau}}G_{MD}(B,s_0)\big[\hat{F}^{-1}_{Y|X}(\tau|\cdot)-F^{-1}_{Y|X}(\tau|\cdot)\big]\big|\big|_2$$

and

$$\big|\big|D_{f_{m_\beta}}G_{MD}(B,s_0)\big[\hat{F}^{-1}_{Y|X}(\beta|\cdot)-F^{-1}_{Y|X}(\beta|\cdot)\big]\big|\big|_2$$

can be treated similarly to each other, so that only the first term is considered in the following. Recall

$$\begin{aligned}
&D_{f_{m_\tau}}k_B(s_0,x,e)\big[\hat{F}^{-1}_{Y|X}(\tau|\cdot)-F^{-1}_{Y|X}(\tau|\cdot)\big]\\
&=-\frac{B(1-e)h(F^{-1}_{Y|X}(\tau|x))(\hat{F}^{-1}_{Y|X}(\tau|x)-F^{-1}_{Y|X}(\tau|x))}{\sigma(x)\lambda(F^{-1}_{Y|X}(\tau|x))}
\end{aligned}$$

and

$$D_{f_{m_\tau}}G_{MD}(B,s_0)(x,e)\big[\hat{F}^{-1}_{Y|X}(\tau|\cdot)-F^{-1}_{Y|X}(\tau|\cdot)\big]$$

$$= \frac{1}{P(X \in M)} \Big(\int_{M_X} f_\varepsilon(k_B(s_0, w, e)) \big(I_{\{w \leq x\}} - P(X \leq x | X \in M) \big)$$
$$D_{f_{m_\tau}} k_B(s_0, w, e) \big[\hat{F}^{-1}_{Y|X}(\tau|\cdot) - F^{-1}_{Y|X}(\tau|\cdot) \big] f_X(w)\, dw \Big)$$

from the proof of Lemma 4.6.4. Consequently, there exists some constant $C > 0$ such that

$$\big|\big| D_{f_{m_\tau}} G_{MD}(B, s_0) \big[\hat{F}^{-1}_{Y|X}(\tau|\cdot) - F^{-1}_{Y|X}(\tau|\cdot) \big] \big|\big|_2 \leq C || \hat{F}^{-1}_{Y|X}(\tau|\cdot) - F^{-1}_{Y|X}(\tau|\cdot) ||_2,$$

that is, it suffices to prove

$$\int_{M_X} \big(\hat{F}^{-1}_{Y|X}(\tau|x) - F^{-1}_{Y|X}(\tau|x) \big)^2 dx = \mathcal{O}_p\Big(\frac{1}{n h_x^{d_X}} \Big).$$

In equation (4.67) in the proof of Lemma 4.2.15, it will be shown that

$$\hat{F}^{-1}_{Y|X}(\tau|x) - F^{-1}_{Y|X}(\tau|x) = -\frac{1}{f_{Y|X}(F^{-1}_{Y|X}(\tau|x)|x)} \frac{1}{n} \sum_{i=1}^n \Big(\frac{1}{f_X(x)} \mathcal{K}_{h_y}(F^{-1}_{Y|X}(\tau|x) - Y_i)$$
$$- \frac{p(F^{-1}_{Y|X}(\tau|x), x)}{f_X(x)^2} \Big) \mathbf{K}_{h_x}(x - X_i) + o_p\Big(\frac{1}{\sqrt{n}} \Big).$$

Even though under different assumptions, it was proven in Lemma 2.3.2 that (compare (2.21))

$$n h_x^{\frac{d_X}{2}} \int \int v(x) \big(\hat{F}^{-1}_{Y|X}(\tau|x) - F^{-1}_{Y|X}(\tau|x) \big)^2 dx\, \mu(d\tau) = \mathcal{O}_p\Big(h_x^{-\frac{d_X}{2}} \Big)$$

for an appropriate weight function v and some measure μ on $(0, 1)$. When choosing $v(x) = I_{\{x \in M_X\}}$ and $\mu = \delta_\tau$ for the Dirac measure δ_τ, this is exactly the assertion that needs to be proven. Indeed, one can proceed exactly in the same way as there, since the expansion of $\hat{F}^{-1}_{Y|X}(\tau|x) - F^{-1}_{Y|X}(\tau|x)$ from (4.67) corresponds to that of Lemma 2.8.1. Hence, the proof will be only sketched here. Define

$$\kappa(x) = \frac{1}{f_{Y|X}(F^{-1}_{Y|X}(\tau|x)|x)^2 f_X(x)^2}.$$

Then, it holds that

$$\int_{M_X} \big(\hat{F}^{-1}_{Y|X}(\tau|x) - F^{-1}_{Y|X}(\tau|x) \big)^2 dx$$
$$\leq 2 \int_{M_X} \Big(\frac{1}{f_{Y|X}(F^{-1}_{Y|X}(\tau|x)|x)} \frac{1}{n} \sum_{i=1}^n \Big(\frac{1}{f_X(x)} \mathcal{K}_{h_y}(F^{-1}_{Y|X}(\tau|x) - Y_i) - \frac{p(F^{-1}_{Y|X}(\tau|x), x)}{f_X(x)^2} \Big)$$
$$\mathbf{K}_{h_x}(x - X_i) \Big)^2 dx + o_p\Big(\frac{1}{n} \Big)$$
$$= \frac{2}{n^2} \sum_{i=1}^n \int_{M_X} \kappa(x) \Big(\mathcal{K}_{h_y}(F^{-1}_{Y|X}(\tau|x) - Y_i) - \frac{p(F^{-1}_{Y|X}(\tau|x), x)}{f_X(x)} \Big)^2 \mathbf{K}_{h_x}(x - X_i)^2\, dx$$
$$+ \frac{2}{n^2} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \int_{M_X} \kappa(x) \Big(\mathcal{K}_{h_y}(F^{-1}_{Y|X}(\tau|x) - Y_i) - \frac{p(F^{-1}_{Y|X}(\tau|x), x)}{f_X(x)} \Big) \mathbf{K}_{h_x}(x - X_i)$$

$$\bigg(\mathcal{K}_{h_y}(F_{Y|X}^{-1}(\tau|x) - Y_j) - \frac{p(F_{Y|X}^{-1}(\tau|x), x)}{f_X(x)}\bigg)\mathbf{K}_{h_x}(X_j - x)\,dx + o_p\Big(\frac{1}{n}\Big).$$

For some sufficiently large constant $C > 0$ the expectation of the first term can be bounded by

$$\begin{aligned}
&E\bigg[\frac{2}{n^2}\sum_{i=1}^{n}\int_{M_X}\kappa(x)\bigg(\mathcal{K}_{h_y}(F_{Y|X}^{-1}(\tau|x) - Y_i) - \frac{p(F_{Y|X}^{-1}(\tau|x), x)}{f_X(x)}\bigg)^2\mathbf{K}_{h_x}(x - X_i)^2\,dx\bigg] \\
&\leq \frac{C}{nh_x^{d_X}}E\bigg[\int\frac{1}{h_x^{d_X}}K\Big(\frac{x - X_1}{h_x}\Big)^2\,dx\bigg] \\
&= \frac{C}{nh_x^{d_X}}\int K(x)^2\,dx \\
&= \mathcal{O}\Big(\frac{1}{nh_x^{d_X}}\Big).
\end{aligned}$$

Due to Lemma 1.1.1, the expectation of the second term is of order $o\big(\frac{1}{n}\big)$, while it can be shown via the same calculations as were done to obtain (2.49) that the variance is of order $\mathcal{O}\big(n^{-2}h_x^{-2d_X}\big)$. The last assertion of Remark 4.2.3 is a direct consequence of part (ii) of Theorem 4.2.4.

4.6.5 Proof of Theorem 4.2.5

Let $\mathcal{K} \subseteq \mathbb{R}$ be compact. Recall the definition of $\tilde{B}$ from equation (4.3)

$$\tilde{B} = -\frac{\partial}{\partial y}\hat{\lambda}(y)\Big|_{y=\hat{y}_0}.$$

First, a Taylor expansion together with Theorem 4.2.8 (see the proof in Section 4.6.7 below) and Corollary 4.2.14 leads to

$$\begin{aligned}
\tilde{B} &= -\frac{\partial}{\partial y}\hat{\lambda}(y)\Big|_{y=\hat{y}_0} \\
&= -\frac{\partial}{\partial y}\hat{\lambda}(y)\Big|_{y=y_0} - \frac{\partial^2}{\partial y^2}\hat{\lambda}(y)\Big|_{y=y^*}(\hat{y}_0 - y_0) \\
&= -\frac{\partial}{\partial y}\hat{\lambda}(y)\Big|_{y=y_0} - \frac{\partial^2}{\partial y^2}\lambda(y)\Big|_{y=y^*}(\hat{y}_0 - y_0) + o_p(|\hat{y}_0 - y_0|) \\
&= -\frac{\partial}{\partial y}\hat{\lambda}(y)\Big|_{y=y_0} + \mathcal{O}_p\Big(\frac{1}{\sqrt{nh_y}}\Big)
\end{aligned}$$

for some y^* between $\hat{y}_0$ and y_0. Note that since λ is two times continuously differentiable and $\hat{y}_0 - y_0 = o_p(1)$ is implied by Theorem 4.2.8, it holds that $\frac{\partial^2}{\partial y^2}\lambda(y)\Big|_{y=y^*}$ is bounded in probability. Hence, the error due to the estimation of y_0 will be asymptotically negligible and it suffices to consider the asymptotic behaviour of $\frac{\partial}{\partial y}\hat{\lambda}(y)\big|_{y=y_0}$. Equations (4.65) and (4.66) below imply

$$\sqrt{h_y^3}\sup_{u\in\mathcal{K}}v(x)\Big|\frac{\partial}{\partial y}\hat{p}_y(u, x) - \frac{\partial}{\partial y}p_y(u, x)\Big| = \mathcal{O}_p\Bigg(\sqrt{\frac{\log(n)}{nh_x^{d_X}}}\Bigg) \qquad = o_p\Big(n^{-\frac{1}{4}}\Big),$$

$$\sqrt{h_y^3}\sup_{u\in\mathcal{K}} v(x)\left|\frac{\partial}{\partial y}\hat{p}_x(u,x)-\frac{\partial}{\partial y}p_x(u,x)\right| = \mathcal{O}_p\left(\sqrt{\frac{\log(n)h_y^2}{nh_x^{d_X+2}}}\right) \qquad = o_p\left(n^{-\frac{1}{4}}\right).$$

In the following, negligibility is (if not specified further) meant uniformly in $y\in\mathcal{K}$ and $x\in\operatorname{supp}(v)$. Remember (4.33) and (4.34). Due to (with $f=f_X$ and $\hat{f}=\hat{f}_X$)

$$\frac{\partial}{\partial y}\hat{\Phi}_y(u,x)=\frac{\frac{\partial}{\partial y}\hat{p}_y(u,x)}{\hat{f}(x)} \quad \text{and} \quad \frac{\partial}{\partial y}\hat{\Phi}_x(u,x)=\frac{\frac{\partial}{\partial y}\hat{p}_x(u,x)}{\hat{f}(x)}-\frac{\hat{f}_x(x)\hat{p}_y(u,x)}{\hat{f}^2(x)},$$

equation (1.8) can be applied again to obtain (arguments are omitted)

$$\frac{\partial}{\partial y}\hat{\Phi}_y-\frac{\partial}{\partial y}\Phi_y=\frac{1}{f}\left(\frac{\partial}{\partial y}\hat{p}_y-\frac{\partial}{\partial y}p_y\right)-\frac{\frac{\partial}{\partial y}p_y}{f^2}(\hat{f}-f)-\underbrace{\frac{\hat{f}-f}{\hat{f}f}\left(\frac{\partial}{\partial y}\hat{p}_y-\frac{\partial}{\partial y}p_y-\frac{\frac{\partial}{\partial y}p_y(\hat{f}-f)}{f}\right)}_{=o_p\left(\frac{1}{\sqrt{nh_y^3}}\right)}$$

as well as

$$\begin{aligned}\frac{\hat{p}_y\hat{f}_x}{\hat{f}^2}-\frac{p_yf_x}{f^2}&=\frac{1}{f^2}(\hat{p}_y\hat{f}_x-p_yf_x)-\frac{p_yf_x}{f^4}(\hat{f}^2-f^2)+o_p\left(\frac{1}{\sqrt{nh_y^3}}\right)\\&=\frac{1}{f^2}((\hat{p}_y-p_y)\hat{f}_x+p(\hat{f}_x-f_x))-\frac{p_yf_x}{f^4}(\hat{f}-f)(\hat{f}+f)+o_p\left(\frac{1}{\sqrt{nh_y^3}}\right)\\&=\frac{f_x}{f^2}(\hat{p}_y-p_y)+\frac{p_y}{f^2}(\hat{f}_x-f_x)-\frac{2p_yf_x}{f^3}(\hat{f}-f)+o_p\left(\frac{1}{\sqrt{nh_y^3}}\right)\end{aligned}$$

and thus

$$\begin{aligned}\frac{\partial}{\partial y}\hat{\Phi}_x-\frac{\partial}{\partial y}\Phi_x=&-\frac{f_x}{f^2}(\hat{p}_y-p_y)+\frac{1}{f}\left(\frac{\partial}{\partial y}\hat{p}_x-\frac{\partial}{\partial y}p_x\right)+\left(\frac{2p_yf_x}{f^3}-\frac{\frac{\partial}{\partial y}p_x}{f^2}\right)(\hat{f}-f)\\&-\frac{p_y}{f^2}(\hat{f}_x-f_x)+o_p\left(\frac{1}{\sqrt{nh_y^3}}\right).\end{aligned}$$

In total, this leads to

$$\begin{aligned}&\frac{\partial}{\partial y}\left(\frac{\hat{\Phi}_x}{\hat{\Phi}_y}-\frac{\Phi_x}{\Phi_y}\right)\\&=\frac{\frac{\partial}{\partial y}\hat{\Phi}_x}{\hat{\Phi}_y}-\frac{\hat{\Phi}_x\frac{\partial}{\partial y}\hat{\Phi}_y}{\hat{\Phi}_y^2}-\frac{\frac{\partial}{\partial y}\Phi_x}{\Phi_y}+\frac{\Phi_x\frac{\partial}{\partial y}\Phi_y}{\Phi_y^2}\\&=\frac{1}{\Phi_y}\left(\frac{\partial}{\partial y}\hat{\Phi}_x-\frac{\partial}{\partial y}\Phi_x\right)-\frac{\frac{\partial}{\partial y}\Phi_x}{\Phi_y^2}(\hat{\Phi}_y-\Phi_y)-\frac{\hat{\Phi}_y-\Phi_y}{\hat{\Phi}_y\Phi_y}\left(\frac{\partial}{\partial y}\hat{\Phi}_x-\frac{\partial}{\partial y}\Phi_x-\frac{\frac{\partial}{\partial y}\Phi_x(\hat{\Phi}_y-\Phi_y)}{\Phi_y}\right)\\&-\frac{1}{\Phi_y^2}\left(\hat{\Phi}_x\frac{\partial}{\partial y}\hat{\Phi}_y-\Phi_x\frac{\partial}{\partial y}\Phi_y\right)+\frac{\Phi_x\frac{\partial}{\partial y}\Phi_y}{\Phi_y^4}(\hat{\Phi}_y^2-\Phi_y^2)\\&+\frac{\hat{\Phi}_y^2-\Phi_y^2}{\hat{\Phi}_y^2\Phi_y^2}\left(\hat{\Phi}_x\frac{\partial}{\partial y}\hat{\Phi}_y-\Phi_x\frac{\partial}{\partial y}\Phi_y-\frac{\Phi_x\frac{\partial}{\partial y}\Phi_y(\hat{\Phi}_y^2-\Phi_y^2)}{\Phi_y^2}\right)\end{aligned}$$

$$= \frac{1}{\Phi_y}\left(\frac{\partial}{\partial y}\hat{\Phi}_x - \frac{\partial}{\partial y}\Phi_x\right) - \frac{\frac{\partial}{\partial y}\Phi_x}{\Phi_y^2}(\hat{\Phi}_y - \Phi_y) - \frac{\frac{\partial}{\partial y}\Phi_y}{\Phi_y^2}(\hat{\Phi}_x - \Phi_x) - \frac{\Phi_x}{\Phi_y^2}\left(\frac{\partial}{\partial y}\hat{\Phi}_y - \frac{\partial}{\partial y}\Phi_y\right)$$

$$+ \frac{2\Phi_x\frac{\partial}{\partial y}\Phi_y}{\Phi_y^3}(\hat{\Phi}_y - \Phi_y) + o_p\left(\frac{1}{\sqrt{nh_y^3}}\right)$$

$$= \frac{1}{\Phi_y}\left(\frac{\partial}{\partial y}\hat{\Phi}_x - \frac{\partial}{\partial y}\Phi_x\right) + \left(\frac{2\Phi_x\frac{\partial}{\partial y}\Phi_y}{\Phi_y^3} - \frac{\frac{\partial}{\partial y}\Phi_x}{\Phi_y^2}\right)(\hat{\Phi}_y - \Phi_y) - \frac{\frac{\partial}{\partial y}\Phi_y}{\Phi_y^2}(\hat{\Phi}_x - \Phi_x)$$

$$- \frac{\Phi_x}{\Phi_y^2}\left(\frac{\partial}{\partial y}\hat{\Phi}_y - \frac{\partial}{\partial y}\Phi_y\right) + o_p\left(\frac{1}{\sqrt{nh_y^3}}\right)$$

$$= \frac{1}{\Phi_y}\left(- \frac{f_x}{f^2}(\hat{p}_y - p_y) + \frac{1}{f}\left(\frac{\partial}{\partial y}\hat{p}_x - \frac{\partial}{\partial y}p_x\right) + \left(\frac{2p_y f_x}{f^3} - \frac{\frac{\partial}{\partial y}p_x}{f^2}\right)(\hat{f} - f) - \frac{p_y}{f^2}(\hat{f}_x - f_x)\right)$$

$$+ \left(\frac{2\Phi_x\frac{\partial}{\partial y}\Phi_y}{\Phi_y^3} - \frac{\frac{\partial}{\partial y}\Phi_x}{\Phi_y^2}\right)\left(\frac{1}{f}(\hat{p}_y - p_y) - \frac{p_y}{f^2}(\hat{f} - f)\right)$$

$$- \frac{\frac{\partial}{\partial y}\Phi_y}{\Phi_y^2}\left(- \frac{f_x}{f^2}(\hat{p} - p) + \frac{1}{f}(\hat{p}_x - p_x) + \left(\frac{2pf_x}{f^3} - \frac{p_x}{f^2}\right)(\hat{f} - f) - \frac{p}{f^2}(\hat{f}_x - f_x)\right)$$

$$- \frac{\Phi_x}{\Phi_y^2}\left(\frac{1}{f}\left(\frac{\partial}{\partial y}\hat{p}_y - \frac{\partial}{\partial y}p_y\right) - \frac{\frac{\partial}{\partial y}p_y}{f^2}(\hat{f} - f)\right) + o_p\left(\frac{1}{\sqrt{nh_y^3}}\right)$$

$$= \frac{\frac{\partial}{\partial y}\Phi_y f_x}{\Phi_y^2 f^2}(\hat{p} - p) + \left(\frac{2\Phi_x\frac{\partial}{\partial y}\Phi_y}{f\Phi_y^3} - \frac{\frac{\partial}{\partial y}\Phi_x}{f\Phi_y^2} - \frac{f_x}{\Phi_y f^2}\right)(\hat{p}_y - p_y) - \frac{\frac{\partial}{\partial y}\Phi_y}{\Phi_y^2 f}(\hat{p}_x - p_x)$$

$$- \frac{\Phi_x}{\Phi_y^2 f}\left(\frac{\partial}{\partial y}\hat{p}_y - \frac{\partial}{\partial y}p_y\right) + \frac{1}{\Phi_y f}\left(\frac{\partial}{\partial y}\hat{p}_x - \frac{\partial}{\partial y}p_x\right)$$

$$+ \left(\frac{2p_y f_x}{\Phi_y f^3} - \frac{\frac{\partial}{\partial y}p_x}{\Phi_y f^2} - \frac{2\Phi_x p_y\frac{\partial}{\partial y}\Phi_y}{\Phi_y^3 f^2} + \frac{p_y\frac{\partial}{\partial y}\Phi_x}{\Phi_y^2 f^2} - \frac{2pf_x\frac{\partial}{\partial y}\Phi_y}{f^3\Phi_y^2} + \frac{p_x\frac{\partial}{\partial y}\Phi_y}{f^2\Phi_y^2} + \frac{\Phi_x\frac{\partial}{\partial y}p_y}{\Phi_y^2 f^2}\right)(\hat{f} - f)$$

$$+ \left(\frac{p\frac{\partial}{\partial y}\Phi_y}{\Phi_y^2 f^2} - \frac{p_y}{\Phi_y f^2}\right)(\hat{f}_x - f_x) + o_p\left(\frac{1}{\sqrt{nh_y^3}}\right)$$

$$= \tilde{D}_{p,0}(\hat{p} - p) + \tilde{D}_{p,y}(\hat{p}_y - p_y) + \tilde{D}_{p,x}(\hat{p}_x - p_x)$$

$$+ \tilde{D}_{p,yy}\left(\frac{\partial}{\partial y}\hat{p}_y - \frac{\partial}{\partial y}p_y\right) + \tilde{D}_{p,xy}\left(\frac{\partial}{\partial y}\hat{p}_x - \frac{\partial}{\partial y}p_x\right)$$

$$+ \tilde{D}_{f,0}(\hat{f} - f) + \tilde{D}_{f,x}(\hat{f}_x - f_x) + o_p\left(\frac{1}{\sqrt{nh_y^3}}\right)$$

with

$$\tilde{D}_{p,0} = \frac{\frac{\partial}{\partial y}\Phi_y f_x}{\Phi_y^2 f^2}, \qquad \tilde{D}_{p,y} = \frac{2\Phi_x\frac{\partial}{\partial y}\Phi_y}{f\Phi_y^3} - \frac{\frac{\partial}{\partial y}\Phi_x}{f\Phi_y^2} - \frac{f_x}{\Phi_y f^2}, \qquad \tilde{D}_{p,x} = -\frac{\frac{\partial}{\partial y}\Phi_y}{\Phi_y^2 f},$$

$$\tilde{D}_{p,yy} = -\frac{\Phi_x}{\Phi_y^2 f}, \qquad \tilde{D}_{p,xy} = \frac{1}{\Phi_y f}, \qquad \tilde{D}_{f,x} = \frac{p\frac{\partial}{\partial y}\Phi_y}{\Phi_y^2 f^2} - \frac{1}{f}$$

and

$$\tilde{D}_{f,0} = \frac{2f_x}{f^2} - \frac{\frac{\partial}{\partial y}p_x}{\Phi_y f^2} - \frac{2\Phi_x \frac{\partial}{\partial y}\Phi_y}{\Phi_y^2 f} + \frac{\frac{\partial}{\partial y}\Phi_x}{\Phi_y f} - \frac{2pf_x \frac{\partial}{\partial y}\Phi_y}{f^3\Phi_y^2} + \frac{p_x \frac{\partial}{\partial y}\Phi_y}{f^2\Phi_y^2} + \frac{\Phi_x \frac{\partial}{\partial y}p_y}{\Phi_y^2 f^2}.$$

Similar to the proof of Lemma 4.2.1, one has

$$\begin{aligned}
&\tilde{D}_{p,0}p + \tilde{D}_{p,y}p_y + \tilde{D}_{p,x}p_x + \tilde{D}_{p,yy}\frac{\partial}{\partial y}p_y + \tilde{D}_{p,xy}\frac{\partial}{\partial y}p_x + \tilde{D}_{f,0}f + \tilde{D}_{f,x}f_x \\
&= \frac{pf_x \frac{\partial}{\partial y}\Phi_y}{\Phi_y^2 f^2} + \frac{2\Phi_x \frac{\partial}{\partial y}\Phi_y}{\Phi_y^2} - \frac{\frac{\partial}{\partial y}\Phi_x}{\Phi_y} - \frac{f_x}{f} - \frac{p_x \frac{\partial}{\partial y}\Phi_y}{\Phi_y^2 f} - \frac{\Phi_x \frac{\partial}{\partial y}p_y}{\Phi_y^2 f} + \frac{\frac{\partial}{\partial y}p_x}{\Phi_y f} \\
&\quad + \frac{2f_x}{f} + \frac{\frac{\partial}{\partial y}p_x}{\Phi_y f} - \frac{2\Phi_x \frac{\partial}{\partial y}\Phi_y}{\Phi_y^2} - \frac{\frac{\partial}{\partial y}\Phi_x}{\Phi_y} - \frac{2pf_x \frac{\partial}{\partial y}\Phi_y}{f^2\Phi_y^2} + \frac{p_x \frac{\partial}{\partial y}\Phi_y}{f\Phi_y^2} + \frac{\Phi_x \frac{\partial}{\partial y}p_y}{\Phi_y^2 f} + \frac{pf_x \frac{\partial}{\partial y}\Phi_y}{\Phi_y^2 f^2} - \frac{f_x}{f} \\
&= 0
\end{aligned}$$

and

$$E\Big[\tilde{D}_{p,0}\hat{p}+\tilde{D}_{p,y}\hat{p}_y+\tilde{D}_{p,x}\hat{p}_x+\tilde{D}_{p,yy}\frac{\partial}{\partial y}\hat{p}_y+\tilde{D}_{p,xy}\frac{\partial}{\partial y}\hat{p}_x+\tilde{D}_{f,0}\hat{f}+\tilde{D}_{f,x}\hat{f}_x\Big] = o\bigg(\frac{1}{\sqrt{nh_y^3}}\bigg). \quad (4.54)$$

The dominated convergence theorem yields

$$\begin{aligned}
&\sqrt{nh_y^3}\bigg(\frac{\partial}{\partial y}\hat{\lambda}(y)\Big|_{y=y_0} - \frac{\partial}{\partial y}\lambda(y)\Big|_{y=y_0}\bigg) \\
&= \sqrt{nh_y^3}\int v(x)\frac{\partial}{\partial y}\bigg(\frac{\hat{\Phi}_x(y_0,x)}{\hat{\Phi}_y(y_0,x)} - \frac{\Phi_x(y_0,x)}{\Phi_y(y_0,x)}\bigg)\,dx \\
&= \sqrt{nh_y^3}\int v(x)\bigg(\tilde{D}_{p,0}(y_0,x)\hat{p}(y_0,x) + \tilde{D}_{p,y}(y_0,x)\hat{p}_y(y_0,x) + \tilde{D}_{p,x}(y_0,x)\hat{p}_x(y_0,x) \\
&\quad + \tilde{D}_{p,yy}(y_0,x)\frac{\partial}{\partial y}\hat{p}_y(y_0,x) + \tilde{D}_{p,xy}(y_0,x)\frac{\partial}{\partial y}\hat{p}_x(y_0,x) + \tilde{D}_{f,0}(y_0,x)\hat{f}(x) \\
&\quad + \tilde{D}_{f,x}(y_0,x)\hat{f}_x(x)\bigg)\,dx + o_p(1).
\end{aligned}$$

The variance of most of the terms above is asymptotically negligible, since for example for some sufficiently large constant $C>0$ one has (see (1.7))

$$\begin{aligned}
&\operatorname{Var}\bigg(\sqrt{nh_y^3}\int v(x)\tilde{D}_{p,xy}(y_0,x)\frac{\partial}{\partial y}\hat{p}_x(y_0,x)\,dx\bigg) \\
&= nh_y^3\operatorname{Var}\bigg(\frac{1}{n}\sum_{i=1}^n K_{h_y}(y-Y_i)\int v(x)\tilde{D}_{p,xy}(y_0,x)\frac{\partial}{\partial w_1}\mathbf{K}_{h_x}(w)\Big|_{w=x-X_i}\,dx\bigg) \\
&= h_y^3 E\bigg[K_{h_y}(y-Y_1)^2\bigg(\int v(x)\tilde{D}_{p,xy}(y_0,x)\frac{\partial}{\partial w_1}\mathbf{K}_{h_x}(w)\Big|_{w=x-X_1}\,dx\bigg)^2\bigg] + o(1) \\
&\le Ch_y E\bigg[\bigg(\int \mathbf{K}_{h_x}(x)\frac{\partial}{\partial x_1}v(x)\tilde{D}_{p,xy}(y_0,x)\,dx\bigg)^2\bigg] + o(1) \\
&= Ch_y E\bigg[\bigg(\int \mathbf{K}(x)\frac{\partial}{\partial w_1}v(w)\tilde{D}_{p,xy}(y_0,w)\Big|_{w=X_i+h_x x}\,dx\bigg)^2\bigg] + o(1)
\end{aligned}$$

$$= o(1).$$

A similar reasoning for the other terms except $\tilde{D}_{p,0}(y_0,x)\hat{p}(y_0,x)$ leads to

$$\sqrt{nh_y^3}\bigg(\frac{\partial}{\partial y}\hat{\lambda}(y)\Big|_{y=y_0} - \frac{\partial}{\partial y}\lambda(y)\Big|_{y=y_0}\bigg)$$

$$= \sqrt{nh_y^3}\int v(x)\bigg(E\Big[\tilde{D}_{p,0}(y_0,x)\hat{p}(y_0,x) + \tilde{D}_{p,y}(y_0,x)\hat{p}_y(y_0,x) + \tilde{D}_{p,x}(y_0,x)\hat{p}_x(y_0,x)\Big]$$

$$+ \tilde{D}_{p,yy}(y_0,x)\frac{\partial}{\partial y}\hat{p}_y(y_0,x) + E\Big[\tilde{D}_{p,xy}(y_0,x)\frac{\partial}{\partial y}\hat{p}_x(y_0,x) + \tilde{D}_{f,0}(y_0,x)\hat{f}(x)$$

$$+ \tilde{D}_{f,x}(y_0,x)\hat{f}_x(x)\Big]\bigg)\,dx + o_p(1)$$

$$\overset{(4.54)}{=} \sqrt{nh_y^3}\int v(x)\bigg(\tilde{D}_{p,yy}(y_0,x)\frac{\partial}{\partial y}\hat{p}_y(y_0,x) - \tilde{D}_{p,yy}(y_0,x)E\Big[\frac{\partial}{\partial y}\hat{p}_y(y_0,x)\Big]\bigg)\,dx + o_p(1).$$

$$= \sqrt{nh_y^3}\int v(x)\tilde{D}_{p,yy}(y_0,x)\bigg(\frac{\partial}{\partial y}\hat{p}_y(y_0,x) - \frac{\partial}{\partial y}p_y(y_0,x)\bigg)\,dx + o_p(1)$$

$$= \sqrt{nh_y^3}\bigg(\frac{1}{nh_y^2}\sum_{i=1}^n \frac{\partial}{\partial u}K(u)\Big|_{u=\frac{y_0-Y_i}{h_y}}\int v(x)\tilde{D}_{p,yy}(y_0,x)\mathbf{K}_{h_x}(x-X_i)\,dx$$

$$- \int v(x)\tilde{D}_{p,yy}(y_0,x)\frac{\partial}{\partial y}p_y(y_0,x)\bigg) + o_p(1)$$

$$= \sum_{i=1}^n (Z_{n,i} - E[Z_{n,i}]) + o_p(1),$$

where

$$Z_{n,i} = \frac{1}{\sqrt{nh_y}}\frac{\partial}{\partial u}K(u)\Big|_{u=\frac{y_0-Y_i}{h_y}}\int v(X_i+h_xx)\tilde{D}_{p,yy}(y_0,X_i+h_xx)\mathbf{K}(x)\,dx = o(1) \quad (4.55)$$

and the third from last equality follows as in (4.64) below. It holds that

$$\mathrm{Var}\bigg(\sum_{i=1}^n (Z_{n,i} - E[Z_{n,i}])\bigg)$$

$$= \frac{1}{h_y}\mathrm{Var}\bigg(\frac{\partial}{\partial u}K(u)\Big|_{u=\frac{y_0-Y_1}{h_y}}\int v(X_1+h_xx)\tilde{D}_{p,yy}(y_0,X_1+h_xx)\mathbf{K}(x)\,dx\bigg)$$

$$= \frac{1}{h_y}E\bigg[\bigg(\frac{\partial}{\partial u}K(u)\Big|_{u=\frac{y_0-Y_1}{h_y}}\int v(X_1+h_xx)\tilde{D}_{p,yy}(y_0,X_1+h_xx)\mathbf{K}(x)\,dx\bigg)^2\bigg]$$

$$- \frac{1}{h_y}E\bigg[\frac{\partial}{\partial u}K(u)\Big|_{u=\frac{y_0-Y_1}{h_y}}\int v(X_1+h_xx)\tilde{D}_{p,yy}(y_0,X_1+h_xx)\mathbf{K}(x)\,dx\bigg]^2$$

$$= \frac{1}{h_y}E\bigg[\bigg(\frac{\partial}{\partial u}K(u)\Big|_{u=\frac{y_0-Y_1}{h_y}}\int v(X_1+h_xx)\tilde{D}_{p,yy}(y_0,X_1+h_xx)\mathbf{K}(x)\,dx\bigg)^2\bigg]$$

$$- \frac{1}{h_y}\bigg(\int\int \frac{\partial}{\partial u}K(u)\Big|_{u=\frac{y_0-z}{h_y}}\int v(w+h_xx)$$

$$\tilde{D}_{p,yy}(y_0, w + h_x x)\mathbf{K}(x)\,dx\,f_{Y,X}(z,w)\,dz\,dw\Bigg)^2$$

$$= \frac{1}{h_y}\int\int\left(\frac{\partial}{\partial u}K(u)\Big|_{u=\frac{y_0-z}{h_y}}\right)^2\left(\int v(w+h_x x)\tilde{D}_{p,yy}(y_0, w+h_x x)\mathbf{K}(x)\,dx\right)^2$$

$$f_{Y,X}(z,w)\,dz\,dw - h_y\Bigg(\int\int\frac{\partial}{\partial u}K(u)\Big|_{u=z}\int v(w+h_x x)\tilde{D}_{p,yy}(y_0, w+h_x x)\mathbf{K}(x)\,dx$$

$$f_{Y,X}(y_0 - h_y z, w)\,dz\,dw\Bigg)^2$$

$$= \int\int\left(\frac{\partial}{\partial z}K(z)\right)^2\left(\int v(w+h_x x)\tilde{D}_{p,yy}(y_0, w+h_x x)\mathbf{K}(x)\,dx\right)^2 f_{Y,X}(y_0 - h_y z, w)\,dz\,dw$$

$$+ o(1)$$

$$= \int\left(\frac{\partial}{\partial z}K(z)\right)^2 dz\int v(w)^2\tilde{D}_{p,yy}(y_0,w)^2 f_{Y,X}(y_0,w)\,dw + o(1)$$

$$= \sigma^2_{\tilde{B}} + o(1).$$

If

$$\frac{1}{\operatorname{Var}\left(\sum_{i=1}^n (Z_{n,i} - E[Z_{n,i}])\right)}$$

$$\sum_{i=1}^n E\left[(Z_{n,i} - E[Z_{n,1}])^2 I_{\left\{|Z_{n,i}-E[Z_{n,1}]|^2 > \varepsilon \operatorname{Var}\left(\sum_{j=1}^n (Z_{n,j}-E[Z_{n,j}])\right)\right\}}\right] \to 0$$

holds for all $\varepsilon > 0$, the Lindeberg-Feller Theorem yields asymptotic normality. Due to $\sigma^2_{\tilde{B}} > 0$ and (4.55), the indicator function in the expectation above equals zero for sufficiently large $n \in \mathbb{N}$, which implies applicability of the Lindeberg-Feller Theorem. The assertion follows from $\tilde{D}_{p,yy} = D_{p,y}$.

4.6.6 Proof of Theorem 4.2.6

Proof of (i): In the proof of Theorem 4.2.4, it was shown that (see (**C6'**) and (4.50))

$$\sqrt{n}(\hat{B} - B) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_{\Gamma_2}(Y_i, X_i) + o_p\left(\frac{1}{\sqrt{n}}\right). \tag{4.56}$$

Now, let $\mathcal{K} \subseteq (y_0, \infty)$ be compact. Recall (4.8) and (4.9), that is,

$$h_c(y) = \exp\left(-c\int_{y_1}^{y}\frac{1}{\lambda(u)}\,du\right) = \exp\left(c\log(h_1(y))\right) \quad \text{and} \quad \bar{h}_c(y) = \exp\left(c\log(\bar{h}_1(y))\right)$$

for all $c \in [B_1, B_2]$. Note that

$$\sup_{y\in\mathcal{K}}|\bar{h}_1(y) - h_1(y)|^2 = o_p\left(\frac{1}{\sqrt{n}}\right),$$

$$\sup_{y\in\mathcal{K}}|\log(\bar{h}_1(y)) - \log(h_1(y))|^2 = o_p\left(\frac{1}{\sqrt{n}}\right)$$

due to (**C4**) and $\mathcal{K} \subseteq (y_0, \infty)$. Therefore, a Taylor expansion yields for some h^* between $\bar{h}_1$ and h_1

$$\begin{aligned}
\bar{h}_c(y) - h_c(y) &= \exp\big(c\log(\bar{h}_1(y))\big) - \exp\big(c\log(h_1(y))\big) \\
&= c\exp\big(c\log(h_1(y))\big)\big(\log(\bar{h}_1(y)) - \log(h_1(y))\big) \\
&\quad + c^2\exp\big(c\log(h_1(y))\big)\big(\log(\bar{h}_1(y)) - \log(h_1(y))\big)^2 \\
&\overset{(\mathbf{C4})}{=} c\exp\big(c\log(h_1(y))\big)\big(\log(\bar{h}_1(y)) - \log(h_1(y))\big) + o_p\Big(\frac{1}{\sqrt{n}}\Big) \\
&= ch_c(y)\frac{\bar{h}_1(y) - h_1(y)}{h_1(y)} + o_p\Big(\frac{1}{\sqrt{n}}\Big)
\end{aligned}$$

uniformly in $y \in \mathcal{K}$ and $c \in [B_1, B_2]$, that is, once h_1 can be estimated with a certain rate, this rate transfers to h_c uniformly in $c \in [B_1, B_2]$. The only thing left is to replace c by the estimator $\hat{B}$ and to apply (4.56), so that

$$\begin{aligned}
&H_n(y) \\
&= \sqrt{n}(\bar{h}(y) - h(y)) \\
&= \sqrt{n}(\bar{h}_{\hat{B}}(y) - h_{\hat{B}}(y) + h_{\hat{B}}(y) - h_B(y)) \\
&= \hat{B}h_{\hat{B}}(y)\frac{\sqrt{n}(\bar{h}_1(y) - h_1(y))}{h_1(y)} + \sqrt{n}\Big(\exp\big(\hat{B}\log(h_1(y))\big) - \exp\big(B\log(h_1(y))\big)\Big) + o_p(1) \\
&= Bh_B(y)\frac{\sqrt{n}(\bar{h}_1(y) - h_1(y))}{h_1(y)} + \exp\big(B\log(h_1(y))\big)\log(h_1(y))\sqrt{n}(\hat{B} - B) + o_p(1) \\
&\overset{(4.56)}{=} \frac{h(y)}{\sqrt{n}}\sum_{i=1}^n\Big(\frac{B\psi_h(Y_i, X_i, y)}{h_1(y)} + \log(h_1(y))\psi_{\Gamma_2}(Y_i, X_i)\Big). \qquad (4.57)
\end{aligned}$$

Convergence of the finite dimensional distributions follows from the Central Limit Theorem. Since h_1 is continuous and bounded away from zero on $\mathcal{K}$, asymptotic equicontinuity (see condition (2.1.8) of Van der Vaart and Wellner (1996) for a definition) is implied by that of $\big(\sqrt{n}(\bar{h}_1(y) - h(y))\big)_{y\in\mathcal{K}}$. Hence, Corollary 2.3.12 of Van der Vaart and Wellner (1996)) yields

$$(H_n(y))_{y\in\mathcal{K}} \rightsquigarrow (Z_h(y))_{y\in\mathcal{K}},$$

where the stated expression for the covariance function directly follows from (4.57).
For the estimator $\hat{h}_1$, which was given in (4.8), it was shown in the proof of Theorem 4.2.2 that

$$\begin{aligned}
\hat{h}_c(y) - h_c(y) &= -c\exp\Big(-c\int_{y_1}^y \frac{1}{\lambda(u)}\,du\Big)\Big(\int_{y_1}^y \frac{1}{\hat{\lambda}(u)}\,du - \int_{y_1}^y \frac{1}{\lambda(u)}\,du\Big) \\
&\quad + o_p\Big(\int_{y_1}^y \frac{1}{\hat{\lambda}(u)}\,du - \int_{y_1}^y \frac{1}{\lambda(u)}\,du\Big) \\
&= -\frac{ch_c(y)}{n}\sum_{i=1}^n \eta_i(y) + o_p\Big(\frac{1}{\sqrt{n}}\Big)
\end{aligned}$$

with η as in (4.36), which leads to

$$\begin{aligned}
H_n(y) &= Bh_B(y)\frac{\sqrt{n}(\hat{h}_1(y)-h_1(y))}{h_1(y)} + \exp\big(B\log(h_1(y))\big)\log(h_1(y))\sqrt{n}(\hat{B}-B) + o_p(1) \\
&= -Bh(y)\sqrt{n}\bigg(\int_{y_1}^{y}\frac{1}{\hat{\lambda}(u)}\,du - \int_{y_1}^{y}\frac{1}{\lambda(u)}\,du\bigg) - h(y)\int_{y_1}^{y}\frac{1}{\lambda(u)}\,du\sqrt{n}(\hat{B}-B) + o_p(1) \\
&= -h(y)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\bigg(B\eta_i(y) + \int_{y_1}^{y}\frac{1}{\lambda(u)}\,du\;\psi_{\Gamma_2}(Y_i,X_i)\bigg) + o_p(1). \qquad (4.58)
\end{aligned}$$

Note that $\hat{s}\in\tilde{\mathcal{H}}$ for $\gamma_h=2$ was shown in the proof of Lemma 4.2.3.

Proof of (ii): Part (ii) of Theorem 4.2.4 and a Taylor expansion lead similarly to before to

$$\begin{aligned}
&\bar{h}(y)-h(y) \\
&= \bar{h}_{\hat{B}}(y) - h_{\hat{B}}(y) + h_{\hat{B}}(y) - h_B(y) \\
&= \hat{B}h_{\hat{B}}(y)\frac{\bar{h}_1(y)-h_1(y)}{h_1(y)} + \exp\big(\hat{B}\log(h_1(y))\big) - \exp\big(B\log(h_1(y))\big) \\
&\quad + o_p\bigg(\frac{1}{\sqrt{n}}\bigg) \\
&= Bh_B(y)\frac{\bar{h}_1(y)-h_1(y)}{h_1(y)} + \exp\big(B\log(h_1(y))\big)\log(h_1(y))(\hat{B}-B) + o_p\bigg(\frac{1}{\sqrt{n}} + a_n\bigg) \\
&= o_p(a_n)
\end{aligned}$$

uniformly on compact sets $\mathcal{K}\subseteq(y_0,\infty)$.

Proof of (iii): Similar to the proof of (ii),

$$\begin{aligned}
\hat{h}(y)-h(y) &= Bh_B(y)\frac{\hat{h}_1(y)-h_1(y)}{h_1(y)} + \exp\big(B\log(h_1(y))\big)\log(h_1(y))(\tilde{B}-B) \\
&\quad + o_p\bigg(\frac{1}{\sqrt{n}} + |\tilde{B}-B|\bigg) \\
&= -h(y)\int_{y_1}^{y}\frac{1}{\lambda(u)}\,du(\tilde{B}-B) + o_p\bigg(\frac{1}{\sqrt{nh_y^3}}\bigg)
\end{aligned}$$

uniformly on compact sets $\mathcal{K}\subseteq(y_0,\infty)$. The assertion follows from Theorem 4.2.5.

4.6.7 Proof of Lemma 4.2.8

Recall the definition of $\hat{y}_0$ from equation (4.2),

$$\hat{y}_0 = \operatorname*{arg\,min}_{y:\,\hat{\lambda}(y)=0}|y|.$$

Corollary 4.2.14 yields

$$\sup_{u\in[z_a,z_b],x\in\mathbb{R}^{d_X}}\big|\hat{\lambda}(u)-\lambda(u)\big| = o_p(1)$$

as well as

$$\sup_{u\in[z_a,z_b],x\in\mathbb{R}^{d_X}}\left|\frac{\partial}{\partial u}\hat{\lambda}(u)-\frac{\partial}{\partial u}\lambda(u)\right|=o_p(1).$$

Let $\varepsilon>0$. Since $\hat{\lambda}$ is continuous and $\frac{\partial}{\partial u}\lambda(u)\big|_{u=y_0}=-B\neq 0$ there exists with probability converging to one exactly one root $\hat{y}_0$ of $\hat{\lambda}$ on each interval of the form $[-|y_0|-r,|y_0|+r]$ (for fixed $r>0$). Indeed, this root coincides with the estimator $\hat{y}_0$ from above and fulfils

$$\hat{y}_0=y_0+o_p(1).$$

One has for some y^* between $\hat{y}_0$ and y_0

$$\begin{aligned}\hat{\lambda}(y_0)&=\hat{\lambda}(y_0)-\hat{\lambda}(\hat{y}_0)\\&=\frac{\partial}{\partial u}\hat{\lambda}(u)\bigg|_{u=y^*}(y_0-\hat{y}_0)\\&=-\frac{\partial}{\partial u}\lambda(u)\bigg|_{u=y_0}(\hat{y}_0-y_0)+o_p(|\hat{y}_0-y_0|),\end{aligned}$$

so that

$$\hat{y}_0-y_0=\frac{\hat{\lambda}(y_0)}{B}+o_p(|\hat{y}_0-y_0|).$$

Lemma 4.2.1 leads to

$$\begin{aligned}&\sqrt{nh_y}(\hat{y}_0-y_0)\\&=\frac{\sqrt{h_y}}{B\sqrt{n}}\sum_{i=1}^n\Bigg(v(X_i)D_{p,0}(y_0,X_i)\mathcal{K}_{h_y}(y_0-Y_i)-\frac{\partial v(X_i)D_{p,x}(y_0,X_i)}{\partial x_1}\mathcal{K}_{h_y}(y_0-Y_i)\\&\quad+v(X_i)D_{p,y}(y_0,X_i)K_{h_y}(y_0-Y_i)+v(X_i)D_{f,0}(y_0,X_i)-\frac{\partial v(X_i)D_{f,x}(y_0,X_i)}{\partial x_1}\Bigg)\\&\quad+o_p(1).\end{aligned}$$

In the following, a Lindeberg-Feller Theorem is applied to prove asymptotic normality of $\sqrt{nh_y}(\hat{y}_0-y_0)$. By the same reasoning as in (4.37), one has

$$\begin{aligned}&E\Bigg[v(X_1)D_{p,0}(y_0,X_1)\mathcal{K}_{h_y}(y_0-Y_1)-\frac{\partial v(X_1)D_{p,x}(y_0,X_1)}{\partial x_1}\mathcal{K}_{h_y}(y_0-Y_1)\\&\quad+v(X_1)D_{p,y}(y_0,X_1)K_{h_y}(y_0-Y_1)+v(X_1)D_{f,0}(y_0,X_1)-\frac{\partial v(X_1)D_{f,x}(y_0,X_1)}{\partial x_1}\Bigg]\\&=o_p\bigg(\frac{1}{\sqrt{nh_y}}\bigg).\end{aligned}$$

Hence, the asymptotic variance of the dominating term can be calculated as follows:

$$\begin{aligned}&\frac{h_y}{B^2}\operatorname{Var}\Bigg(v(X_1)D_{p,0}(y_0,X_1)\mathcal{K}_{h_y}(y_0-Y_1)-\frac{\partial v(X_1)D_{p,x}(y_0,X_1)}{\partial x_1}\mathcal{K}_{h_y}(y_0-Y_1)\\&\quad+v(X_1)D_{p,y}(y_0,X_1)K_{h_y}(y_0-Y_1)+v(X_1)D_{f,0}(y_0,X_1)-\frac{\partial v(X_1)D_{f,x}(y_0,X_1)}{\partial x_1}\Bigg)\end{aligned}$$

$$= \frac{h_y}{B^2} E\Big[\Big(v(X_1)D_{p,0}(y_0,X_1)\mathcal{K}_{h_y}(y_0-Y_1) - \frac{\partial v(X_1)D_{p,x}(y_0,X_1)}{\partial x_1}\mathcal{K}_{h_y}(y_0-Y_1)$$

$$+ v(X_1)D_{p,y}(y_0,X_1)K_{h_y}(y_0-Y_1) + v(X_1)D_{f,0}(y_0,X_1) - \frac{\partial v(X_1)D_{f,x}(y_0,X_1)}{\partial x_1}\Big)^2\Big]$$

$$+ o(1).$$

Note that apart from the third one all of the terms inside the expectation are bounded and hence (after multiplying with h_y) asymptotically negligible, so that

$$\mathrm{Var}\left(\frac{\sqrt{h_y}}{B\sqrt{n}}\sum_{i=1}^n \Big(v(X_i)D_{p,0}(y_0,X_i)\mathcal{K}_{h_y}(y_0-Y_i) - \frac{\partial v(X_i)D_{p,x}(y_0,X_i)}{\partial x_1}\mathcal{K}_{h_y}(y_0-Y_i)\right.$$

$$\left.+ v(X_i)D_{p,y}(y_0,X_i)K_{h_y}(y_0-Y_i) + v(X_i)D_{f,0}(y_0,X_i) - \frac{\partial v(X_i)D_{f,x}(y_0,X_i)}{\partial x_1}\Big)\right)$$

$$= \frac{h_y}{B^2}E\big[\big(v(X_1)D_{p,y}(y_0,X_1)K_{h_y}(y_0-Y_1)\big)^2\big] + o(1)$$

$$= \frac{h_y}{B^2}\int\int D_{p,y}(y_0,w)^2K_{h_y}(y_0-z)^2v(w)^2f_{Y,X}(z,w)\,dz\,dw + o(1)$$

$$= \frac{1}{B^2}\int v(w)^2D_{p,y}(y_0,w)^2\int \frac{1}{h_y}K\left(\frac{y_0-z}{h_y}\right)^2 f_{Y,X}(z,w)\,dz\,dw + o(1)$$

$$= \frac{1}{B^2}\int v(w)^2D_{p,y}(y_0,w)^2\int K(z)^2 f_{Y,X}(y_0-h_yz,w)\,dz\,dw + o(1)$$

$$= \sigma_{y_0}^2 + o(1).$$

Let

$$Z_{n,i} := \frac{\sqrt{h_y}}{B\sqrt{n}}\Big(v(X_i)D_{p,0}(y_0,X_i)\mathcal{K}_{h_y}(y_0-Y_i) - \frac{\partial v(X_i)D_{p,x}(y_0,X_i)}{\partial x_1}\mathcal{K}_{h_y}(y_0-Y_i)$$

$$+ v(X_i)D_{p,y}(y_0,X_i)K_{h_y}(y_0-Y_i) + v(X_i)D_{f,0}(y_0,X_i) - \frac{\partial v(X_i)D_{f,x}(y_0,X_i)}{\partial x_1}\Big),$$

that is, $\sqrt{nh_y}(\hat{y}_0-y_0) = \sum_{i=1}^n Z_{n,i} + o_p(1)$. Then, the assertion is implied by the Lindeberg Feller Theorem, if

$$\frac{1}{\mathrm{Var}\big(\sum_{i=1}^n(Z_{n,i}-E[Z_{n,i}])\big)}$$

$$\sum_{i=1}^n E\Big[(Z_{n,i}-E[Z_{n,1}])^2 I_{\{|Z_{n,i}-E[Z_{n,1}]|^2>\varepsilon\,\mathrm{Var}\,(\sum_{j=1}^n(Z_{n,j}-E[Z_{n,j}]))\}}\Big] \to 0$$

holds for all $\varepsilon>0$. Since $|Z_{n,i}| \le \frac{C}{\sqrt{nh_y}}$ for a sufficiently large constant $C>0$ this in turn directly follows from $\mathrm{Var}\big(\sum_{i=1}^n(Z_{n,i}-E[Z_{n,i}])\big) = \sigma_{y_0}^2 + o(1)$ and $\sigma_{y_0}^2>0$.

4.6.8 Proof of Theorem 4.2.9

It is started with the first assertion. Define $\tilde{t}_n = t_n + \hat{y}_0 - y_0$. Since referring to Theorem 4.2.8, $\hat{y}_0$ converges to y_0 at a faster rate than t_n, one has $\tilde{t}_n = t_n(1+o_p(1))$. Moreover, a

Taylor expansion leads for some y^* between $\hat{y}_0$ and y_0 to

$$\tilde{h}(\hat{y}_0+t_n)-h(y_0+t_n)=\tilde{h}(y_0+\tilde{t}_n)-\Big(h(y_0+\tilde{t}_n)+\frac{\partial}{\partial y}h(y)\big|_{y=y^*}(y_0-\hat{y}_0)\Big)$$
$$=\tilde{h}(y_0+\tilde{t}_n)-h(y_0+\tilde{t}_n)+\mathcal{O}_p\Big(\frac{1}{\sqrt{nh_y}}\Big),$$

so that it suffices to prove

$$\tilde{h}(y_0+t_n)-h(y_0+t_n)=\mathcal{O}_p\left(\frac{\log(t_n)t_n}{\sqrt{nh_y^3}}+\frac{\log(n)}{\sqrt{nh_y}}\right).$$

To this end, write

$$\begin{aligned}
&\tilde{h}(y_0+t_n)\\
&=\exp\Big(-\tilde{B}\int_{y_1}^{y_0+t_n}\frac{1}{\hat{\lambda}(u)}\,du\Big)\\
&=\exp\Big(-B\int_{y_1}^{y_0+t_n}\frac{1}{\lambda(u)}\,du+(B-\tilde{B})\int_{y_1}^{y_0+t_n}\frac{1}{\lambda(u)}\,du-B\int_{y_1}^{y_0+t_n}\underbrace{\frac{1}{\hat{\lambda}(u)}-\frac{1}{\lambda(u)}}_{=\frac{\lambda(u)-\hat{\lambda}(u)}{\lambda(u)\hat{\lambda}(u)}}\,du\\
&\qquad+(B-\tilde{B})\int_{y_1}^{y_0+t_n}\frac{\lambda(u)-\hat{\lambda}(u)}{\lambda(u)\hat{\lambda}(u)}\,du\Big).
\end{aligned}$$

Due to $\frac{\partial}{\partial u}\lambda(u)\big|_{u=y_0}=-B<0$, one has $|\frac{\partial}{\partial u}\lambda(u)|\geq c|u-y_0|$ for some $c>0$ and u in a sufficiently small neighbourhood of y_0. This in turn implies for all u in this neighbourhood of y_0 and some corresponding y^* between u and y_0 that

$$|\lambda(u)|=\Big|\lambda(y_0)+\frac{\partial}{\partial u}\lambda(u)\big|_{u=y^*}(u-y_0)\Big|\geq c|u-y_0|. \tag{4.59}$$

Further, apply Theorem 4.2.1 to obtain

$$\begin{aligned}
\int_{y_1}^{y_0+t_n}\frac{\lambda(u)-\hat{\lambda}(u)}{\lambda(u)\hat{\lambda}(u)}\,du\ &=\int_{y_1}^{y_0+t_n}\frac{\lambda(u)-\hat{\lambda}(u)}{\lambda(u)(\lambda(u)+\hat{\lambda}(u)-\lambda(u))}\,du\\
&\overset{(4.22)}{=}\int_{y_1}^{y_0+t_n}\frac{\lambda(u)-\hat{\lambda}(u)}{\lambda(u)(\lambda(u)+o_p(t_n))}\,du\\
&\overset{(4.59)}{=}\int_{y_1}^{y_0+t_n}\frac{\lambda(u)-\hat{\lambda}(u)}{\lambda(u)^2}\,du\,(1+o_p(1)).
\end{aligned}$$

For $n\in\mathbb{N}$ sufficiently large and every fixed $\tilde{y}$ in such a small neighbourhood of y_0 that (4.59) is valid, this integral can be bounded for some constant $C>0$ via

$$\begin{aligned}
\int_{y_1}^{y_0+t_n}\frac{\lambda(u)-\hat{\lambda}(u)}{\lambda(u)^2}\,du\ &\leq C\Big|\int_{\tilde{y}}^{y_0+t_n}\frac{1}{(u-y_0)^2}\,du\Big|\sup_{u\in[y_0,y_1]}|\hat{\lambda}(u)-\lambda(u)|(1+o_p(1))\\
&=C\Big|\frac{1}{y_0+t_n-y_0}-\frac{1}{\tilde{y}-y_0}\Big|\sup_{u\in[y_0,y_1]}|\hat{\lambda}(u)-\lambda(u)|(1+o_p(1))
\end{aligned}$$

$$\stackrel{(4.22)}{=} \mathcal{O}_p\bigg(\frac{\log(n)}{t_n\sqrt{nh_y}}\bigg). \tag{4.60}$$

Because $\int_{y_1}^{y_0+t_n} \frac{1}{\lambda(u)}\, du = \mathcal{O}(\log(t_n))$, this results in

$$\begin{aligned}
&\tilde{h}(y_0+t_n)\\
&= h(y_0+t_n)\exp\left((B-\tilde{B})\int_{y_1}^{y_0+t_n}\frac{1}{\lambda(u)}\,du - B\int_{y_1}^{y_0+t_n}\frac{\lambda(u)-\hat{\lambda}(u)}{\lambda(u)^2}\,du\,(1+o_p(1))\right)\\
&= h(y_0+t_n)\left(1+\mathcal{O}_p\left(\frac{\log(t_n)}{\sqrt{nh_y^3}}+\frac{\log(n)}{t_n\sqrt{nh_y}}\right)\right).
\end{aligned} \tag{4.61}$$

The fact

$$h(y_0+t_n) = \mathcal{O}(t_n) \tag{4.62}$$

yields the first assertion.
By the same reasoning as above, it can be shown that

$$\exp\bigg(-\tilde{B}\int_{y_2}^{\hat{y}_0-t_n}\frac{1}{\hat{\lambda}(u)}\,du\bigg) - \exp\bigg(-B\int_{y_2}^{y_0-t_n}\frac{1}{\lambda(u)}\,du\bigg) = \mathcal{O}_p\bigg(\frac{\log(t_n)t_n}{\sqrt{nh_y^3}}+\frac{\log(n)}{\sqrt{nh_y}}\bigg). \tag{4.63}$$

This leads to

$$\begin{aligned}
\tilde{\lambda}_2 &= -\frac{\exp\Big(-\tilde{B}\int_{y_1}^{\hat{y}_0+t_n}\frac{1}{\hat{\lambda}(u)}\,du\Big)}{\exp\Big(-\tilde{B}\int_{y_2}^{\hat{y}_0-t_n}\frac{1}{\hat{\lambda}(u)}\,du\Big)}\\
&= -\frac{\exp\Big(-B\int_{y_1}^{y_0+t_n}\frac{1}{\lambda(u)}\,du\Big)+\mathcal{O}_p\Big(\frac{\log(t_n)t_n}{\sqrt{nh_y^3}}+\frac{\log(n)}{\sqrt{nh_y}}\Big)}{\exp\Big(-B\int_{y_2}^{y_0-t_n}\frac{1}{\lambda(u)}\,du\Big)+\mathcal{O}_p\Big(\frac{\log(t_n)t_n}{\sqrt{nh_y^3}}+\frac{\log(n)}{\sqrt{nh_y}}\Big)}\\
&\stackrel{(4.62)}{=} -\lambda_2\frac{h(y_0+t_n)}{h(y_0-t_n)}+\mathcal{O}_p\left(\frac{\log(t_n)}{\sqrt{nh_y^3}}+\frac{\log(n)}{t_n\sqrt{nh_y}}\right)\\
&= -\lambda_2\frac{h(y_0)+\frac{\partial}{\partial y}h(y_0)t_n+\frac{\frac{\partial^2}{\partial y^2}h(y_0)}{2}t_n^2+o(t_n^2)}{h(y_0)-\frac{\partial}{\partial y}h(y_0)t_n+\frac{\frac{\partial^2}{\partial y^2}h(y_0)}{2}t_n^2+o(t_n^2)}+\mathcal{O}_p\left(\frac{\log(t_n)}{\sqrt{nh_y^3}}+\frac{\log(n)}{t_n\sqrt{nh_y}}\right)\\
&= \lambda_2\frac{\frac{\partial}{\partial y}h(y_0)+\frac{\frac{\partial^2}{\partial y^2}h(y_0)}{2}t_n+o(t_n)}{\frac{\partial}{\partial y}h(y_0)-\frac{\frac{\partial^2}{\partial y^2}h(y_0)}{2}t_n+o(t_n)}+\mathcal{O}_p\left(\frac{\log(t_n)}{\sqrt{nh_y^3}}+\frac{\log(n)}{t_n\sqrt{nh_y}}\right)\\
&= \lambda_2\left(1+\frac{\frac{\partial^2}{\partial y^2}h(y_0)t_n+o(t_n)}{\frac{\partial}{\partial y}h(y_0)-\frac{\frac{\partial^2}{\partial y^2}h(y_0)}{2}t_n+o(t_n)}\right)+\mathcal{O}_p\left(\frac{\log(t_n)}{\sqrt{nh_y^3}}+\frac{\log(n)}{t_n\sqrt{nh_y}}\right)\\
&= \lambda_2+\frac{\lambda_2\frac{\partial^2}{\partial y^2}h(y_0)t_n}{\frac{\partial}{\partial y}h(y_0)}+\mathcal{O}_p\left(\frac{\log(t_n)}{\sqrt{nh_y^3}}+\frac{\log(n)}{t_n\sqrt{nh_y}}\right)+o_p(t_n),
\end{aligned}$$

that is, (4.30) holds.
To obtain the optimal rate in (4.30) try to choose t_n such that at least two of the terms t_n, $\frac{\log(t_n)}{\sqrt{nh_y^3}}$ and $\frac{\log(n)}{t_n\sqrt{nh_y}}$ converge to zero at the same rate. Note that h_x, h_y and thus the optimal t_n will converge to zero at a polynomial rate (with respect to n). Therefore, the choice

$$\frac{\log(\bar{t}_n)}{\sqrt{nh_y^3}} = \frac{\log(n)}{\bar{t}_n\sqrt{nh_y}}$$

leads to $\bar{t}_n = h_y$ and $\tilde{\lambda}_2 - \lambda_2 = \mathcal{O}_p(h_y)$, while

$$\tilde{t}_n = \frac{\log(\tilde{t}_n)}{\sqrt{nh_y^3}}$$

leads to $\tilde{t}_n = \frac{\log(n)}{\sqrt{nh_y^3}}$ and $\tilde{\lambda}_2 - \lambda_2 = \mathcal{O}_p(h_y)$. The remaining possibility

$$t_n = \frac{\log(n)}{t_n\sqrt{nh_y}},$$

that is

$$t_n = \left(\frac{\log(n)^2}{nh_y}\right)^{\frac{1}{4}} = o(h_y),$$

results in

$$\frac{\log(t_n)}{\sqrt{nh_y^3}} = \mathcal{O}\left(\frac{\log(n)}{\sqrt{nh_y^3}}\right) = \mathcal{O}\left(t_n\left(\frac{\log(n)^2}{nh_y^5}\right)^{\frac{1}{4}}\right) \overset{\textbf{(B3')}}{=} o(t_n),$$

which proves the last part of Theorem 4.2.9.

4.6.9 Proof of Theorem 4.2.11

The proof mainly follows from Theorems 4.2.6 and 4.2.9. While the assertion in (i)a is a direct consequence of part (i) in Theorem 4.2.6, the proof of Theorem 4.2.9 has to be slightly adjusted to show the assertion in (i)b. Let $y \in \mathcal{K}\backslash(-\infty, y_0 + t_n)$. By the same reasoning as there, one can show for some sufficiently large constant $C > 0$ that

$$\begin{aligned}
&\sup_{y\in\mathcal{K}\backslash(-\infty,y_0+t_n)} |\hat{h}(y) - h(y)| \\
&= \sup_{y\in\mathcal{K}\backslash(-\infty,y_0+t_n)} h(y)\left|\exp\left((B-\hat{B})\int_{y_1}^{y}\frac{1}{\lambda(u)}\,du - B\int_{y_1}^{y}\frac{\lambda(u)-\hat{\lambda}(u)}{\lambda(u)^2}\,du\,(1+o_p(1))\right) - 1\right| \\
&\leq \sup_{y\in\mathcal{K}\backslash(-\infty,y_0+t_n)} \frac{Ch(y)}{y-y_0}\sup_{u\in\mathcal{K}}|\hat{\lambda}(u)-\lambda(u)| + \mathcal{O}_p\left(\frac{\log(t_n)}{\sqrt{n}}\right) \\
&= \mathcal{O}_p\left(\frac{\log(t_n)}{\sqrt{n}} + \frac{\log(n)}{\sqrt{nh_y}}\right),
\end{aligned}$$

where the second to last equation follows similarly to (4.60). Remember $\hat{B} - B = \mathcal{O}_p\big(n^{-\frac{1}{2}}\big)$. Due to $h(y) = \mathcal{O}_p(t_n)$ for all $y \in \mathcal{K}\cap[0, y_0 + t_n)$ and $t_n \sim \frac{\log(n)}{\sqrt{nh_y}}$, this implies (i)b.

To prove the assertion in (ii), recall the definition of $\tilde{h}$:

$$\tilde{h}(y) = \begin{cases} \exp\Big(- \tilde{B} \int_{y_1}^{y} \frac{1}{\hat{\lambda}(u)}\, du \Big), & y \geq \hat{y}_0 + t_n, \\ \frac{y - \hat{y}_0}{t_n} \tilde{h}(\hat{y}_0 + t_n), & y \in (\hat{y}_0, \hat{y}_0 + t_n), \\ 0, & y = \hat{y}_0, \\ \frac{\hat{y}_0 - y}{t_n} \tilde{h}(\hat{y}_0 - t_n), & y \in (\hat{y}_0 - t_n, \hat{y}_0), \\ \tilde{\lambda}_2 \exp\Big(- \tilde{B} \int_{y_2}^{y} \frac{1}{\hat{\lambda}(u)}\, du \Big), & y \leq \hat{y}_0. \end{cases}$$

Let $\varepsilon > 0$. In the proof of Theorem 4.2.9, it was shown that (compare (4.61))

$$\sup_{y \in \mathcal{K} \setminus (-\infty, y_0 + t_n]} |\tilde{h}(y) - h(y)| = \mathcal{O}_p \left(\frac{\log(t_n)}{\sqrt{n h_y^3}} + \frac{\log(n)}{t_n \sqrt{n h_y}} \right) = o_p(t_n)$$

as well as (compare (4.63))

$$\sup_{y \in \mathcal{K} \setminus [y_0 - t_n, \infty)} \left| \frac{\tilde{h}(y)}{\tilde{\lambda}_2} - \frac{h(y)}{\lambda_2} \right| = \mathcal{O}_p \left(\frac{\log(t_n)}{\sqrt{n h_y^3}} + \frac{\log(n)}{t_n \sqrt{n h_y}} \right) = o_p(t_n).$$

The rate of the convergence here is equal to that of $\tilde{B}$ to B. Therefore,

$$\begin{aligned} \sup_{y \in \mathcal{K} \setminus [y_0 - \varepsilon, \infty)} \left| \tilde{h}(y) - h(y) \right| &= \sup_{y \in \mathcal{K} \setminus [y_0 - \varepsilon, \infty)} \left| \tilde{\lambda}_2 \frac{\tilde{h}(y)}{\tilde{\lambda}_2} - \lambda_2 \frac{h(y)}{\lambda_2} \right| \\ &= \sup_{y \in \mathcal{K} \setminus [y_0 - t_n, \infty)} \left| (\tilde{\lambda}_2 - \lambda_2) \frac{\tilde{h}(y)}{\tilde{\lambda}_2} + \lambda_2 \Big(\frac{\tilde{h}(y)}{\tilde{\lambda}_2} - \frac{h(y)}{\lambda_2} \Big) \right| \\ &= \mathcal{O}_p\big(|\tilde{\lambda}_2 - \lambda_2| + t_n \big) \\ &= \mathcal{O}_p(t_n). \end{aligned}$$

The fact that $|h(y)| = \mathcal{O}_p(t_n)$ for all $y \in [y_0 - t_n, y_0 + t_n]$ and $t_n \sim \big(\frac{\log(n)^2}{n h_y} \big)^{\frac{1}{4}}$ complete the proof.

4.6.10 Proof of Lemma 4.2.12

Recall

$$a_n = \left(\frac{\log(n)}{n h_x^{d_X}} \right)^{\frac{1}{2}} = o\big(n^{-\frac{1}{4}}\big)$$

and let $(c_n)_{n \in \mathbb{N}}$ be a sequence such that $c_n \to \infty$ and there exist constants $\delta_1, \delta_2 > 0$ with $|c_n| \leq \delta_1 n^{\delta_2}$.

First, the expectations can be calculated as in Section 1.1 to obtain

$$\sup_{y \in [z_a, z_b], x \in \mathbb{R}^{d_X}} v(x) |E[\hat{p}(y, x)] - p(y, x)| = \mathcal{O}\big(h_x^m + h_y^m\big) = o_p\big(n^{-\frac{1}{4}}\big),$$

$$\sup_{y \in [z_a, z_b], x \in \mathbb{R}^{d_X}} v(x) |E[\hat{p}_x(y, x)] - p_x(y, x)| = \mathcal{O}\big(h_x^m + h_y^m\big) = o_p\big(n^{-\frac{1}{4}}\big),$$

$$\sup_{y \in [z_a, z_b], x \in \mathbb{R}^{d_X}} v(x) |E[\hat{p}_y(y, x)] - p_y(y, x)| = \mathcal{O}\big(h_x^m + h_y^m\big) = o_p\big(n^{-\frac{1}{4}}\big),$$

$$\sup_{x\in\mathbb{R}^{d_X}} v(x)|E[\hat{f}(x)] - f(x)| = \mathcal{O}\big(h_x^m\big) \qquad = o_p\big(n^{-\frac{1}{4}}\big),$$

$$\sup_{x\in\mathbb{R}^{d_X}} v(x)|E[\hat{f}_x(x)] - f_x(x)| = \mathcal{O}\big(h_x^m\big) \qquad = o_p\big(n^{-\frac{1}{4}}\big).$$

As already mentioned, basically the results of Hansen (2008) are applied. Note that all of the estimators in the lemma are very similar to the estimator $\hat{\Psi}$ in (4.31) although the corresponding Z_i may be different. For each estimator, one has to consider a different "dimension" $\tilde{d}$ and "kernel" of $\hat{\Psi}$. In a few moments, it will be clear, how this is meant. First, write

$$\hat{p}(y,x) = \frac{1}{nh_x^{d_X}}\sum_{i=1}^n \underbrace{\mathcal{K}_{h_y}(y-Y_i)}_{=Z_i}\mathbf{K}\left(\frac{x-X_i}{h_x}\right) \qquad \rightsquigarrow \tilde{d} = d_X,$$

$$\hat{p}_x(y,x) = \frac{1}{nh_x^{d_X}}\sum_{i=1}^n \underbrace{\mathcal{K}_{h_y}(y-Y_i)}_{=Z_i}\frac{\partial}{\partial x_1}\mathbf{K}\left(\frac{x-X_i}{h_x}\right) \qquad \rightsquigarrow \tilde{d} = d_X,$$

$$\hat{p}_y(y,x) = \frac{1}{nh_x^{d_X}}\sum_{i=1}^n K_{h_y}(y-Y_i)\mathbf{K}\left(\frac{x-X_i}{h_x}\right) \qquad \rightsquigarrow \tilde{d} = d_X+1, Z_i \equiv 1,$$

$$\hat{f}(x) = \frac{1}{nh_x^{d_X}}\sum_{i=1}^n \mathbf{K}\left(\frac{x-X_i}{h_x}\right) \qquad \rightsquigarrow \tilde{d} = d_X, Z_i \equiv 1,$$

$$\hat{f}_x(x) = \frac{1}{nh_x^{d_X+1}}\sum_{i=1}^n \frac{\partial}{\partial x_1}\mathbf{K}\left(\frac{x-X_i}{h_x}\right) \qquad \rightsquigarrow \tilde{d} = d_X, Z_i \equiv 1.$$

Apart from some assumptions on the dependency structure (that are automatically fulfilled here due to the independence), the Z_i are not allowed to depend on n or y and have to fulfil some boundedness conditions, like for example

$$\sup_{x\in\mathbb{R}^{d_X}} E[|Z_1|\,|X_1 = x]f_X(x)v(x) < \infty.$$

In the last three cases, both of the requirements are fulfilled so that the assertion directly follows by Hansen (2008) and assumption (**B3**). In the first two cases, $Z_i = \mathcal{K}_{h_y}(y-Y_i)$ depends on n as well as on y, so that the results of Hansen (2008) can not be applied directly. The idea there was to construct an appropriate grid of the set $\{(y,x) : ||(y,x)|| \leq c_n\}$ such that the proof of equation (4.31) can be reduced to a non uniform version.
Since in the remaining cases $|Z_i|$ is still bounded by some constant $K > 0$, one can show for a sufficiently large $M > 0$

$$P\big(|\hat{\Psi}(y,x) - E[\hat{\Psi}(y,x)]| > Ma_n\big) \leq 4n^{-\frac{M}{64+6K}} \quad \text{for all } y \in [z_a, z_b], x \in \operatorname{supp}(v)$$

analogously to Hansen (2008). In contrast to the proof there, the construction of the grid used in the next step is slightly more sophisticated. The set $\{(y,x) : ||(y,x)|| \leq c_n\}$ can be split into

$$N \leq \frac{c_n^{d_X+1}}{a_n^{(d_X+1)}h_y h_x^{d_X(d_X+2)}}$$

subsets of the form $A_j = \{(y,x) : |y - y_j| \leq a_n h_y h_x^{d_X}, ||x - x_j|| \leq a_n h_x^{d_X+1}\}$. Then, one has for $(y,x) \in A_j$ and an appropriate constant $\bar{K} > 0$

$$\begin{aligned}
|\hat{p}(y,x) - \hat{p}(y_j,x_j)| &\leq \frac{1}{nh_x^{d_X}} \Bigg| \sum_{i=1}^n \mathcal{K}_{h_y}(y - Y_i)\mathbf{K}\Big(\frac{x - X_i}{h_x}\Big) - \mathcal{K}_{h_y}(y_j - Y_i)\mathbf{K}\Big(\frac{x_j - X_i}{h_x}\Big)\Bigg| \\
&= \frac{1}{nh_x^{d_X}} \Bigg| \sum_{i=1}^n \big(\mathcal{K}_{h_y}(y - Y_i) - \mathcal{K}_{h_y}(y_j - Y_i)\big)\mathbf{K}\Big(\frac{x - X_i}{h_x}\Big) \\
&\qquad + \mathcal{K}_{h_y}(y_j - Y_i)\Big(\mathbf{K}\Big(\frac{x - X_i}{h_x}\Big) - \mathbf{K}\Big(\frac{x_j - X_i}{h_x}\Big)\Big)\Bigg| \\
&\leq \bar{K}^2 a_n,
\end{aligned}$$

which means that one can proceed analogously to Hansen (2008) to obtain for a sufficiently large $M > 0$

$$\begin{aligned}
&P\Big(\sup_{||(y,x)|| \leq c_n} |\hat{\Psi}(y,x) - E[\hat{\Psi}(y,x)]| > Ma_n\Big) \\
&\leq N \max_{j=1,\dots,N} P\Big(\big|\hat{\Psi}(y_j,x_j) - E[\hat{\Psi}(y_j,x_j)]\big| > \frac{M}{2}a_n\Big) \\
&\leq 4c_n^{d_X+1} a_n^{-(d_X+1)} h_y^{-1} h_x^{-d_X(d_X+2)} n^{-\frac{M}{128+12K}} \\
&\to 0.
\end{aligned}$$

As was shown above and in Section 1.1, it holds that

$$\sup_{y \in \mathbb{R},\, x \in \mathbb{R}^{d_X}} v(x)|E[\hat{p}(y,x)] - p(y,x)| = \mathcal{O}\big(h_y^m + h_x^m\big),$$

so that

$$\sup_{y \in [z_a,z_b],\, x \in \mathbb{R}^{d_X}} v(x)|\hat{p}(y,x) - p(y,x)]| = \mathcal{O}\Bigg(h_y^m + h_x^m + \sqrt{\frac{\log(n)}{nh_x^{d_X}}}\Bigg) \overset{((\mathbf{B3}))}{=} o\big(n^{-\frac{1}{4}}\big)$$

follows. The same reasoning can be applied for $\hat{p}_x$. □

4.6.11 Proof of Corollary 4.2.14

The proof is very similar to that of Lemma 4.2.12. Again, only derivatives with respect to x_1 are considered. First, convergence of the expectations can be shown as before using integration by parts:

$$\begin{aligned}
E\Big[v(x)\frac{\partial}{\partial u}\hat{p}_y(u,x)\Big] &= v(x)E\Big[\frac{\partial}{\partial u}K_{h_y}(u - Y_1)\mathbf{K}_{h_x}(x - X_1)\Big] \\
&= -v(x)\int\int \frac{\partial}{\partial z}K_{h_y}(u - z)\mathbf{K}_{h_x}(x - w)f_{Y,X}(z,w)\,dz\,dw \\
&= v(x)\int\int K_{h_y}(u - z)\mathbf{K}_{h_x}(x - w)\frac{\partial}{\partial z}f_{Y,X}(z,w)\,dz\,dw
\end{aligned}$$

$$= v(x)\int\int K(z)\mathbf{K}(w)\frac{\partial}{\partial u}f_{Y,X}(u-h_y z, x-h_x w)\,dz\,dw$$

$$= v(x)\frac{\partial}{\partial u}f_{Y,X}(u,x)+o\left(\frac{1}{\sqrt{n}}\right), \tag{4.64}$$

$$\begin{aligned}
E\left[v(x)\frac{\partial}{\partial u}\hat{p}_x(u,x)\right] &= v(x)E\left[K_{h_y}(u-Y_1)\frac{\partial}{\partial x_1}\mathbf{K}_{h_x}(x-X_1)\right]\\
&= -v(x)\int\int K_{h_y}(u-z)\frac{\partial}{\partial w_1}\mathbf{K}_{h_x}(x-w)f_{Y,X}(z,w)\,dz\,dw\\
&= v(x)\int\int K_{h_y}(u-z)\mathbf{K}_{h_x}(x-w)\frac{\partial}{\partial w_1}f_{Y,X}(z,w)\,dz\,dw\\
&= v(x)\int\int K(z)\mathbf{K}(w)\frac{\partial}{\partial x_1}f_{Y,X}(u-h_y z, x-h_x w)\,dz\,dw\\
&= v(x)\frac{\partial}{\partial x_1}f_{Y,X}(u,x)+o\left(\frac{1}{\sqrt{n}}\right)
\end{aligned}$$

and

$$\begin{aligned}
E\left[v(x)\frac{\partial}{\partial x_1}\hat{f}_x(x)\right] &= v(x)E\left[\frac{\partial^2}{\partial x_1^2}\mathbf{K}_{h_x}(x-X_1)\right]\\
&= v(x)\int\int \frac{\partial^2}{\partial w_1^2}\mathbf{K}_{h_x}(x-w)f_X(w)\,dw\\
&= -v(x)\int\int \frac{\partial}{\partial w_1}\mathbf{K}_{h_x}(x-w)\frac{\partial}{\partial w_1}f_X(w)\,dw\\
&= v(x)\int\int \mathbf{K}_{h_x}(x-w)\frac{\partial^2}{\partial w_1^2}f_X(w)\,dw\\
&= v(x)\int\int K(z)\mathbf{K}(w)\frac{\partial^2}{\partial x_1^2}f_X(x-h_x w)\,dw\\
&= \frac{\partial^2}{\partial x_1^2}f_X(x)+o\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

For each estimator (i.e. for each $\hat{\Psi}$), one simply has to redefine the corresponding "dimension" $\tilde{d}$, the "kernel" or the random variables Z_i. For example, when treating

$$\frac{\partial}{\partial y}\hat{p}_y(y,x) = \frac{1}{nh_x^{d_X}}\sum_{i=1}^n \frac{\partial}{\partial y}K_{h_y}(y-Y_i)\mathbf{K}\left(\frac{x-X_i}{h_x}\right) = \frac{1}{nh_x^{d_X}h_y^2}\sum_{i=1}^n K'\left(\frac{y-Y_i}{h_y}\right)\mathbf{K}\left(\frac{x-X_i}{h_x}\right)$$

or

$$\frac{\partial}{\partial y}\hat{p}_x(y,x) = \frac{1}{nh_x^{d_X}}\sum_{i=1}^n K_{h_y}(y-Y_i)\frac{\partial}{\partial x_1}\mathbf{K}\left(\frac{x-X_i}{h_x}\right),$$

set $\tilde{d} = d_X+1, Z_i = 1$ and define the "kernel" functions by

$$(y,x)\mapsto K'(y)\mathbf{K}(x) \quad \text{and} \quad (y,x)\mapsto K(y)\frac{\partial}{\partial x_1}\mathbf{K}(x),$$

respectively. Now, Theorem 2 of Hansen (2008) can be applied directly so that

$$\sup_{u\in[z_a,z_b]} v(x)\left|\frac{\partial^2}{\partial x_1^2}\hat{f}_X(x)-\frac{\partial^2}{\partial x_1^2}f_X(x)\right| = \mathcal{O}_p\left(\sqrt{\frac{\log(n)}{nh_x^{d_X+4}}}\right) \qquad = o_p(1),$$

$$\sup_{u\in[z_a,z_b]} v(x)\left|\frac{\partial}{\partial u}\hat{p}_y(u,x)-\frac{\partial}{\partial u}p_y(u,x)\right| = \mathcal{O}_p\left(\sqrt{\frac{\log(n)}{nh_y^3h_x^{d_X}}}\right) = o_p(1), \tag{4.65}$$

$$\sup_{u\in[z_a,z_b]} v(x)\left|\frac{\partial}{\partial u}\hat{p}_x(u,x)-\frac{\partial}{\partial u}p_x(u,x)\right| = \mathcal{O}_p\left(\sqrt{\frac{\log(n)}{nh_yh_x^{d_X+2}}}\right) = o_p(1). \tag{4.66}$$

Consider

$$\hat{\lambda}(u|x) = \frac{\hat{\Phi}_x(u,x)}{\hat{\Phi}_y(u,x)}.$$

Due to

$$\frac{\partial}{\partial u}\hat{\Phi}_y(u,x) = \frac{\frac{\partial}{\partial u}\hat{p}_y(u,x)}{\hat{f}(x)}, \qquad \frac{\partial}{\partial u}\hat{\Phi}_x(u,x) = \frac{\frac{\partial}{\partial u}\hat{p}_x(u,x)}{\hat{f}(x)} - \frac{\frac{\partial}{\partial u}\hat{p}(u,x)\hat{f}_x(x)}{\hat{f}(x)^2}$$

and the fact that not the values of the weight function, but only the compactness of its support was used, the foregoing convergence leads to

$$\begin{aligned}
&\sup_{u\in[z_a,z_b],x\in\mathbb{R}^{d_X}} v(x)\left|\frac{\partial}{\partial u}\hat{\lambda}(u|x)-\frac{\partial}{\partial u}\lambda(u|x)\right| \\
&= \sup_{u\in[z_a,z_b],x\in\mathbb{R}^{d_X}} v(x)\left|\frac{\partial}{\partial u}\left(\frac{\hat{\Phi}_x(u,x)}{\hat{\Phi}_y(u,x)}\right)-\frac{\partial}{\partial u}\left(\frac{\Phi_x(u,x)}{\Phi_y(u,x)}\right)\right| \\
&= \sup_{u\in[z_a,z_b],x\in\mathbb{R}^{d_X}} v(x)\left|\frac{\hat{\Phi}_y(u,x)\frac{\partial}{\partial u}\hat{\Phi}_x(u,x)-\hat{\Phi}_x(u,x)\frac{\partial}{\partial u}\hat{\Phi}_y(u,x)}{\hat{\Phi}_y(u,x)^2}\right. \\
&\qquad \left. -\frac{\Phi_y(u,x)\frac{\partial}{\partial u}\Phi_x(u,x)-\Phi_x(u,x)\frac{\partial}{\partial u}\Phi_y(u,x)}{\Phi_y(u,x)^2}\right| \\
&= o_p(1).
\end{aligned}$$

Since $\frac{\partial}{\partial u}\lambda(u|x)$ is bounded on $[z_a,z_b]\times\operatorname{supp}(v)$ and $\frac{\partial}{\partial u}\hat{\lambda}(u|x)$ is uniformly consistent, this in turn implies

$$\begin{aligned}
\sup_{u\in[z_a,z_b]} v(x)\left|\frac{\partial}{\partial u}\hat{\lambda}(u)-\frac{\partial}{\partial u}\lambda(u)\right| &= \sup_{u\in[z_a,z_b]} v(x)\left|\frac{\partial}{\partial u}\int v(x)\hat{\lambda}(u|x)\,dx-\frac{\partial}{\partial u}\int v(x)\lambda(u|x)\,dx\right| \\
&\leq \int \sup_{u\in[z_a,z_b],x\in\mathbb{R}^{d_X}} v(x)\left|\frac{\partial}{\partial u}\hat{\lambda}(u|x)-\frac{\partial}{\partial u}\lambda(u|x)\right|dx \\
&= o_p(1).
\end{aligned}$$

Basically, the last part of Corollary 4.2.14 can be shown analogously to the other parts. First, the expectations need to be calculated, this time using integration by parts twice. Second,

$$\begin{aligned}
\frac{\partial^2}{\partial y^2}\hat{p}_y(y,x) &= \frac{1}{nh_x^{d_X}}\sum_{i=1}^n \frac{\partial^2}{\partial y^2}K_{h_y}(y-Y_i)\mathbf{K}\left(\frac{x-X_i}{h_x}\right) \\
&= \frac{1}{nh_x^{d_X}h_y^3}\sum_{i=1}^n K''\left(\frac{y-Y_i}{h_y}\right)\mathbf{K}\left(\frac{x-X_i}{h_x}\right)
\end{aligned}$$

and

$$\frac{\partial^2}{\partial y^2}\hat{p}_x(y,x) = \frac{1}{nh_x^{d_X}h_y^2}\sum_{i=1}^n K'\Big(\frac{y-Y_i}{h_y}\Big)\frac{\partial}{\partial x_1}\mathbf{K}\Big(\frac{x-X_i}{h_x}\Big)$$

can be treated as before to prove

$$\sup_{u\in[z_a,z_b]} v(x)\left|\frac{\partial^2}{\partial u^2}\hat{p}_y(u,x) - \frac{\partial^2}{\partial u^2}p_y(u,x)\right| = \mathcal{O}_p\left(\sqrt{\frac{\log(n)}{nh_y^5h_x^{d_X}}}\right) = o_p(1),$$

$$\sup_{u\in[z_a,z_b]} v(x)\left|\frac{\partial^2}{\partial u^2}\hat{p}_x(u,x) - \frac{\partial^2}{\partial u^2}p_x(u,x)\right| = \mathcal{O}_p\left(\sqrt{\frac{\log(n)}{nh_y^3h_x^{d_X+2}}}\right) = o_p(1).$$

Third, show uniform consistency of $\frac{\partial^2}{\partial u^2}\hat{\Phi}_y(u,x)$ and $\frac{\partial^2}{\partial u^2}\hat{\Phi}_x(u,x)$ to obtain first the uniform consistency of $\frac{\partial^2}{\partial u^2}\hat{\lambda}(u|x)$. Then, this can be used together with the boundedness of $\frac{\partial^2}{\partial u^2}\lambda(u|x)$ to obtain

$$\sup_{u\in[z_a,z_b]} v(x)\left|\frac{\partial^2}{\partial u^2}\hat{\lambda}(u) - \frac{\partial^2}{\partial u^2}\lambda(u)\right| = o_p(1).$$

□

4.6.12 Proof of Lemma 4.2.15

Note that equation (1.8) can be applied to the conditional distribution function, so that Lemma 4.2.12 yields

$$\begin{aligned}
&v(x)(\hat{F}_{Y|X}(y|x) - F_{Y|X}(y|x))\\
&= v(x)\Big(\frac{\hat{p}(y,x)}{\hat{f}_X(x)} - \frac{p(y,x)}{f_X(x)}\Big)\\
&= \frac{v(x)}{f_X(x)}(\hat{p}(y,x) - p(y,x)) - \frac{p(y,x)v(x)}{f_X(x)^2}(\hat{f}_X(x) - f_X(x)) + o_p\Big(\frac{1}{\sqrt{n}}\Big)\\
&= \frac{1}{n}\sum_{i=1}^n v(x)\Big(\frac{1}{f_X(x)}\mathcal{K}_{h_y}(y-Y_i)\mathbf{K}_{h_x}(x-X_i) - \frac{p(y,x)}{f_X(x)^2}\mathbf{K}_{h_x}(x-X_i)\Big) + o_p\Big(\frac{1}{\sqrt{n}}\Big)
\end{aligned}$$

uniformly in $x \in \mathbb{R}^{d_X}$. Now, the framework is the same as for Lemma 4.2.12, so that one can proceed exactly as there. It remains to extend the uniform convergence to the estimated quantile function, which can be shown similar to the proof of Lemma 2.8.1. Due to Lemma 4.2.12, one has

$$v(x)\Big(\frac{\partial}{\partial y}\hat{F}_{Y|X}(y|x) - \frac{\partial}{\partial y}F_{Y|X}(y|x)\Big) = v(x)\Big(\frac{\hat{p}_y(y,x)}{\hat{f}(x)} - \frac{p_y(y,x)}{f(x)}\Big) = o_p(1)$$

uniformly on compact sets. Let $z_a = F^{-1}_{Y|X}(\tau_a|x)$ and $z_b = F^{-1}_{Y|X}(\tau_b|x)$. Since $F_{Y|X}(y|x)$ is strictly increasing in y, or more precisely

$$\inf_{y\in[z_a,z_b],x\in\text{supp}(v)}\frac{\partial}{\partial y}F_{Y|X}(y|x) = \inf_{y\in[z_a,z_b],x\in\text{supp}(v)} f_\varepsilon\Big(\frac{h(y)-g(x)}{\sigma(x)}\Big)\frac{h'(y)}{\sigma(x)} > 0,$$

$\hat{F}_{Y|X}(y|x)$ is strictly increasing with respect to $y \in [z_a, z_b]$ for all $x \in \text{supp}(v)$ with probability converging to one. Hence, one can assume $\hat{F}_{Y|X}(y|x)$ to be invertible. Applying a Taylor expansion with respect to $\tau \in [\tau_a, \tau_b]$ leads for some appropriate function $\tilde{F}(\tau|\cdot)$ between $\hat{F}^{-1}_{Y|X}(\tau|\cdot)$ and $F^{-1}_{Y|X}(\tau|\cdot)$ to

$$\begin{aligned}
0 &= \hat{F}_{Y|X}(\hat{F}^{-1}_{Y|X}(\tau|x)|x) - F_{Y|X}(F^{-1}_{Y|X}(\tau|x)|x) \\
&= \hat{F}_{Y|X}(F^{-1}_{Y|X}(\tau|x)|x) + \hat{f}_{Y|X}(\tilde{F}(\tau,x)|x)\big(\hat{F}^{-1}_{Y|X}(\tau|x) - F^{-1}_{Y|X}(\tau|x)\big) \\
&\quad - F_{Y|X}(F^{-1}_{Y|X}(\tau|x)|x) \\
&= \hat{F}_{Y|X}(F^{-1}_{Y|X}(\tau|x)|x) - F_{Y|X}(F^{-1}_{Y|X}(\tau|x)|x) \\
&\quad + f_{Y|X}(F^{-1}_{Y|X}(\tau|x)|x)\big(\hat{F}^{-1}_{Y|X}(\tau|x) - F^{-1}_{Y|X}(\tau|x)\big) + o_p\big(\hat{F}^{-1}_{Y|X}(\tau|x) - F^{-1}_{Y|X}(\tau|x)\big)
\end{aligned}$$

uniformly in $x \in \operatorname{supp}(v)$, which in turn results in

$$\begin{aligned}
&\hat{F}^{-1}_{Y|X}(\tau|x) - F^{-1}_{Y|X}(\tau|x) \\
&= -\frac{\hat{F}_{Y|X}(F^{-1}_{Y|X}(\tau|x)|x) - F_{Y|X}(F^{-1}_{Y|X}(\tau|x)|x)}{f_{Y|X}(F^{-1}_{Y|X}(\tau|x)|x)}(1 + o_p(1)) \\
&= -\frac{(1 + o_p(1))}{f_{Y|X}(F^{-1}_{Y|X}(\tau|x)|x)} \frac{1}{n} \sum_{i=1}^{n} \bigg(\frac{1}{f_X(x)} \mathcal{K}_{h_y}(F^{-1}_{Y|X}(\tau|x) - Y_i)\mathbf{K}_{h_x}(x - X_i) \\
&\quad - \frac{p(F^{-1}_{Y|X}(\tau|x), x)}{f_X(x)^2} \mathbf{K}_{h_x}(x - X_i) \bigg) \qquad (4.67)
\end{aligned}$$

uniformly in $x \in \operatorname{supp}(v)$. Because of $f_{Y|X}(F^{-1}_{Y|X}(\tau|x)|x) > 0$ for all $x \in \operatorname{supp}(v)$, the second assertion follows from the first one. □

5

Testing for a Parametric Transformation Function

This Chapter has arisen from from a research project with Natalie Neumeyer and Ingrid Van Keilegom. A preprint (Kloodt, Neumeyer, and Van Keilegom (2019)), which is based on the findings of this chapter and which contains most of the results, can be found on https://arxiv.org/pdf/1907.01223.pdf.

As so often in this thesis, this chapter starts with a model equation namely

$$h(Y) = g(X) + \varepsilon. \tag{5.1}$$

As before, refer to h and g as the transformation and regression function, respectively, as before ε is assumed to be independent of X with $E[\varepsilon] = 0$ and as before, several properties like, e.g., smoothness or monotonicity of h will be assumed to build the theory on these. But contrary to the last chapters, the aim of this chapter does not consist in testing for a parametric regression function or in identification and estimation results for heteroscedastic errors, but instead in providing a goodness of fit test for the null hypothesis of a parametric transformation function.

Up to now, testing for a parametric transformation function has not attracted too much attention in the literature. Mu and He (2007) developed a testing procedure in the context of quantile regression in fully parametric transformation models. Neumeyer et al. (2016) provided a procedure to test for a parametric transformation function in (heteroscedastic) semiparametric transformation models with nonparametric regression and variance functions. The main idea there is to test for independence of the covariate and the (appropriately estimated and standardized) residuals ϵ. This can be done in several ways, e.g., by comparing the appropriately estimated distribution or characteristic function of the joint distribution of (X, ϵ) to the product of the estimated marginal distribution or characteristic functions. Similar techniques were used in the context of nontransformed regression models before, e.g., see Akritas and Van Keilegom (2001) or Jiménez Gamero, Muñoz García, and Pino Mejías (2005) for only one reference to each of the approaches. Neumeyer et al. (2016) used an estimator of the distribution function to test for independence of X and ϵ, while

Hušková et al. (2018) provided a similar test in the homoscedastic case, which is based on the comparison of estimated characteristic functions. The latter test was extended to heteroscedastic errors by Hušková et al. (2019) very recently. Under the assumption of a parametric regression function, Szydłowski (2017) used an L^2-distance between the nonparametric estimator of Chen (2002) and the parametric class, which is tested for, to create a goodness of fit test for a parametric transformation function.
Throughout this chapter, homoscedastic errors are assumed although the methods presented here can possibly be combined with the results of the Chapters 3 and 4 to obtain a more general procedure in fully nonparametric and heteroscedastic models (see Section 5.9.2). For the first time, local alternatives will be considered in the context of testing for a parametric transformation function. The presented test will be based on the ideas on estimating the transformation parameter of Colling and Van Keilegom (2018).
Keeping the goal of transforming a model in order to simplify relations and apply simple and fast procedures afterwards (as mentioned in the introduction) in mind, it might be sensible from a practical point of view to apply a parametric estimator even if the model does not hold exactly. With a good choice of the transformation parameter such a transformation, e.g., can reduce the dependence between covariates and errors enormously. Estimating an appropriate parameter is much easier than estimating the transformation h nonparametrically. Consequently, one might prefer the semiparametric transformation model with a parametric transformation function over a completely nonparametric one. It is then of interest to examine how far away the parametric class is from the true transformation function. To the author's knowledge this issue has not been discussed in the literature yet. Later, a test for the precise null hypothesis that a certain distance between the true transformation function and the parametric class exceeds a given threshold is presented. Then, rejecting the null yields evidence for applying the parametric transformation model. As a side effect, this procedure is accompanied with a quantification of the distance between the parametric class and the true transformation function making the results well interpretable.
This chapter is organized as follows. First, the test statistic is described in Section 5.1, before its asymptotic distribution is developed in Section 5.2. Interchanged (also called relevant or precise) hypotheses are considered in Section 5.3. A bootstrap approach is given in Section 5.4, while a simulation study is presented in Section 5.5. Finally, the results are summarized in Section 5.6. The assumptions and proofs are postponed to Sections 5.7 and 5.8, respectively.

5.1 Model and Test Statistic

Consider independent and identically distributed random variables $(Y_i, X_i), i = 1, ..., n$, which fulfil equation (5.1) and define $S_i = g(X_i) + \varepsilon_i$. The main question of this chapter is if a given parametric class of transformation functions is considered, does the transformation function h belong to this class? Hence, a parameter space $\Theta \subseteq \mathbb{R}^{d_\Theta}$ with a corresponding transformation class $\tilde{\mathcal{H}}_\Theta = \{\Lambda_\theta : \theta \in \Theta\}$ is needed. Then, the null hypothesis could be for example $H_0 : h \in \tilde{\mathcal{H}}_\Theta$ and the alternative $H_1 : h \notin \tilde{\mathcal{H}}_\Theta$. Having introduced these notations the idea behind the test statistic is quite simple. An appropriate distance $d(\hat{h}, \Lambda_\theta)$ between

an appropriate nonparametric estimator $\hat{h}$ for h and each function of the transformation class will be considered and the minimum will be used as the test statistic. Finally, the null hypothesis will be rejected if this distance is too large.
Although the presented idea is rather intuitive, a lot of questions arise when thinking about applying the test statistic or a corresponding test. So far, it is not clear how to identify or estimate the transformation function h and Λ_θ or which distance should be used. In principle, any nonparametric estimator of h with an asymptotic representation as in (5.13) below is applicable. Later, the pretransformed estimator of Colling and Van Keilegom (2019), which was described in Section 1.4, will be used. A sketch of how this approach could be extended to the heteroscedastic models by using the estimator from Chapter 4 is given in Section 5.9.2.
One has to be careful before applying any distance to $\hat{h}$ and Λ_θ since both functions are probably not comparable. As mentioned before in this thesis, transformation models are in general not identified unless some identification constraints are required. Therefore, one has to ensure that the identification conditions on which the nonparametric estimator is based meet the identification conditions of the parametric class, since both estimators otherwise may deviate from each other only because they are based on different identification constraints and h is thus a linear transform of some Λ_{θ_0} (which means that basically both functions fulfil the model equation (5.1)). Throughout this chapter, the identification conditions

$$h(0) = 0 \quad \text{and} \quad h(1) = 1, \tag{5.2}$$

which meet the conditions in Section 1.4, are used for the nonparametric transformation function and its estimator. Note that for example the Yeo-Johnson-transforms (Yeo and Johnson (2000))

$$\Lambda_\theta(Y) = \begin{cases} \frac{(Y+1)^\theta - 1}{\theta}, & \text{if } Y \geq 0, \theta \neq 0, \\ \log(Y+1), & \text{if } Y \geq 0, \theta = 0, \\ -\frac{(1-Y)^{2-\theta} - 1}{2-\theta}, & \text{if } Y < 0, \theta \neq 2, \\ -\log(1-Y), & \text{if } Y < 0, \theta = 2 \end{cases} \tag{5.3}$$

fulfil $\Lambda_\theta(0) = 0$ for all $\theta \in \mathbb{R}$, but in general not $\Lambda_\theta(1) = 1$. Therefore, when assuming (5.2) the null hypothesis of h belonging to the Yeo-Johnson-transforms has to be reformulated as $H_0 : h \in \left\{\frac{\Lambda_\theta}{\Lambda_\theta(1)} : \theta \in \mathbb{R}\right\}$. For general parameter spaces Θ and corresponding transformation classes, this can be extended to

$$H_0 : h \in \left\{\frac{\Lambda_\theta - \Lambda_\theta(0)}{\Lambda_\theta(1) - \Lambda_\theta(0)} : \theta \in \Theta\right\} =: \mathcal{H}_\Theta. \tag{5.4}$$

When comparing a nonparametric estimator of h and a parametric member of $\mathcal{H}_\Theta$ this can be done by computing the difference on the one hand between $y \mapsto h(y)$ and $y \mapsto \frac{\Lambda_\theta(y) - \Lambda_\theta(0)}{\Lambda_\theta(1) - \Lambda_\theta(0)}$ itself or on the other hand between $y \mapsto h(y)(\Lambda_\theta(1) - \Lambda_\theta(0)) + \Lambda_\theta(0)$ and $y \mapsto \Lambda_\theta(y)$. Although in principle both of these approaches are conceivable, only the second one will be considered in the following, since this approach should be exemplary for the other and was already used by Colling and Van Keilegom (2018) to develop an estimator of θ under the null hypothesis. So far, the question of how to choose a distance appropriately has not been answered. Let w be a weight function with compact support. Let f and g be real valued and monotone functions with $f(0) = 0$ and $f(1) = 1$. Colling and Van Keilegom (2018) considered the weighted $\mathcal{L}^2$-distance

$$\tilde{d}(f, g) = E[w(Y)(f(Y)(g(1) - g(0)) + g(0) - g(Y))^2] \tag{5.5}$$

and the modified version

$$d(f, g) = \min_{c_1 \in \mathbb{R}^+, c_2 \in \mathbb{R}} E[w(Y)(f(Y)c_1 + c_2 - g(Y))^2]. \tag{5.6}$$

Note that both distances are zero if and only if g is almost surely a linear transform of f, which going back to transformation functions would correspond to $d(h, \Lambda_\theta) = 0$ for some $\theta \in \Theta$ and thus to the null hypothesis. Colling and Van Keilegom (2018) found that estimators of θ that are based on d seem to outperform those based on $\tilde{d}$. Hence, only d will be considered in the following, but again it should be possible to treat a test statistic based on $\tilde{d}$ similarly to the one that will be proposed here. Because the expectation in general has to be estimated as well, d will be replaced by its empirical counterpart

$$d_n(\hat{h}, \Lambda_\theta) := \min_{c_1 \in \mathbb{R}^+, c_2 \in \mathbb{R}} \frac{1}{n} \sum_{i=1}^{n} w(Y_i)(\hat{h}(Y_i)c_1 + c_2 - \Lambda_\theta(Y_i))^2, \tag{5.7}$$

where $\hat{h}$ denotes some nonparametric estimator of h with an asymptotic representation as in (5.13) below, e.g. the estimator of Colling and Van Keilegom (2019). Now, a test statistic can be defined as

$$T_n := \min_{\theta \in \Theta} n d_n(\hat{h}, \Lambda_\theta), \tag{5.8}$$

where the prefactor results from the asymptotic behaviour derived in Section 5.2. An asymptotic level α test for (5.4) may consequently look like

$$\Phi(Y_1, X_1, ...Y_n, X_n) = I_{\{T_n > c_\alpha\}} \tag{5.9}$$

with some critical value c_α that is obtained from the asymptotic distribution in Theorem 5.2.2.

5.2 Asymptotic Behaviour of the Test Statistic

In this section, it is examined how the test statistic given in (5.8) behaves asymptotically. First, define

$$U_i = \mathcal{T}(Y_i) \quad \text{with} \quad \mathcal{T}(y) = \frac{F_Y(y) - F_Y(0)}{F_Y(1) - F_Y(0)} \quad \text{and} \quad Z_i = (U_i, X_i). \tag{5.10}$$

Assume F_Y to be strictly increasing on the support of Y and due to the identification constraints (5.2) assume w.l.o.g. that 0 and 1 belong to the support of Y (otherwise replace them by arbitrary values $a < b \in \mathbb{R}$ here and in (5.2), which belong to the support of Y), so that $\mathcal{T}$ is well defined and invertible. Using this definition, the model equation (5.1) can be written as

$$S := h(Y) = Q(\mathcal{T}(Y)) = g(X) + \varepsilon \tag{5.11}$$

with $Q(\cdot) = h(\mathcal{T}^{-1}(\cdot))$.

Remark 5.2.1 *Due to the identification constraints (5.2), U_i can be expressed alternatively as*

$$U_i = \frac{F_S(h(Y_i)) - F_S(h(0))}{F_S(h(1)) - F_S(h(0))} = \frac{F_S(S_i) - F_S(0)}{F_S(1) - F_S(0)}. \tag{5.12}$$

It was already mentioned in Section 5.1, that an asymptotic representation for $\hat{h}(y) - h(y)$ is needed. To be precise, assume

$$\hat{h}(y) - h(y) = \frac{1}{n}\sum_{i=1}^{n} \psi(Z_i, \mathcal{T}(y)) + o_p\Big(\frac{1}{\sqrt{n}}\Big) \tag{5.13}$$

uniformly on compact sets where ψ has to fulfil assumption (**A7**) in Section 5.7. An example for such an estimator is the nonparametric estimator of Colling and Van Keilegom (2019), which was described in Section 1.4.
As already mentioned, the null hypothesis (5.4) can be rewritten as

$$H_0 : h(\cdot)c_1 + c_2 = \Lambda_{\theta_0}(\cdot) \quad \text{for some } \theta_0 \in \Theta, c_1 \in \mathbb{R}^+, c_2 \in \mathbb{R}.$$

Keeping X_i and ε_i (and thus S_i and Z_i as well) fixed, local alternatives can for example be formulated as

$$H_{1,n} : h(\cdot)c_1 + c_2 = \Lambda_{\theta_0}(\cdot) + n^{-\frac{1}{2}} r(\cdot) \quad \text{for some } \theta_0 \in \Theta, c_1 \in \mathbb{R}^+, c_2 \in \mathbb{R}, \tag{5.14}$$

where r is some continuous function and θ_0 is assumed to be fixed. This leads to $Y_i = h^{-1}(S_i)$. Applying (5.2) again leads to $c_2 = \Lambda_{\theta_0}(0) + n^{-\frac{1}{2}} r(0)$ and $c_1 = \Lambda_{\theta_0}(1) - \Lambda_{\theta_0}(0) + n^{-\frac{1}{2}}(r(1) - r(0))$, so that

$$h(\cdot) = \frac{\Lambda_{\theta_0}(\cdot) - \Lambda_{\theta_0}(0) + n^{-\frac{1}{2}}(r(\cdot) - r(0))}{\Lambda_{\theta_0}(1) - \Lambda_{\theta_0}(0) + n^{-\frac{1}{2}}(r(1) - r(0))} = h_0(\cdot) + n^{-\frac{1}{2}} r_0(\cdot) + \mathcal{O}\Big(\frac{1}{n}\Big) \tag{5.15}$$

uniformly on compact sets, where

$$h_0(\cdot) = \frac{\Lambda_{\theta_0}(\cdot) - \Lambda_{\theta_0}(0)}{\Lambda_{\theta_0}(1) - \Lambda_{\theta_0}(0)} \quad \text{and} \quad r_0(\cdot) = \frac{r(\cdot) - r(0) - h_0(\cdot)(r(1) - r(0))}{\Lambda_{\theta_0}(1) - \Lambda_{\theta_0}(0)}. \tag{5.16}$$

Note that, because $r \equiv 0$ is possible as well, the null hypothesis is included in this framework and that due to Remark 5.2.1 the distribution of U_i is independent of n. Hence, it is reasonable to assume that

$$h(Y_j) - \hat{h}(Y_j) = \frac{1}{n}\sum_{i=1}^{n} \psi(Z_i, U_j) + o_p\Big(\frac{1}{\sqrt{n}}\Big)$$

holds under local alternatives as well. For the estimator of Colling and Van Keilegom (2019), this will be shown in Lemma 5.8.1.
Before any result can be stated some notations have to be introduced. First of all, denote $\gamma = (c_1, c_2, \theta^t)^t \in \Upsilon := C_1 \times C_2 \times \Theta$, where Υ is assumed to be compact (see (**A1**) in Section 5.7), and

$$\tilde{\gamma} = \arg\min_{\gamma \in \Upsilon} \sum_{k=1}^{n} w(Y_k)(\hat{h}(Y_k)c_1 + c_2 - \Lambda_\theta(Y_k))^2. \tag{5.17}$$

For all minimizers $\tilde{\gamma}$, the test statistic can be written as

$$T_n = \min_{\gamma \in \Upsilon} \sum_{i=1}^{n} w(Y_i)(\hat{h}(Y_i)c_1 + c_2 - \Lambda_\theta(Y_i))^2 = \sum_{i=1}^{n} w(Y_i)(\hat{h}(Y_i)\tilde{c}_1 + \tilde{c}_2 - \Lambda_{\tilde{\theta}}(Y_i))^2. \quad (5.18)$$

Moreover, mark the derivatives with respect to θ with a "$\cdot$" and define

$$R(s) = (s, 1, -\dot{\Lambda}_{\theta_0}(h_0^{-1}(s)))^t, \quad (5.19)$$

$$\Gamma = E[w(h_0^{-1}(S_1))R(S_1)R(S_1)^t], \quad (5.20)$$

$$\varphi(z) = E[w(h_0^{-1}(S_2))\psi(Z_1, U_2)R(S_2) \mid Z_1 = z] \quad (5.21)$$

as well as

$$\zeta(z_1, z_2) = E\Big[w(h_0^{-1}(S_3))\big(\psi(Z_1, U_3) - \varphi(Z_1)^t\Gamma^{-1}R(S_3)\big)$$
$$\big(\psi(Z_2, U_3) - \varphi(Z_2)^t\Gamma^{-1}R(S_3)\big) \mid Z_1 = z_1, Z_2 = z_2\Big], \quad (5.22)$$

$$\bar{r}(s) = r_0(h_0^{-1}(s)) - E[w(h_0^{-1}(S_1))r_0(h_0^{-1}(S_1))R(S_1)]^t\Gamma^{-1}R(s), \quad (5.23)$$

$$\tilde{\zeta}(z) = 2E[w(h_0^{-1}(S_2))\psi(Z_1, U_2)\bar{r}(S_2) \mid Z_1 = z]. \quad (5.24)$$

Here, Z_i is defined as in (5.10). Further, let P^Z and F^Z denote the distribution law and distribution function of Z_i, respectively. Under some assumptions stated later in Section 5.7, it can be shown that the minimizer $\tilde{\gamma}$ is uniquely determined with probability converging to one.

Theorem 5.2.2 *Assume **(A1)**–**(A7)** given in Section 5.7. Let $(\lambda_k)_{k\in\{1,2,\dots\}}$ be the eigenvalues of the operator*

$$K\rho(z_1) := \int \rho(z_2)\zeta(z_1, z_2)\, dF_Z(z_2)$$

with corresponding eigenfunctions $(\rho_k)_{k\in\{1,2,\dots\}}$, which are orthonormal in the L^2-space corresponding to the distribution law P^Z. Let $(V_k)_{k\in\{1,2,\dots\}}$ be independent and standard normally distributed random variables and let V_0 be centred normally distributed with variance $E[\tilde{\zeta}(Z_1)^2]$ such that for all $K \in \mathbb{N}$ the random vector $(V_0, V_1, \dots, V_K)^t$ follows a multivariate normal distribution with $\mathrm{Cov}(V_0, V_k) = E[\tilde{\zeta}(Z_1)\rho_k(Z_1)]$ for all $k = 1, \dots, K$. Then, under the local alternative $H_{1,n}$ in (5.14), T_n converges in distribution to

$$(\Lambda_{\theta_0}(1) - \Lambda_{\theta_0}(0))^2\left(\sum_{k=1}^{\infty}\lambda_k V_k^2 + V_0 + E\left[w(h_0^{-1}(S_1))\bar{r}(S_1)^2\right]\right). \quad (5.25)$$

In particular, under H_0 (i.e. for $r \equiv 0$), T_n converges in distribution to

$$T = (\Lambda_{\theta_0}(1) - \Lambda_{\theta_0}(0))^2 \sum_{k=1}^{\infty}\lambda_k V_k^2. \quad (5.26)$$

The proof can be found in Section 5.8.1.

Remark 5.2.3 *Note that* $\zeta(z_1, z_2) = E[I(z_1)I(z_2)]$ *with*

$$I(z) := w(h_0^{-1}(S_1))^{1/2} \left(\psi(z, U_1) - \varphi(z)^t \Gamma^{-1} R(S_1)\right).$$

Thus, the operator K *defined in Theorem 5.2.2 is positive semi-definite, because the inner product* $\langle \cdot, \cdot \rangle$ *of the Hilbert space* $\mathcal{L}^2(\mathbb{R}^{d_X+1}, P^Z)$ *can be used to write*

$$\begin{aligned}\langle K\rho, \rho \rangle &= \int \int \rho(z_1)\zeta(z_1, z_2)\, dF_Z(z_1)\, \rho(z_2)\, dF_Z(z_2) \\ &= E\left[\left(\int \rho(z_1)I(z_1)\, dF_Z(z_1)\right)^2\right] \;\geq\; 0.\end{aligned}$$

Remark 5.2.4 *When considering transformation classes of finitely many (by d distinguishable) transformation functions the results of Theorem 5.2.2 can be modified as in Section 5.9.1, since the estimation (here classification) of the transformation parameter does no longer influence the asymptotic distribution of* T_n.

Remark 5.2.5 *Note that the minimization is also affected by the local alternative. Often in regression models, local alternatives are assumed to fulfil some orthogonality condition, since then the minimizing parameter does not depend on* n*, see for example Härdle and Mammen (1993). As was pointed out in Section 2.3.2, an orthogonality condition is needed to ensure that the test from Chapter 2 detects the local alternatives. If the minimizing parameter in the context of local alternatives with respect to the transformation function is assumed to be independent of* n*, this would mean that*

$$\tilde{\gamma}_0 = \begin{pmatrix} \tilde{c}_{1,0} \\ \tilde{c}_{2,0} \\ \tilde{\theta}_0 \end{pmatrix} := \underset{c_1, c_2, \theta}{\arg\min}\, E[w(Y)(h(Y)c_1 + c_2 - \Lambda_\theta(Y))^2]$$

is independent of n*. Since* h *converges to* h_0*, one should obtain* $\tilde{\gamma}_0 = \gamma_0$ *in this case with* γ_0 *such that*

$$h_0(y)c_{1,0} + c_{2,0} = \Lambda_{\theta_0}(y).$$

Consequently, this would lead to

$$\begin{aligned} 0 &= \left(D_\gamma E[w(Y)(h(Y)c_1 + c_2 - \Lambda_\theta(Y))^2]\big|_{\gamma = \tilde{\gamma}_0}\right)^t \\ &= 2E\left[w(Y)\left(h_0(Y)c_{1,0} + \frac{1}{\sqrt{n}} r(Y)c_{1,0} + c_{2,0} - \Lambda_{\theta_0}(Y)\right)\begin{pmatrix} h(Y) \\ 1 \\ \dot{\Lambda}_{\theta_0}(Y)^t \end{pmatrix}\right].\end{aligned}$$

Thus,

$$E[w(Y)r(Y)] = 0, \quad E[w(Y)r(Y)\dot{\Lambda}_{\theta_0}(Y)^t] = 0 \quad \text{and} \quad E[w(Y)r(Y)h(Y)] = 0.$$

Because not only the distribution of Y*, but also* $h = h_0 + \frac{1}{\sqrt{n}}r$ *itself depends on* n*, these (and especially the last) restrictions might be too strong, so that this assumption would be no longer sensible. Consequently, the minimizing parameter in general depends on* n.

Theoretically, the critical values c_α mentioned in (5.9) can now be obtained using the $(1-\alpha)$-quantile of the asymptotic distribution in (5.26). In Section 5.4, a bootstrap procedure is suggested to estimate these quantiles.

Theorem 5.2.6 *Assume (**A1**)–(**A3**),(**A4'**),(**A5'**) and let $\hat{h}$ estimate h uniformly consistently on compact sets. Consider the fixed alternative*

$$H_1: \quad d(h, \Lambda_\theta) > 0 \quad \text{for all } \theta \in \Theta \tag{5.27}$$

and denote the $(1-\alpha)$-quantiles of T (as in (5.26)) by c_α. Then, for the test provided in (5.9) and all $\alpha \in (0,1)$

$$P(\Phi(Z_1, ..., Z_n) = 1) = P(T_n > c_\alpha) \overset{n\to\infty}{\longrightarrow} 1,$$

that is, the proposed test is consistent.

The proof can be found in Section 5.8.2.

5.3 Testing Precise Hypotheses

So far, the null hypothesis of a parametric transformation model was considered and the guidance of the proposed test would be to use a parametric estimator unless H_0 was not rejected. Nevertheless, the transformation model with a parametric transformation class might be useful in applications even if the model does not hold exactly. Since parametric in general outperform nonparametric estimators (if the model holds), it could still make sense in some cases to prefer a parametric estimator to a nonparametric one. This gives rise to the question if a transformation function is too far away from a given parametric class or if the class is still acceptable. Hence, the interchanged hypotheses

$$H_0': \min_{\theta\in\Theta} d(h, \Lambda_\theta) \geq \eta \quad \text{and} \quad H_1': \min_{\theta\in\Theta} d(h, \Lambda_\theta) < \eta \tag{5.28}$$

are considered in this section. Here, Θ and Λ_θ are as in (5.4) and η is some threshold fixed beforehand by the experimenter. Referring to Berger and Delampady (1987), the hypotheses (5.28) are called "precise hypotheses" in the following. If a suitable test rejects H_0' for some small η the model is "good enough" to work with, even if it does not hold exactly. Dette, Kokot, and Volgushev (2018) considered precise hypotheses in the context of comparing mean functions in the context of functional time series. Note that the idea of precise hypotheses is related to that of equivalence tests, which originate from the field of pharmacokinetics (see Lakens (2017)).
Fortunately, the same test statistic as before can be used although one has to standardize differently. Assume H_0' holds, that is, let h be a (fixed) transformation which does **not** belong to some given parametric transformation class. Further, let

$$M(\gamma) = M(c_1, c_2, \theta^t)^t = E[w(Y)(h(Y)c_1 + c_2 - \Lambda_\theta(Y))^2], \tag{5.29}$$

let

$$\gamma_0 = (c_{1,0}, c_{2,0}, \theta_0) := \underset{(c_1,c_2,\theta^t)^t\in\Upsilon}{\arg\min} \; M(c_1, c_2, \theta^t)^t$$

be unique and let $\tilde{\gamma}$ be defined as in (5.17). Note that $\min_{c_1 \in C_1, c_2 \in C_2} M(\gamma) = d(h, \Lambda_\theta)$ for all $\theta \in \Theta$. Assume that

$$\Gamma' = E\left[w(Y_1)\begin{pmatrix} h(Y_1)^2 & h(Y_1) & -h(Y_1)\dot{\Lambda}_{\theta_0}(Y_1) \\ h(Y_1) & 1 & -\dot{\Lambda}_{\theta_0}(Y_1) \\ -h(Y_1)\dot{\Lambda}_{\theta_0}(Y_1)^t & -\dot{\Lambda}_{\theta_0}(Y_1)^t & \Gamma'_{3,3} \end{pmatrix}\right] \tag{5.30}$$

is positive definite, where $\Gamma'_{3,3} = \dot{\Lambda}_{\theta_0}(Y)^t \dot{\Lambda}_{\theta_0}(Y_1) - \ddot{\Lambda}_{\theta_0}(Y_1)\tilde{R}_1$ and $\tilde{R}_i$ is defined as

$$\tilde{R}_i = h(Y_i)c_{1,0} + c_{2,0} - \Lambda_{\theta_0}(Y_i), \quad i = 1, ..., n. \tag{5.31}$$

Theorem 5.3.1 *Assume **(A1)**–**(A3)**,**(A4')**,**(A5')**,**(A7')** in Section 5.7 and let* Γ' *be positive definite. Then*

$$n^{1/2}(T_n/n - M(\gamma_0)) \xrightarrow{\mathcal{D}} \mathcal{N}\big(0, \sigma^2\big), \tag{5.32}$$

where $\sigma^2 = \operatorname{Var}\big(w(Y_1)\tilde{R}_1^2 + \delta(Z_1)\big)$ *and* $\delta(z) = 2E[w(Y_2)\psi(Z_1, U_2)\tilde{R}_2 \mid Z_1 = z]$.

The proof can be found in Section 5.8.3. Since in this scenario it is not clear how to obtain appropriate quantiles via some bootstrap algorithm, it would be desirable to simply plug a consistent estimator of σ^2 into (5.32) or to be precise standardise the left hand side to obtain a standard normal distribution. Then, a consistent asymptotic level-α-test would reject H_0' if $(T_n - n\eta)/(n\hat{\sigma}^2)^{1/2} < u_\alpha$, where u_α is the α-quantile of the standard normal distribution. For this purpose, let $(m)_{n\in\mathbb{N}} = (m_n)_{n\in\mathbb{N}}$ be an intermediate sequence, that is

$$m_n \to \infty, \quad \frac{n}{m_n} \to \infty,$$

and define $q := \lceil \frac{n}{m_n} \rceil - 1$. Moreover, let $\mathcal{Y}_w$ be the compact support of w and denote for some $\nu \in \{1, ..., q\}$ the nonparametric estimator of h depending on $Z_{(\nu-1)m+1}, ..., Z_{\nu r}$ by $\hat{h}^{(\nu)}$, that is,

$$\hat{h}^{(\nu)}(y) - h(y) = \frac{1}{m}\sum_{j=(\nu-1)m+1}^{\nu r} \psi(Z_j, \mathcal{T}(y)) + o_p\left(\frac{1}{\sqrt{m}}\right)$$

holds uniformly on $\mathcal{Y}_w$.

Lemma 5.3.2 *Assume **(A1)**–**(A3)**,**(A4')**,**(A5')**,**(A7')** in Section 5.7 and let* Γ' *be positive definite. Define*

$$H_{m,\nu} := \sup_{y\in\mathcal{Y}_w}\left|\sqrt{m}(\hat{h}^{(\nu)}(y) - h(y)) - \frac{1}{\sqrt{m}}\sum_{j=(\nu-1)m+1}^{\nu m}\psi(Z_j, \mathcal{T}(y))\right| = o_p(1) \tag{5.33}$$

for $m \to \infty$. *Then, there exists an intermediate sequence* $(m_n)_{n\in\mathbb{N}}$ *and a null sequence* $(a_n)_{n\in\mathbb{N}}$ *such that*

$$\max_{\nu=1,\dots,q} H_{m,\nu} = \mathcal{O}_p(a_n) \tag{5.34}$$

and

$$P(|H_{m,1}| > \varepsilon) = o\left(\frac{1}{q}\right) \tag{5.35}$$

holds for all $\varepsilon > 0$. Then, a consistent estimator of the asymptotic variance σ^2 is given by

$$\hat{\sigma}^2 := \frac{1}{q}\sum_{\nu=1}^{q}\Bigg(\frac{2\sqrt{m_n}}{n}\sum_{k=1}^{n} w(Y_k)\big(\hat{h}^{(\nu)}(Y_k)-\hat{h}(Y_k)\big)\big(\hat{h}(Y_k)\tilde{c}_1+\tilde{c}_2-\Lambda_{\tilde{\theta}}(Y_k)\big)$$
$$+\frac{1}{\sqrt{m_n}}\sum_{j=(\nu-1)m_n+1}^{\nu m_n}\Bigg(w(Y_j)(\hat{h}(Y_j)\tilde{c}_1+\tilde{c}_2-\Lambda_{\tilde{\theta}}(Y_j))^2$$
$$-\frac{1}{n}\sum_{i=1}^{n} w(Y_i)(\hat{h}(Y_i)\tilde{c}_1+\tilde{c}_2-\Lambda_{\tilde{\theta}}(Y_i))^2\Bigg)\Bigg)^2$$

with $(\tilde{c}_1,\tilde{c}_2,\tilde{\theta}^t)=\tilde{\gamma}^t$ from equation (5.17). One has $\hat{\sigma}^2-\sigma^2=\mathcal{O}_p\big(\sqrt{\frac{m_n}{n}}+a_n+n^{-\frac{1}{4}}\big)$. The optimal, that is fastest, rate of convergence is $\mathcal{O}_p(n^{-\frac{1}{4}})$, which is obtained if (5.34) and (5.35) hold for $m_n=\mathcal{O}_p(\sqrt{n})$ and $a_n=\mathcal{O}_p(n^{-\frac{1}{4}})$.

The proof can be found in Section 5.8.4.

5.4 Bootstrap

Although the results presented in Theorem 5.2.2 give an idea of how the test statistic behaves asymptotically, it is hard to extract any information about how to choose the critical values c_α in (5.9) from the limiting distribution in equation (5.26) appropriately. The main reasons for this are that first for any function ζ the eigenvalues of the operator mentioned in (5.2.2) are unknown, that second this function is unknown and needs to be estimated as well and that third even ψ (which would be needed to estimate ζ) is unknown and rather complex (see Section 1.4). Therefore, calculating the quantiles of the distribution in (5.26) directly might not be a good idea and some bootstrap technique should be applied instead. In this section, a bootstrap algorithm is developed and consistency of the procedure is proven.

A bootstrap algorithm is required to fulfil two properties: First, under the null hypothesis the algorithm has to provide, conditioned on the original data $(Y_1,X_1),...,(Y_n,X_n)$, consistent estimates of the quantiles of T_n or more precisely its asymptotic distribution (5.26). To specify, what is meant by this, let $(\Omega,\mathcal{A},P)$ denote the underlying probability space and assume that $(\Omega,\mathcal{A})$ can be written as $\Omega=\Omega_1\times\Omega_2$ and $\mathcal{A}=\mathcal{A}_1\otimes\mathcal{A}_2$ for some measurable spaces $(\Omega_1,\mathcal{A}_1)$ and $(\Omega_2,\mathcal{A}_2)$. Further, assume that P is characterized as the product of a probability measure P_1 on $(\Omega_1,\mathcal{A}_1)$ and a Markov kernel

$$P_2^1:\Omega_1\times\mathcal{A}_2\to[0,1],$$

that is $P=P_1\otimes P_2^1$. The expectation with respect to $P_2^1(\omega,\cdot)$ is written as $E[\cdot|\omega]$. In the following, bootstrap data $(Y_1^*,X_1^*),...,(Y_m^*,X_m^*)$ is generated. While randomness with respect to the original data is modelled by P_1, randomness with respect to the bootstrap data and conditional on the original data is modelled by P_2^1. Moreover, assume

$$P_2^1(\omega,A)=P\big(\Omega_1\times A|(Y_i,X_i)=(Y_i(\omega),X_i(\omega))\ \forall i=1,...,n\big)\text{ for all }\omega\in\Omega_1,A\in\mathcal{A}_2.$$

With these notations in mind, if the bootstrap statistic is denoted by $T^*_{n,m}$, where m is the sample size of the bootstrap data, then for all $q \in (0, \infty)$ it would be desirable to obtain

$$P_1\Big(\omega \in \Omega_1 : \limsup_{m\to\infty} \big|P_2^1(\omega, \{T^*_{n,m} \le q\}) - P(T_n \le q)\big| > \delta\Big) = o(1) \tag{5.36}$$

for all $\delta > 0$ and $n \to \infty$. Here, the notation

$$P_2^1(\omega, \{T^*_{n,m} \le q\}) = P_2^1\big(\omega, \{\tilde{\omega} \in \Omega_2 : (\omega, \tilde{\omega}) \in \{T^*_{n,m} \le q\}\}\big)$$

is used. On the other hand, to be consistent under H_1 the bootstrap quantiles have to stabilize or at least converge to infinity with a rate less than that of T_n. To be precise, it is needed that

$$P_1\Big(\omega \in \Omega_1 : \limsup_{m\to\infty} P_2^1(\omega, \{T_n \le T^*_{n,m}\}) > \delta\Big) = o(1) \tag{5.37}$$

for all $\delta > 0$ and $n \to \infty$. The main problem of estimating the quantiles in (5.26) consists in mimicking the asymptotic behaviour of $\hat{h}$. In the following, an algorithm is given which ensures both properties (5.36) and (5.37).

Algorithm 5.4.1 *Let $(Y_1, X_1), ..., (Y_n, X_n)$ denote the observed data and for some $\alpha \in (0,1)$ let q_α denote the quantile that needs to be estimated. Further, let $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ be sequences with $a_n, b_n \searrow 0$. Moreover, define*

$$h_\theta(y) = \frac{\Lambda_\theta(y) - \Lambda_\theta(0)}{\Lambda_\theta(1) - \Lambda_\theta(0)} \quad \text{and} \quad g_\theta(x) = E[h_\theta(Y)|X = x].$$

(1) *Calculate $\tilde{\gamma} = (\tilde{c}_1, \tilde{c}_2, \tilde{\theta})^t = \arg\min_{\gamma\in\Upsilon} \sum_{i=1}^n w(Y_i)(\hat{h}(Y_i)c_1 + c_2 - \Lambda_\theta(Y_i))^2$. Let θ_0 be defined as in **(A5)** (under the null hypothesis) or as in **(A5')** (under the alternative) in Section 5.7 and let $\hat{g}$ be an estimator of g_{θ_0}.*

(2) *Let $m \in \mathbb{N}$ and let W_i and $\xi_i, i = 1, ..., m$, be independent and absolute continuous random variables. Estimate the parametric residuals $\varepsilon(\theta) = h_\theta(Y) - g_\theta(X)$ by*

$$\hat{\varepsilon}_i(\tilde{\theta}) = h_{\tilde{\theta}}(Y_i) - \hat{g}(X_i),\ \ i = 1, ..., n.$$

Draw some indices $j_X^*, j_\varepsilon^*, j = 1, ..., m$, with replacement from $\{1, ..., n\}$ (independently) and define the bootstrap data by

$$X_j^* = X_{j_X^*} + b_n W_j$$

and

$$Y_j^* = (h^*)^{-1}\Big(\hat{g}(X_j^*) + \hat{\varepsilon}_{j_\varepsilon^*}(\tilde{\theta}) - \frac{1}{n}\sum_{l=1}^n \hat{\varepsilon}_l(\tilde{\theta}) + a_n\xi_j\Big) \quad \text{for} \quad h^*(\cdot) = \frac{\Lambda_{\tilde{\theta}}(\cdot) - \Lambda_{\tilde{\theta}}(0)}{\Lambda_{\tilde{\theta}}(1) - \Lambda_{\tilde{\theta}}(0)}. \tag{5.38}$$

(3) *Calculate the bootstrap estimate $\hat{h}^*$ for h^* from $(Y_j^*, X_j^*), j = 1, ..., m$.*

(4) *Calculate the bootstrap statistic $T^*_{n,m} = \min_{\gamma\in\Upsilon} \sum_{j=1}^m w(Y_j^*)(\hat{h}^*(Y_j^*)c_1 + c_2 - \Lambda_\theta(Y_j^*))^2$.*

(5) *Let $B \in \mathbb{N}$. Repeat steps (2)–(4) B times to obtain the bootstrap statistics $T^*_{n,m,1}, ...,$ $T^*_{n,m,B}$. Let q^*_α denote the quantile of $T^*_{n,m}$ conditional on $(Y_i, X_i), i = 1, ..., n$. Estimate q^*_α by*

$$\hat{q}^*_\alpha = \min \left\{ z \in \{T^*_{n,m,1}, ..., T^*_{n,m,B}\} : \frac{1}{B} \sum_{k=1}^{B} I_{\{T^*_{n,m,k} \leq z\}} \geq \alpha \right\}.$$

To apply the estimating approach developed by Colling and Van Keilegom (2019), further assumptions are needed to induce the validity of Algorithm 5.4.1.

Assumptions 5.4.2 *In the following, the notations from Algorithm 5.4.1 are employed. Let $\frac{1}{na_n}, \frac{\log(n)}{nb_n^{d_X+4}} \to 0$.*

(1) *Let $\hat{g}$ be $(q+2)$-times continuously differentiable (same q as in (**B4**) in Section 5.7).*

(2) *Denote the densities of W_i and $\xi_i, i = 1, ..., m$, by f_W and f_ξ, respectively. Assume*

- *f_W and f_ξ are $(q+2)$-times continuously differentiable,*
- *f_W has bounded support,*
- *$f_W(0) > 0$ and $f_\xi > 0$,*
- *either $\left|\frac{\partial}{\partial u} f_\xi(u)\right| \leq K$,*

$$|\tilde{f}(u)| < |u|^{-\nu} \quad \text{and} \quad \left|\frac{\partial}{\partial u}\tilde{f}(u)\right| \leq K|u|^{-\nu} \tag{5.39}$$

for all $|u| > L$, $\tilde{f} \in \left\{f_\xi, \frac{\partial}{\partial u} f_\xi\right\}$ or

$$\left|\left|\frac{\partial}{\partial x}\tilde{f}(x)\right|\right| \leq K \quad \text{and} \quad \left|\left|\frac{\partial}{\partial x}\tilde{f}(x)\right|\right| I_{\{||x||>L\}} \leq K||x||^{-\nu} I_{\{||x||>L\}}$$

*for some $\nu > 1, K, L \in (0, \infty)$ and all $\tilde{f} \in \left\{f_W, \frac{\partial}{\partial x_i} f_W, \frac{\partial^2}{\partial x_i^2} f_W\right\}$, where the same i was used as in (**B3**). From now on, the case $i = 1$ is considered w.l.o.g.*

Remark 5.4.3 *1. The properties $\frac{nb_n^{d_X}}{\log(n)} \to \infty$ and $f_W(0) > 0$ ensure that conditional on the original data $(Y_1, X_1), ..., (Y_n, X_n)$ the support of X_1^* contains the support of v (from assumption (**B7**) in Section 5.7) with probability converging to one. Thus, v can be used for calculating $\hat{h}^*$ as well.*

2. To proceed as in (5.38) it may be necessary to modify h^* so that $S_j^* := \hat{g}(X_j^*) + \hat{\varepsilon}_{j_\varepsilon^*}(\tilde{\theta}) - \frac{1}{n}\sum_{l=1}^n \hat{\varepsilon}_l(\tilde{\theta}) + a_n\xi_j$ belongs to the domain of $(h^*)^{-1}$ for all $j = 1, ..., m$. As long as these modifications do not have influence on $h^*(y)$ for $y \in \mathcal{Y}_w$, the influence on $\hat{h}^*$ and $T^*_{n,m}$ should be asymptotically negligible (which can be proven for the estimator of Colling and Van Keilegom (2019)).

3. Later, it is assumed that the map $y \mapsto \Lambda_\theta(y)$ is $(q+2)$-times continuously differentiable for every $\theta \in \tilde{\Theta}$, where $\tilde{\Theta}$ is an arbitrarily small neighbourhood of θ_0. This assumption can be relaxed as long as (uniformly in $\theta \in \tilde{\Theta}$) Λ_θ can be approximated by some $(q+2)$-times differentiable function $\tilde{\Lambda}_\theta$ such that

$$\sup_{y \in \mathcal{Y}_w} |\tilde{\Lambda}_\theta(y) - \Lambda_\theta(y)| = o_p(a_n^2).$$

For any realisation $\omega \in \Omega_1$ define

$$F_{Y^*}(y) = P_2^1(\omega, \{Y_1^* \leq y\}), \quad \mathcal{T}^*(y) = \frac{F_{Y^*}(y) - F_{Y^*}(0)}{F_{Y^*}(1) - F_{Y^*}(0)}, \quad U^* = \mathcal{T}^*(Y^*) \tag{5.40}$$

and

$$S^* = h^*(Y^*).$$

Together with assumptions 5.4.2, Algorithm 5.4.1 is constructed in a way such that for the estimator of Colling and Van Keilegom (2019) a bootstrap counterpart to (5.13) is valid, namely

$$P_1\Bigg(\omega \in \Omega_1 : \forall \delta > 0 : \limsup_{m\to\infty} P_2^1\Bigg(\omega, \Bigg\{\sup_{y\in\mathcal{K}}\Bigg|\hat{h}^*(y) - h^*(y) - \frac{1}{m}\sum_{j=1}^{m}\psi^*(S_j^*, X_j^*, \mathcal{T}^*(y))\Bigg| > \frac{\delta}{\sqrt{m}}\Bigg\}\Bigg) = 0\Bigg)$$
$$= 1 + o(1) \tag{5.41}$$

for all compact sets $\mathcal{K} \subseteq \mathbb{R}$ and $n \to \infty$, where ψ^* fulfils assumption (**A7***). In the following, the conditional density of $\varepsilon(\theta_0)$ given X is written as $f_{\varepsilon(\theta_0)|X}$.

Lemma 5.4.4 *Let $\mathcal{K} \subset \mathbb{R}$ be compact. Assume (**A9**) in Section 5.7 and*

$$\frac{\max\limits_{i=1,\ldots,n} |\hat{g}(X_i) - g_{\theta_0}(X_i)|^r}{a_n^{r+1}} = o_p(1) \quad \text{and} \quad \frac{||\tilde{\theta} - \theta_0||^r}{a_n^{r+1}} = o_p(1). \tag{5.42}$$

Then,

$$\sup_{u\in\mathcal{K}}\Bigg|\frac{1}{n}\sum_{i=1}^{n} F_\xi\Bigg(\frac{u - \hat{\varepsilon}_i(\tilde{\theta})}{a_n}\Bigg) - \frac{1}{n}\sum_{i=1}^{n} F_\xi\Bigg(\frac{u - \varepsilon_i(\theta_0)}{a_n}\Bigg)\Bigg| = o_p(1)$$

and

$$\sup_{u\in\mathcal{K}}\Bigg|\frac{1}{na_n}\sum_{i=1}^{n} f_\xi\Bigg(\frac{u - \hat{\varepsilon}_i(\tilde{\theta})}{a_n}\Bigg) - \frac{1}{na_n}\sum_{i=1}^{n} f_\xi\Bigg(\frac{u - \varepsilon_i(\theta_0)}{a_n}\Bigg)\Bigg| = o_p(1).$$

The proof can be found in Section 5.8.5.

Remark 5.4.5 *(i) Later, it will be shown that $\frac{1}{n}\sum_{l=1}^{n} \hat{\varepsilon}_l(\tilde{\theta}) = o_p(1)$ in the proof of Lemma 5.8.2. Hence, the assertion of Lemma 5.4.4 can be extended to the centred residuals $\hat{\varepsilon}_i(\tilde{\theta}) - \frac{1}{n}\sum_{l=1}^{n} \hat{\varepsilon}_l(\tilde{\theta})$ by considering a slightly bigger set $\mathcal{K}$.*

*(ii) Many of the assumptions in (**A9**) can be replaced by less complex, but more restrictive versions. For example, due to (5.42) assumption (5.63) below is implied by*

$$E\Bigg[\sup_{||\theta-\theta_0||<\delta} ||\operatorname{Hess}\Lambda_\theta(Y_i)||^j\Bigg] < \infty$$

for all $j = 1, ..., r-1$ or

$$\max_{k=1,\ldots,n}\sup_{||\theta-\theta_0||<\delta} ||\operatorname{Hess}\Lambda_\theta(Y_i)|| = \mathcal{O}_p\big(||\tilde{\theta} - \theta_0||^{-1}\big)$$

$\big(= \mathcal{O}_p\big(n^{\frac{1}{4}}\big)$ under the alternative$\big)$.

(iii) One has $||\tilde{\theta} - \theta_0|| = \mathcal{O}_p\big(n^{-\frac{1}{4}}\big)$ under the alternative. Unfortunately, the experimenter in advance does not know, if the null hypothesis or the alternative holds, so that in general (5.42) limits a_n to $a_n^{-1} = o\big(n^{\frac{r}{4(r+1)}}\big) = o\big(n^{\frac{1}{4}}\big)$.

(iv) The regression function g_{θ_0} can be estimated using the Nadaraya-Watson approach as in the paper of Heuchenne, Samb, and Van Keilegom (2015). Under some additional assumptions, their Proposition 6.1 and its extension in the supplement of Colling and Van Keilegom (2016, p. 7) yield $\hat{g}(x) - g_{\theta_0}(x) = \mathcal{O}_p\Big(\sqrt{\frac{\log(n)}{nh_x^{d_X}}} + ||\tilde{\theta} - \theta_0||\Big)$ uniformly on compact sets. When assuming the existence of some compact set $\mathcal{K} \subseteq \operatorname{supp}(f_X)$ such that $\operatorname{supp}(v)$ is contained in the interior of $\mathcal{K}$, equation (5.42) (and a counterpart of (5.43) below for compact sets) can be obtained when discarding those $(Y_i, X_i), (Y_j^*, X_j^*)$, for which $X_i, X_j^* \notin \mathcal{K}$ holds. Note that an equation similar to (5.13) can still be derived, although in general with another ψ. Then, (5.42) requires $a_n^{-1} = o_p\Big(\big(\frac{nh_x^{d_X}}{\log(n)}\big)^{\frac{r}{2(r+1)}}\Big)$ for $||\tilde{\theta} - \theta_0|| = \mathcal{O}_p\big(n^{-\frac{1}{2}}\big)$ and $a_n^{-1} = o\Big(\big(\frac{nh_x^{d_X}}{\log(n)}\big)^{\frac{r}{2(r+1)}} + n^{\frac{r}{4(r+1)}}\Big)$ for $||\tilde{\theta} - \theta_0|| = \mathcal{O}_p\big(n^{-\frac{1}{4}}\big)$.

Lemma 5.4.6 *Let assumptions 5.4.2 and (5.42) be fulfilled. Further, assume **(A1)**–**(A7)**, **(A9)**,**(B1)**–**(B10)** in Section 5.7 and*

$$\sup_{x \in \mathbb{R}^{d_X}} |\hat{g}(x) - g_{\theta_0}(x)| = o_p(1) \quad \text{and} \quad \sup_{x \in \mathbb{R}^{d_X}} \left| \frac{\partial}{\partial x_1} \hat{g}(x) - \frac{\partial}{\partial x_1} g_{\theta_0}(x) \right| = o_p(1), \tag{5.43}$$

Moreover, assume the existence of a neighbourhood $\tilde{\Theta}$ of θ_0 such that the map $y \mapsto \Lambda_\theta(y)$ is $(q+2)$-times continuously differentiable for all $\theta \in \tilde{\Theta}$. Let $\hat{h}^*$ be the estimator from (5.46) based on the bootstrap data $(Y_j^*, X_j^*), j = 1, ..., m$. Assume that the density of $\varepsilon(\theta_0)$ (denoted by $f_{\varepsilon(\theta_0)}$) is continuous and

$$\sup_{e \in \mathbb{R}} |e f_{\varepsilon(\theta_0)}(e)| < \infty. \tag{5.44}$$

*Then, assumptions **(A7*)** and **(A8*)** are fulfilled and especially equation (5.41) is valid.*

The proof can be found in Section 5.8.6. The assertion can be extended to other $i \in \{1, ..., d_X\}$ as long as assumption (**B3**) is fulfilled.

Theorem 5.4.7 *Assume H_0,**(A1)**–**(A7)**,**(A7*)**,**(A8*)** in Section 5.7. Then, the bootstrap statistic $T_{n,m}^*$ computed by algorithm 5.4.1 fulfils (5.36). If q_α^* denotes for all $\alpha \in (0,1)$ the corresponding bootstrap quantile described in 5.4.1, one has*

$$P_1\Big(\omega \in \Omega_1 : \limsup_{m \to \infty} |q_\alpha^* - q_\alpha| > \delta\Big) = o(1)$$

for all $\delta > 0$ and $n \to \infty$.

The proof can be found in Section 5.8.7. Under the null hypothesis the proof of Lemma 5.4.6 is based on the convergence of ψ^* to ψ in probability and uniformly in its input argument, which will be specified later. If the alternative holds, it is not even clear if ψ^* stabilizes in some sense (see assumption (5.57)). Hence, additional assumptions are needed. For that purpose, define

$$F_{\varepsilon(\theta)}(e) = P(\varepsilon(\theta) \leq e),$$

$$F_S^B(u) = \int F_{\varepsilon(\theta_0)}(u - g_{\theta_0}(x)) f_X(x)\, dx,$$

$$\mathcal{T}_S^B(u) = \frac{F_S^B(u) - F_S^B(0)}{F_S^B(1) - F_S^B(0)},$$

$$\tilde{\Phi}(u|x) = F_{\varepsilon(\theta_0)}\big((\mathcal{T}_S^B)^{-1}(u) - g_{\theta_0}(x)\big).$$

While doing so, assume $F_S^B(0) < F_S^B(1)$ to ensure that $\mathcal{T}_S^B$ is well defined, and define

$$(\mathcal{T}_S^B)^{-1}(u) = \begin{Bmatrix} -\infty \\ \infty \end{Bmatrix}, \quad \text{if} \quad T_S^B(y) \begin{Bmatrix} > \\ < \end{Bmatrix} u \quad \text{for all } y \in \mathbb{R}.$$

Note that $\tilde{\Phi}$ plays a similar role under the alternative as $\Phi(u|x) = F_{U|X}(u|x)$ under the null hypothesis and thus is assumed to be continuously differentiable on $\mathcal{U}_0 \times \operatorname{supp}(v)$ with

$$\inf_{(u,x)\in\mathcal{U}_0\times\operatorname{supp}(v)} \frac{\partial}{\partial u}\tilde{\Phi}(u|x) > 0 \quad \text{and} \quad \inf_{(u,x)\in\mathcal{U}_0\times\operatorname{supp}(v)} \frac{\partial}{\partial x_1}\tilde{\Phi}(u|x) > 0 \tag{5.45}$$

in the following. Again, the derivative with respect to any other component x_i, $i \in \{1, ..., d_X\}$, could have been used as well.

Lemma 5.4.8 *Let the assumptions 5.4.2 and (5.42) be fulfilled. Moreover, assume H_1, **(A1)**–**(A3)**, **(A4')**, **(A5')**, **(A6)**, **(A7')**, **(A9)**, **(B1)**–**(B10)** in Section 5.7 as well as (5.43)–(5.45). Further, assume the existence of a neighbourhood $\tilde{\Theta}$ of θ_0 such that the map $y \mapsto \Lambda_\theta(y)$ is $(q+2)$-times continuously differentiable for all $\theta \in \tilde{\Theta}$. Let $\hat{h}^*$ be the estimator from (5.49) below, which is based on the bootstrap data $(Y_j^*, X_j^*), j = 1, ..., m$. Then, assumption **(A7*)** is fulfilled.*

The proof can be found in Section 5.8.8.

Theorem 5.4.9 *Under H_1 assume **(A1)**–**(A3)**,**(A4')**,**(A5')**,**(A7*)** in Section 5.7. Then, the bootstrap statistic $T_{n,m}^*$ computed by algorithm 5.4.1 fulfils (5.37). If q_α^* denotes for all $\alpha \in (0,1)$ the corresponding bootstrap quantile described in Algorithm 5.4.1, one has*

$$P_1\Big(\omega \in \Omega_1 : T_n > \limsup_{m\to\infty} q_\alpha^*\Big) = 1 + o(1).$$

The proof can be found in Section 5.8.9.

Remark 5.4.10 *Whatever estimator of g is used, there are in general further assumptions (for example integrability conditions on Y) needed to ensure $\sup_{x\in\mathbb{R}^{d_X}} |\hat{g}(x) - g(x)| = o_p(1)$. The reader is referred to Heuchenne et al. (2015) or Neumeyer et al. (2016) for one Nadaraya-Watson and one local polynomial approach, respectively, in the context of transformation models.*

5.5 Simulations

In this section, the question of how the performance of the proposed test for finite sample sizes could be analysed is answered and some simulations are conducted to illustrate the behaviour of the test for finitely many observations. The testing procedure mainly consists of three parts:

- the nonparametric estimation of h,
- calculating the minimum and the test statistic T_n,
- estimating critical values (via bootstrap).

While the bootstrap algorithm has already been explained in the previous section, both of the following subsections deal with one of the first two issues. The last subsection contains the actual simulation, that means the rejection probabilities and some explanatory figures for some chosen examples under the null hypothesis and under several alternatives. Consider $n \in I\!N$ independent random variables $(Y_1, X_1), ..., (Y_n, X_n)$ identically distributed as some (Y, X). Throughout this section, the nonparametric, but homoscedastic transformation model (5.1) is assumed and the test from (5.9) is considered. Simulations and calculations are conducted with *R* (R Core Team (2017)).

5.5.1 Nonparametric Estimation of *h*

The nonparametric estimator $\hat{h}$ of Colling and Van Keilegom (2019) is applied, see Section 1.4 for some motivation and a detailed treatment of this estimator. Denote the conditional distribution function of U_1, given $X_1 = x$, with U_1 from (5.10) by $F_{U|X}(\cdot|x)$ and the empirical distribution function of $Y_1, ..., Y_n$ by $\hat{F}_Y$. Remember

$$\hat{h}(y) = \hat{Q}(\hat{\mathcal{T}}(y)), \tag{5.46}$$

where

$$\hat{\mathcal{T}}(y) := \frac{\hat{F}_Y(y) - \hat{F}_Y(0)}{\hat{F}_Y(1) - \hat{F}_Y(0)},$$

and $\hat{U}_i := \hat{\mathcal{T}}(Y_i)$ estimates U_i from (5.10) for all $i = 1, ..., n$. For some appropriate weight function v, let

$$\tilde{Q}(u) = \operatorname*{arg\,min}_{q \in \mathbb{R}} \int v(x) \left| \frac{\hat{s}_1(u, x)}{\hat{s}_1(1, x)} - q \right| dx \tag{5.47}$$

with

$$\hat{s}_1(u, x) := \int_0^u \frac{\frac{\partial \hat{F}_{U|X}(r|x)}{\partial r}}{\frac{\partial \hat{F}_{U|X}(r|x)}{\partial x_1}} \, dr \tag{5.48}$$

and

$$\hat{F}_{U|X}(u|x) := \frac{\sum_{i=1}^n K_{h_x}(X_i - x) \mathcal{K}_{h_u}(u - \hat{U}_i)}{\sum_{i=1}^n K_{h_x}(X_i - x)}.$$

Here, h_x and h_u are bandwidths and K is an appropriate kernel function (as in the assumptions (**B3**) and (**B4**)) and

$$K_h(u) := \frac{1}{h} K\left(\frac{u}{h}\right), \qquad \mathcal{K}_h(u) := \int_{-\infty}^u K_h(r) \, dr = \int_{-\infty}^{\frac{u}{h}} K(r) \, dr.$$

Moreover, let L and b be some appropriate distribution function and bandwidth, respectively, and $L_b(\cdot) = L\left(\frac{\cdot}{b}\right)$. To ensure the derived asymptotic properties, the estimator in (5.47) actually has to be replaced by a smoothed version

$$\hat{Q}(u) = \operatorname*{arg\,min}_{q \in \mathbb{R}} \int v(x) \left(\frac{\hat{s}_1(u, x)}{\hat{s}_1(1, x)} - q \right) \left(2L_b\left(\frac{\hat{s}_1(u, x)}{\hat{s}_1(1, x)} - q \right) - 1 \right) dx \tag{5.49}$$

(see Colling and Van Keilegom (2019)). Nevertheless, the median is used as in the paper of Colling and Van Keilegom (2019) for conducting simulations since on the one hand the expression in (5.49) converges to that in (5.47) for $b \to 0$ (which indeed can be written as the median) and on the other hand applying the estimator in (5.47) avoids the choice of another smoothing parameter. Additionally, it is anyway discretized with respect to the x-component later.

Remark 5.5.1 *1. In principle, the weighted L^1-distance in (5.47) could be replaced by an L^2-distance (or possibly even another distance) as well. Here, only the L^1-distance is used since according to Colling and Van Keilegom (2019) the resulting estimator seems to outperform its counterpart which is based on a square loss.*

*2. The derivative with respect to any other component x_i fulfilling assumption (**B3**) can be used as well (similar to Chiappori et al. (2015)). W.l.o.g. only the case $i = 1$ is considered here.*

As in the paper of Colling and Van Keilegom (2019) v is chosen to be the density of a uniform distribution in the simulations. Moreover, for some natural number N_x, the estimator is approximated by the empirical median of points $\hat{\lambda}(u, x_l), l = 1, ..., N_x$, where $x_1, ..., x_{N_x}$ form an equidistant grid between $\min\limits_{i=1,...,n} X_i$ and $\min\limits_{i=1,...,n} X_i$. After rejecting those values for which the integral (calculated by *integrate* in *R*) in (5.48) diverges or which are near to such values, the median of all remaining values is considered.

5.5.2 Calculating the Test Statistic

The three-dimensional minimization problem can be reduced to a one-dimensional one. For this purpose, define

$$f_\theta(c_1, c_2) := \sum_{k=1}^{n} w(Y_k)(\hat{h}(Y_k)c_1 + c_2 - \Lambda_\theta(Y_k))^2$$

and write

$$T_n = \min_{\theta \in \Theta} \min_{(c_1,c_2) \in C_1 \times C_2} f_\theta(c_1, c_2).$$

Taking the derivative of f_θ leads to

$$\begin{aligned} D_{(c_1,c_2)} f_\theta(c_1, c_2) &= 2 \sum_{k=1}^{n} w(Y_k)(\hat{h}(Y_k)c_1 + c_2 - \Lambda_\theta(Y_k)) \begin{pmatrix} \hat{h}(Y_k) & , & 1 \end{pmatrix} \\ &= 2 \begin{pmatrix} c_1 & , & c_2 \end{pmatrix} \sum_{k=1}^{n} w(Y_k) \begin{pmatrix} \hat{h}(Y_k) \\ 1 \end{pmatrix} \begin{pmatrix} \hat{h}(Y_k) & , & 1 \end{pmatrix} \\ &\quad - 2 \sum_{k=1}^{n} w(Y_k)\Lambda_\theta(Y_k) \begin{pmatrix} \hat{h}(Y_k) & , & 1 \end{pmatrix} \\ &\overset{!}{=} 0 \end{aligned}$$

and the corresponding solution

$$\begin{pmatrix} c_1(\theta) \\ c_2(\theta) \end{pmatrix} = R^{-1} \sum_{k=1}^{n} w(Y_k)\Lambda_\theta(Y_k) \begin{pmatrix} \hat{h}(Y_k) \\ 1 \end{pmatrix},$$

where

$$R := \sum_{k=1}^{n} w(Y_k) \begin{pmatrix} \hat{h}(Y_k)^2 & \hat{h}(Y_k) \\ \hat{h}(Y_k) & 1 \end{pmatrix}.$$

After inserting this solution, one has

$$\begin{aligned} T_n = \min_{\theta\in\Theta} &\begin{pmatrix} c_1(\theta) & , & c_2(\theta) \end{pmatrix} R \begin{pmatrix} c_1(\theta) \\ c_2(\theta) \end{pmatrix} \\ &- 2\begin{pmatrix} c_1(\theta) & , & c_2(\theta) \end{pmatrix} \sum_{k=1}^{n} w(Y_k)\Lambda_\theta(Y_k) \begin{pmatrix} \hat{h}(Y_k) \\ 1 \end{pmatrix} + \sum_{k=1}^{n} w(Y_k)\Lambda_\theta(Y_k)^2 \\ = \min_{\theta\in\Theta} &\sum_{k=1}^{n} w(Y_k)\Lambda_\theta(Y_k)^2 \\ &- \sum_{j=1}^{n}\sum_{k=1}^{n} w(Y_j)w(Y_k)\Lambda_\theta(Y_j)\Lambda_\theta(Y_k) \begin{pmatrix} \hat{h}(Y_j) & , & 1 \end{pmatrix} R^{-1} \begin{pmatrix} \hat{h}(Y_k) \\ 1 \end{pmatrix}. \end{aligned} \tag{5.50}$$

This expression can now be used to obtain in a first step the minimizer $\tilde{\theta} \in \Theta$ and in a second step the minimum itself. To calculate the minimum of (5.50) the function *optimize*, which is already implemented in R, is used.

5.5.3 A Simulation Study

Finally, some data is generated and a simulation study is conducted in this section. There are mainly three blocks of unknown quantities that have to be fixed for simulations. First, having a look at the model equation

$$h(Y) = g(X) + \varepsilon,$$

this model already contains four components, namely the regression function g, the transformation function h, the distribution of X and the distribution of ε, that can be chosen differently when conducting simulations. Additionally, the same simulation parameters as in most simulation studies such as bandwidths, kernel functions, the sample size etc. have to be chosen. The third group consists of the bootstrap components, e.g., the distributions of W and ξ or the number of bootstrap repetitions. Although it is conjectured by the author that all of these choices might have an impact on the performance of the test in (5.9), examining the influence of each of these choices would go beyond the scope of this thesis. Further, any comparison to other testing approaches is difficult since to the authors knowledge only Neumeyer et al. (2016) and Hušková et al. (2019) provided comparable tests, but considered a parametric transformation and alternatives with respect to the error distribution in their simulations, which are not covered by model (5.1). Consequently, possibly the first simulations for a hypothesis test in model (5.1) testing for a parametric transformation function are given in the following.

Remark 5.5.2 *It might be possible to adapt the testing procedure to the heteroscedastic model (3.1) by applying the estimation techniques presented in Section 4. The resulting test might be more comparable to other model tests like for example those of Neumeyer et al. (2016) and Hušková et al. (2019).*

Fixing the Model Related Components

Throughout this section,

$$g(X) = 4X - 1, \quad X \sim \mathcal{U}([0,1]) \quad \text{and} \quad \varepsilon \sim \mathcal{N}(0,1)$$

are chosen. Moreover, the null hypothesis of h belonging to the Yeo-Johnson transforms (5.3) with parameter $\theta \in \Theta_0 = [0,2]$ is tested. Actually, these transforms are standardized in advance to match the identification constraints

$$h(0) = 0 \quad \text{and} \quad h(1) = 1.$$

Later, the cases $\theta = 0$, $\theta = 0.5$, $\theta = 1$ and $\theta = 2$ are simulated. It remains to define appropriate alternatives. The choice

$$h(Y) = \frac{\Lambda_{\theta_0}(Y) + cr(Y) - \Lambda_{\theta_0}(0) - cr(0)}{\Lambda_{\theta_0}(1) + cr(1) - \Lambda_{\theta_0}(0) - cr(0)}$$

for some $\theta \in [0,2]$, some strictly increasing function r and some $c > 0$ leads to the problem, that the observations are simulated via $Y = h^{-1}(g(X) + \varepsilon)$, that is, it is necessary to know the inverse of the chosen transformation function, which in general cannot be calculated straightforwardly for such an h. Furthermore, note that the standardization in order to fulfil the identification constraints indirectly leads to a convex combination of Λ_{θ_0} and r. In the following, this convex combination (now for the inverse functions) is done directly via

$$h^{-1}(Y) = \frac{(1-c)(\Lambda_{\theta_0}^{-1}(Y) - \Lambda_{\theta_0}^{-1}(0)) + c(r(Y) - r(0))}{(1-c)(\Lambda_{\theta_0}^{-1}(1) - \Lambda_{\theta_0}^{-1}(0)) + c(r(1) - r(0))} \tag{5.51}$$

for some $\theta_0 \in [0,2]$, some strictly increasing function r and some $c \in [0,1]$. The inverse Yeo-Johnson transforms are given by

$$\Lambda_\theta^{-1}(Y) = \begin{cases} (1+\theta Y)^{\frac{1}{\theta}} - 1, & \text{if } Y \geq 0, \theta \neq 0 \\ \exp(Y) - 1, & \text{if } Y \geq 0, \theta = 0 \\ 1 - (1 + (\theta - 2)Y)^{\frac{1}{2-\theta}}, & \text{if } Y < 0, \theta \neq 2 \\ 1 - \exp(-Y), & \text{if } Y < 0, \theta = 2 \end{cases}. \tag{5.52}$$

Remark 5.5.3 *1. In general, even with the definition (5.51) it is not clear if a growing factor c leads to a growing distance (5.6). Indeed, the opposite might be the case, if r is somehow close to the class of transformation functions considered in the null hypothesis. To illustrate this phenomenon, let $\theta_1 \neq \theta_2 \in [0,2]$ and consider for some $c \geq 0$ the transformation function*

$$h(Y) = \frac{\Lambda_{\theta_1}(Y) - \Lambda_{\theta_1}(0) + c(\Lambda_{\theta_2}(Y) - \Lambda_{\theta_2}(0))}{\Lambda_{\theta_1}(1) - \Lambda_{\theta_1}(0) + c(\Lambda_{\theta_2}(1) - \Lambda_{\theta_2}(0))}.$$

While for $c = 0$ the null hypothesis is valid, the alternative in general holds for $c > 0$. Nevertheless, with growing c the transformation function will be closer to $Y \mapsto \frac{\Lambda_{\theta_2}(Y)-\Lambda_{\theta_2}(0)}{\Lambda_{\theta_2}(1)-\Lambda_{\theta_2}(0)}$, that is, the distance decreases. Hence, a growing c does not necessarily imply a somehow "stronger" violation of the null hypothesis.

2. *If one defines the transformation function via its inverse as in (5.51), it holds that*

$$h \in \left\{ Y \mapsto \Lambda_{\theta_0}(Y(\Lambda_{\theta_0}^{-1}(1) - \Lambda_{\theta_0}^{-1}(0)) + \Lambda_{\theta_0}^{-1}(0)) : \theta \in \Theta \right\}$$

for $c = 0$, that is, the null hypothesis in (5.4) would be violated even for $c = 0$. Since for computational reasons an analytical form of the inverse transform is required, it will be proceeded as in (5.51) to simulate the alternative case and as in (5.4) to simulate the null hypothesis case.

Simulations are conducted for

$$r_1(Y) = 5\Phi(Y), \tag{5.53}$$

$$r_2(Y) = \exp(Y), \tag{5.54}$$

$$r_3(Y) = Y^3, \tag{5.55}$$

where Φ denotes the cumulative distribution function of a standard normal distribution, and $c = 0, 0.2, 0.4, 0.6, 0.8, 1$. Although quite similar to r_1 sometimes the logit function

$$r_4(Y) = \frac{5\exp(Y)}{1+\exp(Y)} \tag{5.56}$$

is used for illustration reasons as well. The prefactor in (5.53) and (5.56) is introduced due to the fact that the values of Φ and the logit function $\frac{\exp(\cdot)}{1+\exp(\cdot)}$ are rather small compared to the values of Λ_θ, that is, even when using the presented convex combination in (5.51), Λ_{θ_0} (except for $c = 1$) would dominate the "alternative part" r of the transformation function. In principle, other prefactors are conceivable as well. Note that r_2 and Λ_0 only differ with respect to a different standardization. Therefore, if h is defined via (5.51) with r_2 from (5.54) the resulting function is close to the null hypothesis case for $c = 1$. Consequently, it is conjectured that the rejection probabilities at some point may decrease for growing c.

Choosing Kernel Functions and Bandwidths

Calculating the test statistic as described in Section 5.5.2 requires a weighting function w and an appropriate nonparametric estimator of h. The weighting function is set equal to one (that is, the test statistic is calculated without weighting the data), while the nonparametric estimator of h is calculated as Colling and Van Keilegom (2019) did. For details of the approach see Section 5.5.1. As in the paper of Colling and Van Keilegom (2019) the Epanechnikov kernel

$$K(y) = \frac{3}{4}(1 - y^2)I_{[-1,1]}(y)$$

is used and the bandwidths are chosen by the normal reference rule (see for example Silverman (1986)):

$$h_u = \left(\frac{40\sqrt{\pi}}{n}\right)^{\frac{1}{5}} \hat{\sigma}_u,$$

$$h_x = \left(\frac{40\sqrt{\pi}}{n}\right)^{\frac{1}{5}} \hat{\sigma}_x,$$

where $\hat{\sigma}_u^2$ and $\hat{\sigma}_x^2$ are estimators for the variance of $U = \mathcal{T}(Y)$ and X, respectively. The number of evaluation points N_x for the nonparametric estimator of h is set equal to 100 (see Section 5.5.1 for details). The integral in (5.48) is computed by applying the function *integrate*, which is already implemented in *R*.

Bootstrap Implementation

After the model components have been fixed and the test statistic can be calculated accordingly, only the question of how to estimate the bootstrap quantiles needs to be answered. Recall the definition of Algorithm 5.4.1 and its notations. In each of $M = 200$ simulation runs $n = 100$ independent and identically distributed random pairs $(Y_1, X_1), ..., (Y_n, X_n)$ are generated as described before. In each simulation run $B = 250$ bootstrap test statistics, which are based on $m = 100$ bootstrap observations $(Y_1^*, X_1^*), ..., (Y_m^*, X_m^*)$, were calculated as in Algorithm 5.4.1 using

$$W \sim \mathcal{U}([-1, 1]), \quad \xi \sim \mathcal{N}(0, 1) \quad \text{and} \quad a_n = b_n = 0.1.$$

Remark 5.5.4 *Among other things, the nonparametric estimation of h, the integration in (5.48), the optimization with respect to θ and the number of bootstrap repetitions cause the simulations to be quite computationally demanding. Hence, an interface for C++ as well as parallelization were used to conduct the simulations.*

Results and Interpretation

The main results of the simulation study are presented in Table 5.1. There, the rejection probabilities of the settings (5.53)–(5.55) for $c = 0, 0.2, 0.4, 0.6, 0.8, 1$ and $\theta_0 = 0, 0.5, 1, 2$ are listed. The significance level was set equal to 0.05 and 0.10. To obtain more precise estimators of the rejection probabilities under the null hypothesis, 800 simulation runs were performed for each choice of θ_0 under the null hypothesis, whereas in the remaining alternative cases 200 runs were conducted.
First, note that the test sticks to the level or is even a bit conservative. Second, the rejection probabilities not only differ between different choices of r, but also between different transformation parameters θ_0 that are inserted in (5.51). Third, the power of the test is, especially for $\theta_0 = 2$ and model (5.55), even for small values of c quite high in some cases. Fourth, for small deviations from the null hypothesis the rejection probabilities are sometimes extremely small and especially smaller than those under the null hypothesis.

Alternative/ Parameter		$\theta = 0$		$\theta = 0.5$		$\theta = 1$		$\theta = 2$	
Level		$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$
null hyp.		0.01000	0.0400	0.03125	0.0875	0.03125	0.07750	0.01625	0.05625
$5\Phi(Y)$	c=0.2	0.000	0.010	0.075	0.105	0.010	0.015	0.000	0.020
	c=0.4	0.000	0.000	0.020	0.045	0.000	0.015	0.120	0.200
	c=0.6	0.100	0.155	0.035	0.050	0.085	0.150	0.415	0.545
	c=0.8	0.685	0.765	0.110	0.210	0.505	0.645	0.785	0.890
	c=1	0.965	0.990	0.925	0.975	0.975	0.985	0.985	0.990
$\exp(Y)$	c=0.2	0.010	0.035	0.030	0.045	0.515	0.640	0.885	0.965
	c=0.4	0.015	0.040	0.000	0.005	0.060	0.135	0.870	0.980
	c=0.6	0.035	0.085	0.000	0.005	0.005	0.005	0.625	0.815
	c=0.8	0.020	0.040	0.010	0.040	0.000	0.005	0.185	0.325
	c=1	0.020	0.065	0.030	0.090	0.025	0.095	0.050	0.105
Y^3	c=0.2	0.330	0.505	0.730	0.855	0.810	0.905	0.930	0.995
	c=0.4	0.730	0.865	0.815	0.945	0.875	0.970	0.915	0.990
	c=0.6	0.880	0.940	0.895	0.960	0.950	0.995	0.940	0.990
	c=0.8	0.895	0.965	0.925	0.975	0.935	0.990	0.915	0.980
	c=1	0.980	0.990	0.960	0.990	0.939	0.990	0.940	0.985

Table 5.1: Rejection probabilities at $\theta_0 = 0, \theta_0 = 0.5, \theta_0 = 1, \theta_0 = 2$ for r chosen as in (5.53)–(5.55).

Apart from other "classical" reasons such as the choice of the model parameters or the bandwidths and kernel functions, there are two reasons, that explain at least some of these observations. First, the class of Yeo-Johnson transforms seems to be quite general and second the testing approach itself is rather flexible due to the minimization with respect to γ. Having a look at the definition of the test statistic in (5.7) and (5.8), it attains small values if the true transformation function is close to a linear transformation of $\Lambda_{\tilde{\theta}}$ for some appropriate $\tilde{\theta} \in [0, 2]$. In the following, this issue will be explored further by analysing some graphics.

All of the figures that occur in the following have the same structure and consist of four subgraphics. The upper left graphic shows the true transformation function h with inverse function as in (5.51). Due to the choice of $g(X) = 4X - 1$ and $X \sim \mathcal{U}([0, 1])$ the vertical axis reaches from -1 to 3, which would be the support of $h(Y)$ if the error is neglected. In the upper right corner the parametric estimator $\Lambda_{\tilde{\theta}}$ of this function, which is based on $\tilde{\theta}$ from (5.17), is displayed. Both of these functions are then plotted against each other in the lower left corner. Finally, the function $Y \mapsto \Lambda_{\theta_0}(Y(\Lambda_{\theta_0}^{-1}(1) - \Lambda_{\theta_0}^{-1}(0)) + \Lambda_{\theta_0}^{-1}(0))$, which somehow represents the part of the true transformation function h, which corresponds to the null hypothesis, is shown in the last graphic. The arguably most informative of these graphics is that in the lower left corner since there one can see whether the true transformation function can be approximated by a linear transform of some $\Lambda_{\tilde{\theta}}$, $\tilde{\theta} \in [0, 2]$, which is an indicator for the test probably not rejecting the null hypothesis.

As already mentioned, the rejection probabilities not only differ between different deviation functions r, but also for different choices of the transformation parameter. For example, when considering r as in (5.53) with $c = 0.6$ the rejection probabilities for $\theta_0 = 0.5$ amount to 0.035 for $\alpha = 0.05$ and to 0.050 for $\alpha = 0.10$, while for $\theta = 2$ they are 0.415 and 0.545. Figures 5.1 and 5.2 explain why the rejection probabilities differ that much. While for $\theta_0 = 0.5$ the transformation function can be approximated quite well by transforming $\Lambda_{\tilde{\theta}} = \Lambda_{1.06}$ linearly, the best approximation for $\theta_0 = 2$ is given by $\Lambda_{\tilde{\theta}} = \Lambda_{1.94}$ and seems to be relatively bad. The best approximation for $c = 1$ can be reached for θ around 1.4.

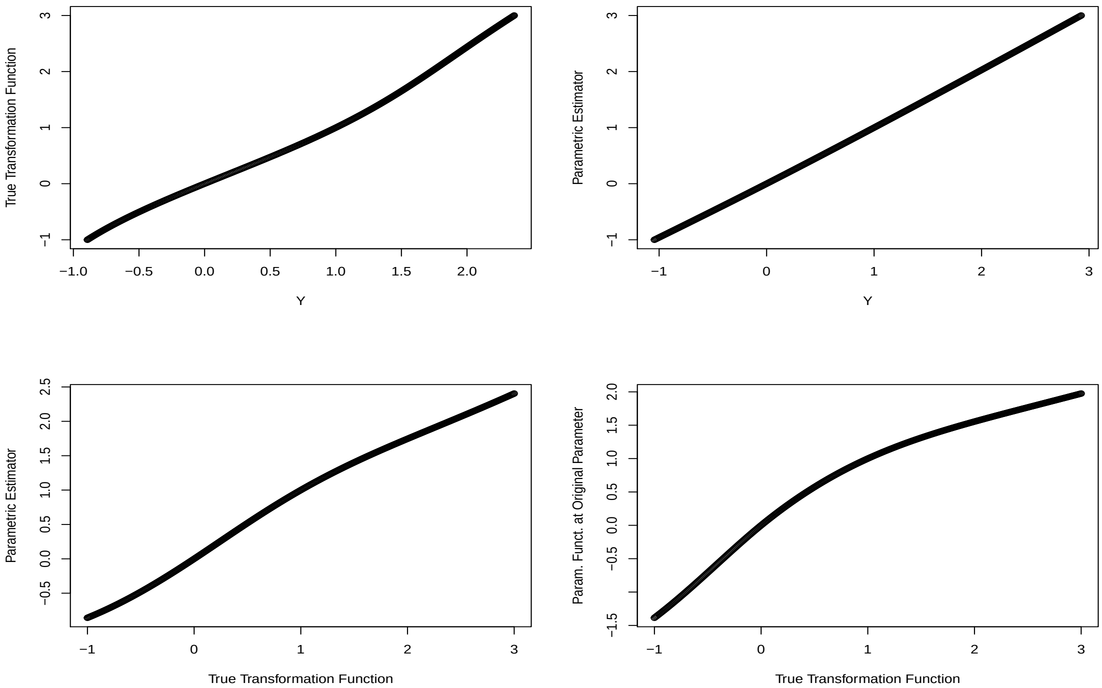

Figure 5.1: Some transformation functions for $\theta = 0.5, c = 0.6$ and r as in (5.53).

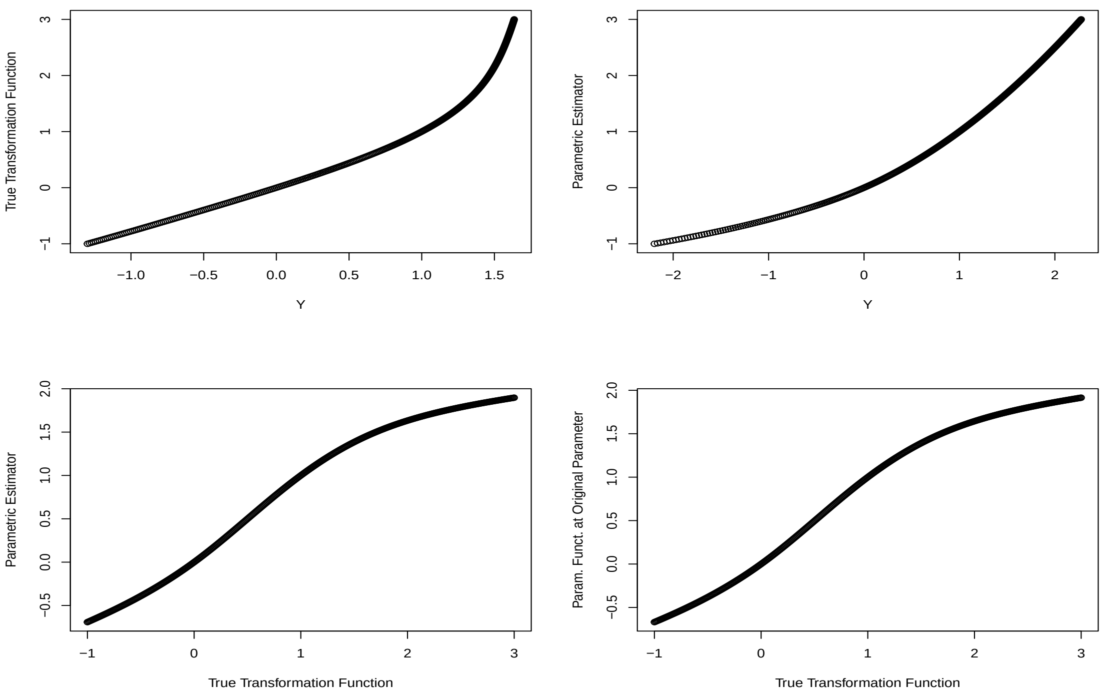

Figure 5.2: Some transformation functions for $\theta = 2, c = 0.6$ and r as in (5.53).

Nevertheless, the differences between the rejection probabilities when using different r seem to be more severe. The graphics in Figure 5.3 deal with the logit function as in (5.56) for $c = 1$. The Yeo-Johnson transforms are so flexible that even for $c = 1$ the logit function can be approximated quite well (for θ around 1.53). Table 5.2 lists the empirical means and variances of some corresponding quantities such as the estimated θ or the value of the test statistic. As can be seen there, the value of the test statistic is even below that of the null hypothesis case for small values of $c \neq 0$, which results in relatively low rejection probabilities. As a side note, the empirical mean of the estimated 0.10-quantiles is below that of the test statistic although the test rejects only in $8,75\%$ of the simulation runs. Here and throughout the whole simulation study, the variance of the estimated transformation parameter is relatively small.

In contrast to the situation for r as in (5.56), considering $\theta_0 = 2$ and r as in (5.55) results in

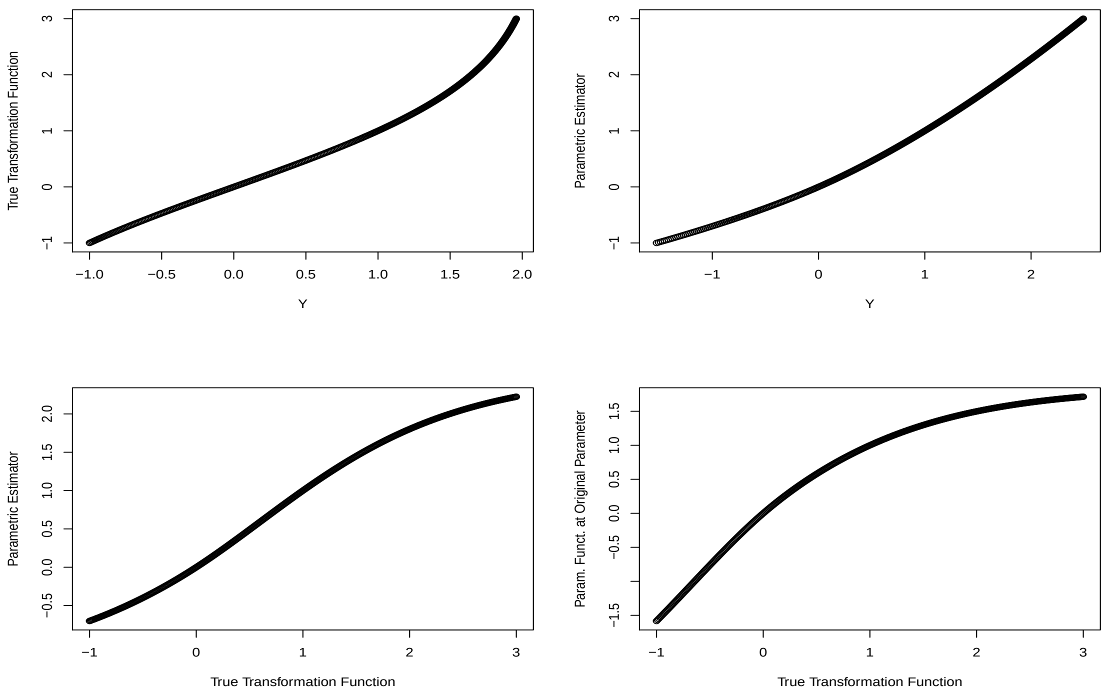

Figure 5.3: Some transformation functions for $c = 1$ and r as in (5.56).

Convex Weights	null hyp.	$c = 0.2$	$c = 0.4$	$c = 0.6$	$c = 0.8$	$c = 1$
Mean of $\tilde{\theta}$	0.461	0.513	0.664	0.831	1.159	1.532
Mean of T_n	2.669	2.364	2.368	2.128	1.849	4.352
Mean of 0.05-bootstrap-quantiles	3.835	3.500	3.700	3.546	3.705	5.036
Mean of 0.10-bootstrap-quantiles	2.663	2.367	2.413	2.171	1.995	3.801
Variance of $\tilde{\theta}$	0.014	0.014	0.017	0.020	0.028	0.042
Variance of T_n	2.381	1.471	1.931	1.402	1.417	6.733
Rej. Prob. for $\alpha = 0.10$	0.0875	0.080	0.045	0.065	0.005	0.270

Table 5.2: Some estimated quantities of the distribution of T_n for $\theta = 0.5$ and r as in (5.56).

a completely different picture. As can be seen in Figure 5.4 even for $c = 0.2$ the resulting h differs so much from the null hypothesis that it can not be linearly transformed into a Yeo-Johnson transform (see the lower left subgraphic). Consequently, the rejection probabilities are rather high. Note that not only the values of the test statistic seem to explode, but those of the bootstrap quantiles as well. Although not intuitive, this is consistent with the findings in Section 5.4 and equation (5.37) since equation (5.37) only requires the bootstrap quantiles to grow less than the test statistic itself.

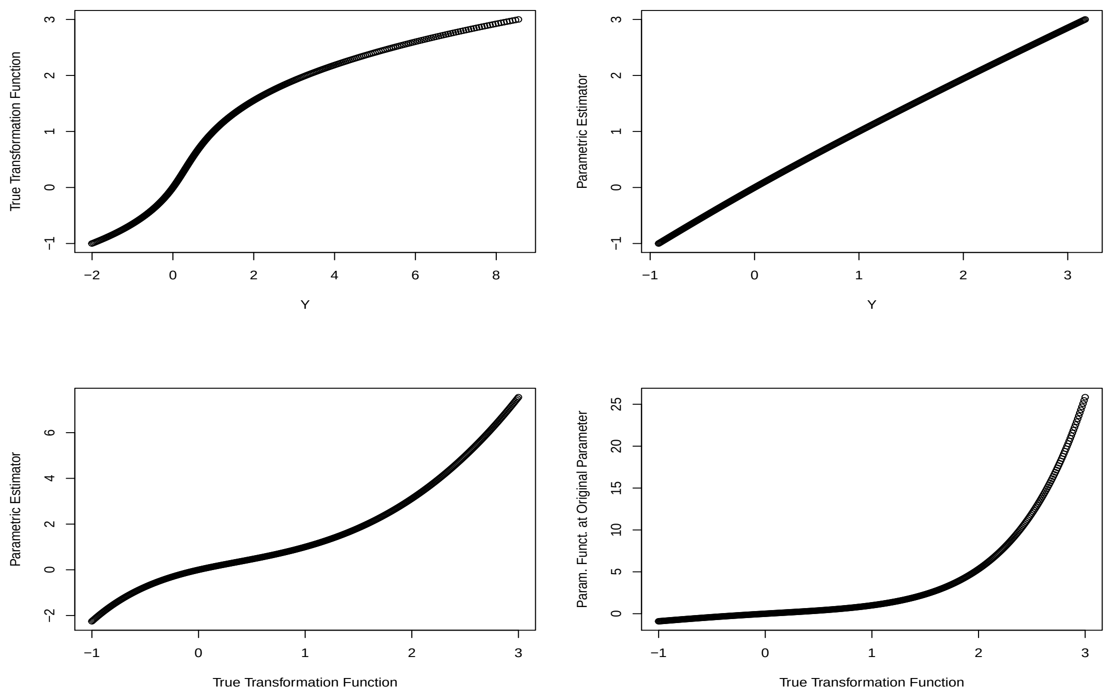

Figure 5.4: Some transformation functions for $\theta = 2, c = 0.2$ and r as in (5.55).

Convex Weights	null hyp.	$c = 0.2$	$c = 0.4$	$c = 0.6$	$c = 0.8$	$c = 1$
Mean of $\tilde{\theta}$	1.978	0.891	0.754	0.695	0.627	0.644
Mean of T_n	10.59	824	1460	1797	1903	2620
Mean of estimated 0.05-quantiles	15.95	621	1105	1320	1447	1944
Mean of estimated 0.10-quantiles	10.77	533	943	1131	1231	1653
Variance of $\tilde{\theta}$	0.002	0.019	0.024	0.021	0.028	0.024
Variance of T_n	31.301	462534	3105125	2066320	4119168	5865643

Table 5.3: Some estimated quantities of the distribution of T_n for $\theta = 2$ and r as in (5.55).

5.6 Discussion

A new goodness of fit test for the null hypothesis of a parametric transformation function in the transformation model (5.1) has been provided. The presented approach allows more general models than Szydłowski (2017). The asymptotic behaviour of the test statistic has been examined and consistency of the corresponding test (5.9) has been proven. In

addition, consistency of a proposed bootstrap algorithm has been shown. In contrast to Neumeyer et al. (2016) or Hušková et al. (2019), (local) alternatives with respect to the transformation function instead of the variance function have been considered. Moreover, a consistent test for the relevant hypotheses (5.28) has been provided for the first time.
There are several opportunities to adjust and examine the presented methods further. First, due to the high generality of the model lots of simulation parameters had to be chosen in Section 5.5.3. Unfortunately, analysing the sensitivity of the test with respect to each of the parameters would have gone beyond the scope of this thesis, so that only the influence of a few of them has been studied. Further, it remains to conduct simulations in the case of the relevant hypotheses. In this context, the choice of η, that is, the threshold up to which a transformation model is expected to fit the data sufficiently well, would be interesting. This issue is strongly related to the interpretation of the distance between two transformation models (which would interest on its own).
At the beginning of this chapter, two possible distances (5.5) and (5.6) were suggested for the comparison of two transformation functions. Although it is conjectured that the corresponding theory for a test statistic based on (5.5) can be deduced similarly, further examination would be interesting. It might even be the case that the corresponding test is accompanied with more power since the minimization with respect to the identification coefficients c_1 and c_2 is replaced by fixing them beforehand.
At last, there are many possibilities to extend the test for example to heteroscedastic models by applying the estimating approaches of Chapter 4 or to the case of finite parameter sets. Some thoughts about these extensions are given in Section 5.9.

5.7 Assumptions

In the following assumptions let $\mathcal{Y}$ denote the support of Y (which may depend on n under local alternatives). Further, F_S denotes the distribution function of S as in (5.11) and $\mathcal{T}_S$ denotes the transformation $s \mapsto (F_S(s) - F_S(0))/(F_S(1) - F_S(0))$.

(A1) The sets C_1, C_2 and Θ are compact.

(A2) The weighting function w is continuous with a compact support $\mathcal{Y}_w \subset \mathcal{Y}$.

(A3) The map $(y, \theta) \mapsto \Lambda_\theta(y)$ is twice continuously differentiable on $\mathcal{Y}_w$ with respect to θ and the (partial) derivatives are continuous in $(y, \theta) \in \mathbb{R} \times \Theta$.

(A4) There exists a unique transformation h such that model (5.1) holds with independent X and ε. $(Y_i, X_i), i = 1, ..., n$, are independent and identically distributed observations from model (5.1). The function h_0 defined in (5.16) is strictly increasing as well as continuously differentiable. Moreover, h is strictly increasing and h and r are continuous on $\mathcal{Y}_w$. F_Y is strictly increasing on the support of Y.

(A5) Minimizing the function $M : \Upsilon \to \mathbb{R}, \gamma = (c_1, c_2, \theta^t)^t \mapsto E\big[w(Y)(h_0(Y)c_1 + c_2 - \Lambda_\theta(Y))^2\big]$ leads to a unique solution $\gamma_0 = (c_{1,0}, c_{2,0}, \theta_0)$ in the interior of Υ. For all $\theta \neq \tilde{\theta}$ one has $\sup\limits_{y \in \mathrm{supp}(w)} \Big|\frac{\Lambda_\theta(y) - \Lambda_\theta(0)}{\Lambda_\theta(1) - \Lambda_\theta(0)} - \frac{\Lambda_{\tilde{\theta}}(y) - \Lambda_{\tilde{\theta}}(0)}{\Lambda_{\tilde{\theta}}(1) - \Lambda_{\tilde{\theta}}(0)}\Big| > 0$.

(**A6**) The Hessian matrix $\Gamma := \operatorname{Hess} M(\gamma_0)$ with M as in (5.29) is positive definite (with γ_0 as in (**A5**) or (**A5'**)).

(**A7**) The estimator $\hat{h}$ of the transformation function h fulfils (5.13) for some function ψ and h from (5.15). For some $\mathcal{U}_0$ (independent of n under local alternatives) with $\mathcal{T}_S(h(\mathcal{Y}_w)) \subset \mathcal{U}_0$ the function class $\{z \mapsto \psi(z,t) : t \in \mathcal{U}_0\}$ is Donsker with respect to P^Z and $E[\psi(Z_1,t)] = 0$ for all $t \in \mathcal{U}_0$. The fourth moment $E[w(h_0^{-1}(S_1))\psi(Z_1,U_1)^4]$ is finite and the conditional moments $E[w(h_0^{-1}(S_1))\psi(Z_1,U_2)^2|Z_1 = z]$ are locally bounded.

Remark 5.7.1 *1. Since the support of w is compact, continuity on $\mathcal{Y}_w$ results in uniform continuity.*

*2. Lemma yields 5.8.1 that (**A7**) is fulfilled for the estimator of Colling and Van Keilegom (2019).*

When considering fixed alternatives or the relevant hypothesis H_0', assumptions (**A4**),(**A5**) and (**A7**) are replaced by the following assumptions (**A4'**),(**A5'**) and (**A7'**) (assumption (**A7'**) is only relevant for H_0'). Note that h is a fixed function then, not depending on n.

(**A4'**) There exists a unique transformation h such that model (5.1) holds with X and ε independent. The function h is strictly increasing and continuous on $\mathcal{Y}_w$.

(**A5'**) Minimizing the function $M : \Upsilon \to \mathbb{R}, \gamma = (c_1, c_2, \theta^t)^t \mapsto E\big[w(Y)(h(Y)c_1 + c_2 - \Lambda_\theta(Y))^2\big]$ leads to a unique solution $\gamma_0 = (c_{1,0}, c_{2,0}, \theta_0)$ in the interior of Υ. For all $\theta \neq \tilde{\theta}$ one has $\sup\limits_{y \in \operatorname{supp}(w)} \left|\frac{\Lambda_\theta(y) - \Lambda_\theta(0)}{\Lambda_\theta(1) - \Lambda_\theta(0)} - \frac{\Lambda_{\tilde{\theta}}(y) - \Lambda_{\tilde{\theta}}(0)}{\Lambda_{\tilde{\theta}}(1) - \Lambda_{\tilde{\theta}}(0)}\right| > 0$.

(**A7'**) The transformation estimator $\hat{h}$ fulfils (5.13) for some function ψ. For some $\mathcal{U}_0 \supset \mathcal{T}_S(h(\mathcal{Y}_w))$ the function class $\{z \mapsto \psi(z,t) : t \in \mathcal{U}_0\}$ is Donsker with respect to P^Z and $E[\psi(Z_1,t)] = 0$ for all $t \in \mathcal{U}_0$. Further, one has $E[\psi(Z_1,U_2)^2] < \infty$.

Now, some assumptions necessary for the bootstrap theory in Section 5.4 are given. The same notation as there is used for the probability space. The expectation with respect to P_2^1 is written as $E[\cdot|\omega]$.

(**A7***) The following properties are meant conditional on the data $(Y_i, X_i), i = 1, ..., n$, and thus define for fixed $n \in \mathbb{N}$ some subsets $A_n \in \mathcal{A}_1$ of Ω_1, where these properties are valid. Thus, let $\omega \in A_n$. Then, the following conditions are assumed to be fulfilled on A_n.

(i) The bootstrap estimator $\hat{h}^*$ of the transformation function fulfils

$$\limsup_{m\to\infty} P_2^1\left(\omega, \left\{\sup_{y\in\mathcal{K}}\left|\hat{h}^*(y) - h^*(y) - \frac{1}{m}\sum_{j=1}^m \psi^*(S_j^*, X_j^*, \mathcal{T}^*(y))\right| > \frac{\delta}{\sqrt{m}}\right\}\right) = 0$$

(see 5.41) for all $\delta > 0$, for some function ψ^* and h^* from algorithm 5.4.1.

(ii) For $\mathcal{U}_0$ from (**A7**) and $\mathcal{Y}_w$ from (**A2**) one has $\mathcal{T}_{S^*}(h^*(\mathcal{Y}_w)) \subset \mathcal{U}_0$.

(iii) Let $Z^* = (U^*, X^*)$. The function class $\{z \mapsto \psi^*(z,t) : t \in \mathcal{U}_0\}$ is Donsker (for fixed n, but $m \to \infty$) with respect to $P_2^1(\omega, \cdot)^{Z^*}$ (distribution of Z^* conditional on ω) and

$$E[\psi^*(Z_1^*, t)|\omega] = \int \psi^*(Z_1^*(\tilde{\omega}), t) P_2^1(\omega, d\tilde{\omega}) = 0 \quad \text{for all} \quad t \in \mathcal{U}_0.$$

(iv) The fourth moment $E[w(h_0^{-1}(S_1^*))\psi^*(Z_1^*, U_1^*)^4|\omega]$ is finite and the conditional moments $E[w(h^{*-1}(S_1^*))\psi^*(Z_1^*, U_2^*)^2|Z_1^* = z, \omega]$ are locally bounded.

(v) For all compact sets $\mathcal{K} \subseteq \mathbb{R}$ one has

$$P_1\Big(\omega \in \Omega_1 : \sup_{y \in \mathcal{K}, z \in \mathbb{R}^{d_X+1}} |w(y)\psi^*(z, \mathcal{T}^*(y))| > \delta\sqrt{n}\Big) = o(1) \quad \text{for all } \delta > 0. \tag{5.57}$$

(vi) One has

$$\lim_{y \to z} E\left[|\psi^*(z, U_1^*) - \psi^*(y, U_1^*)|\, |\omega\right] = 0 \quad \text{for all } z \text{ in the support of } Z^*.$$

For A_n as defined above, assume $P_1(A_n) \to 1$ for $n \to \infty$.

(**A8***) Define the distribution function of Z^* for some $\omega \in \Omega_1$ by $F_{Z^*}(z) = P_2^1(\omega, \{Z^* \leq z\})$ and assume

$$\sup_{z \in \mathcal{R}^{d_X+1}} |F_{Z^*}(z) - F_Z(z)| = o_p(1). \tag{5.58}$$

Moreover, for all compact $\mathcal{K} \subseteq \mathbb{R}^{d_X+1}$ there exists an appropriate $C > 0$, such that for $n \to \infty$

$$\sup_{z \in \mathcal{K}, s \in \mathbb{R}} w((h^*)^{-1}(s))\psi^*(z, \mathcal{T}_{S^*}(s)) \leq C + o_p(1). \tag{5.59}$$

Further,

$$w((h^*)^{-1}(s))(\psi^*(z, \mathcal{T}_{S^*}(s)) - \psi(z, \mathcal{T}_{S^*}(s))) = o_p(1) \tag{5.60}$$

for all $z \in \mathbb{R}^{d_X+1}$, $s \in \mathbb{R}$ and for ψ from (**A7**) for $n \to \infty$.

(**A9**) Denote the conditional density of $\varepsilon(\theta_0)$ as defined in 5.4.1 given X by $f_{\varepsilon(\theta_0)|X}$. Let $\mathcal{K} \subseteq \mathbb{R}$ be compact and f_ξ be bounded and r-times continuously differentiable with bounded derivatives and denote the k-th derivative of f_ξ by $f_\xi^{(k)}$. Further, assume

$$\sup_{u \in \mathcal{K}} E\bigg[\int ||g_{\theta_0}(X) + u - a_n e||^l |f_\xi^{(j)}(e)| f_{\varepsilon(\theta_0)|X}(u - a_n e|X)\, de\bigg] < C \tag{5.61}$$

for $l \in \{0, j\}$ and

$$\sup_{u \in \mathcal{K}} E\bigg[\int ||\dot{\Lambda}_{\theta_0}(\Lambda_{\theta_0}^{-1}(g_{\theta_0}(X) + u - a_n e))||^j |f_\xi^{(j)}(e)| f_{\varepsilon(\theta_0)|X}(u - a_n e|X)\, de\bigg] < C \tag{5.62}$$

as well as

$$\frac{||\tilde{\theta} - \theta_0||^j}{n a_n} \sup_{u \in \mathcal{K}} \sum_{i=1}^n \left| f_\xi^{(j)}\left(\frac{u - \varepsilon_i(\theta_0)}{a_n}\right)\right| \sup_{||\theta - \theta_0|| < \delta} ||\operatorname{Hess} \Lambda_\theta(Y_i)||^j = \mathcal{O}_p(1) \tag{5.63}$$

for sufficiently large $C > 0, n \in \mathbb{N}$ and all $j = 1, ..., r-1$. Moreover, let

$$E\big[|\Lambda_{\theta_0}(Y)|^r + ||\dot{\Lambda}_{\theta_0}(Y)||^r\big] < \infty \tag{5.64}$$

and

$$\frac{||\tilde{\theta} - \theta_0||^r}{n} \sum_{i=1}^n \sup_{||\theta - \theta_0|| < \delta} ||\operatorname{Hess} \Lambda_\theta(Y_i)||^r = \mathcal{O}_p(1). \tag{5.65}$$

5.7.1 Assumptions Needed for the Estimation of h

Since in general it is not clear under which conditions assumption (**A7**) is fulfilled, in the following assumptions are given which ensure (**A7**) for the estimator of Colling and Van Keilegom (2019) in (5.46) (see Section 1.4 for details). Let, as in (**A7**), $\mathcal{U}_0 \supset \mathcal{T}_S(h(\mathcal{Y}_w))$ (independent of n under local alternatives) belong to the interior of the support of U and assume compactness of $\mathcal{U}_0$. Let $\mathcal{X} \subset \mathbb{R}^{d_X}$ denote the support of X.

(**B1**) The cumulative distribution function F_ε of ε is absolutely continuous and has a density f_ε that is continuous on its support. Furthermore, X and ε are independent.

(**B2**) The transformation Q is strictly increasing and continuously differentiable on $\mathcal{U}_0$, which is a connected (and compact) subset of $\mathbb{R}$.

(**B3**) The set

$$\mathcal{X}_{\partial i} := \left\{ x \in \mathcal{X} : \frac{\partial F_{U|X}(u|x)}{\partial x_i} \neq 0 \quad \text{for all } u \in \mathcal{U}_0 \right\}$$

is nonempty for some $i \in \{1, ..., n\}$ (later the case $i = 1$ is considered w.l.o.g.).

(**B4**) The bandwidths h_x and h_u satisfy for an appropriate $q \in I\!N$

$$\sqrt{n} h_x^q \to 0, \sqrt{n} h_u^q \to 0, \frac{\sqrt{n} h_x^{d_X+2}}{\log(n)} \to \infty, \quad \frac{\sqrt{n} h_x^{d_X} h_u^2}{\log(n)} \to \infty.$$

(**B5**) The kernel K is symmetric with a connected and compact support containing some neighbourhood around 0. Further, K is q-times continuously differentiable with K and K' being of bounded variation. Moreover, $\int K(z)\,dz = 1, \int z^l K(z)\,dz = 0$ for all $l = 1, ..., q-1$.

(**B6**) The kernel L is twice continuously differentiable with uniformly bounded derivatives and with median 0, and $b = b_n > 0$ is a bandwidth sequence that satisfies $nb^4 \to \infty$ and $b\sqrt{n} h_x^{d_X} (\min(h_x, h_u))^2 / \log(n) \to \infty$.

(**B7**) v is a weighting function with a compact support $\mathcal{X}_0 \subseteq \mathcal{X}_{\partial i}$ with nonempty interior. Further, $\int_{\mathcal{X}_0} v(x)\,dx = 1$ and v is q-times continuously differentiable and all these derivatives are uniformly bounded in the interior, i.e.,

$$\sup_{x \in \mathcal{X}_0} \left| \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \cdots \partial x_{d_X}^{\alpha_{d_X}}} v(x) \right| < \infty,$$

for all $\alpha_1, ..., \alpha_{d_X} \in \{0, ..., q-1\}$ with $|\alpha| = \sum_{i=1}^{d_X} \alpha_i \leq m$.

(**B8**) The regression function g is continuously differentiable with respect to x_i on $\mathcal{X}$ for $i = 1, ..., d_X$.

(**B9**) The joint density function $f_{Y,X}(y, x)$ of (Y, X) is uniformly bounded, $(q+2)$-times continuously differentiable and all these derivatives are uniformly bounded, i.e.,

$$\sup_{y:\, \mathcal{T}(y) \in \mathcal{U}_0,\, x \in \mathcal{X}_0} \left| \frac{\partial^{|\alpha|}}{\partial y^{\alpha_0} \partial x_1^{\alpha_1} \cdots \partial x_{d_X}^{\alpha_{d_X}}} f_{Y,X}(y, x) \right| < \infty,$$

for all $\alpha_0, \alpha_1, ..., \alpha_{d_X}$ with $|\alpha| = \sum_{i=0}^{d_X} \alpha_i \leq q+2$. Further, assume $\inf_{y\colon \mathcal{T}(y)\in\mathcal{U}_0} f_Y(y) > 0$, where f_Y is the density function of Y.

(**B10**) Assume

$$\inf_{x\in\mathcal{X}_0} f_X(x) > 0, \inf_{(u,x)\in\mathcal{U}_0\times\mathcal{X}_0} \frac{\partial F_{U|X}(u|x)}{\partial x_1} > 0 \quad \text{and} \quad \inf_{x\in\mathcal{X}_0} \int_0^1 \frac{\frac{\partial F_{U|X}(u|x)}{\partial u}(u,x)}{\frac{\partial F_{U|X}(u|x)}{\partial x_1}(u,x)} du > 0.$$

Remark 5.7.2 *Assumptions (**B1**) and (**B9**) imply continuous differentiability of the transformation function h.*

5.8 Proofs

As was shown in Section 1.4, for some fixed h and under (**A4**), (**B1**)–(**B10**) the difference $\hat{h}(y) - h(y)$ can be rewritten as

$$h(y) - \hat{h}(y) = \frac{1}{n}\sum_{i=1}^n \psi(Z_i, \mathcal{T}(y)) + o_p\Big(\frac{1}{\sqrt{n}}\Big),$$

where (with $Q, \delta, \bar{w}_1, \bar{w}_2$ from Section 1.4)

$$\begin{aligned}\psi(Z_i, u) = \delta_j^{\tilde{w}_1}(1,u) - \delta_j^{\tilde{w}_2}(u,1) + \frac{Q'(u)}{F_U(1) - F_U(0)}\big(I_{\{U_j\leq u\}} - I_{\{U_j\leq 0\}} - F_U(u) + F_U(0)\big)\\ - Q'(u)\frac{F_U(u) - F_U(0)}{(F_U(1) - F_U(0))^2}\big(I_{\{U_j\leq 1\}} - I_{\{U_j\leq 0\}} - F_U(1) + F_U(0)\big).\end{aligned}$$

Before Theorem 5.2.2 is proven, an additional lemma, which ensures (5.13) for the estimator of Colling and Van Keilegom (2019) under local alternatives, is shown first.

Lemma 5.8.1 *Assume (**B1**)–(**B10**) and depending on whether local or fixed alternatives are considered assume (**A4**) or (**A4'**). Then, assumption (**A7**) is fulfilled for the estimator of Colling and Van Keilegom (2019).*
*Moreover, if r is an alternative function, such that (**A4**) is fulfilled, let $A \subseteq \mathbb{R}$ be a compact set which contains 0 in its interior and such that*

$$h(y,\alpha) := \frac{\Lambda_{\theta_0}(y) - \Lambda_{\theta_0}(0) + \alpha(r(y) - r(0))}{\Lambda_{\theta_0}(1) - \Lambda_{\theta_0}(0) + \alpha(r(1) - r(0))}$$

is for all $\alpha \in A$ strictly increasing with respect to y. Further, let $h^{-1}(y,\alpha)$ denote the inverse function of $h(y,\alpha)$ with respect to y and define

$$\mathcal{T}_\alpha(y) = \frac{F_S\Big(\frac{\Lambda_{\theta_0}(\cdot) - \Lambda_{\theta_0}(0) + \alpha(r(y) - r(0))}{\Lambda_{\theta_0}(1) - \Lambda_{\theta_0}(0) + \alpha(r(1) - r(0))}\Big) - F_S(0)}{F_S(1) - F_S(0)}.$$

Then, if $\hat{h}(y,\alpha)$ denotes the estimator of Colling and Van Keilegom (2019) of $h(y,\alpha)$, which is based on $(Y_{\alpha,i}, X_i)$ with $Y_{\alpha,i} = h^{-1}(S_i,\alpha)$, $i = 1, ..., n$, it holds that

$$\hat{h}(y,\alpha) - h(y,\alpha) = \frac{1}{n}\sum_{j=1}^n \psi(Z_j, \mathcal{T}_\alpha(y)) + o_p\big(n^{-\frac{1}{2}}\big)$$

uniformly in $y \in \mathcal{Y}_w$ and $\alpha \in A$ and the process

$$Z_n(y,\alpha) = \sqrt{n}(\hat{h}(y,\alpha) - h(y,\alpha)), \quad (y,\alpha) \in \operatorname{supp}(w) \times A$$

converges weakly to a centred Gaussian process with covariance function

$$c((y,\alpha,\tilde{y},\tilde{\alpha})) = E[\psi(Z_1,\mathcal{T}_\alpha(y))\psi(Z_1,\mathcal{T}_{\tilde{\alpha}}(\tilde{y}))].$$

Proof: Note that in the framework here $S = g(X) + \varepsilon$ is fixed. In the case of a fixed transformation h, the assertion is covered by Theorems 5.1 and 5.2 of Colling and Van Keilegom (2019). Therefore, the assertion only needs to be proven for $h(y,\alpha)$. Due to $\{n^{-\frac{1}{2}} : n \in \mathbb{N}, n \geq N\} \subseteq A$ for a sufficiently large N, the statement for local alternatives would follow from this. The expansion is shown uniformly in $y \in \mathcal{Y}_w$ and $\alpha \in A$. Nevertheless, most arguments used for fixed h are still valid.
Verification of (5.13): The proof follows the same lines as for some fixed h or fixed α. The only extra effort is to check whether all previous $o_p(1)$-terms are still negligible. Let $F_{Y_\alpha}, \hat{F}_{Y_\alpha}$ denote the distibution function and the empirical distibution function, respectively, of $Y_{\alpha,i}, i = 1, ..., n$. First, note that as in Remark 5.2.1

$$U_i = \mathcal{T}_\alpha(Y_{\alpha,i}) = \frac{F_{Y_\alpha}(Y_{\alpha,i}) - F_{Y_\alpha}(0)}{F_{Y_\alpha}(1) - F_{Y_\alpha}(0)}$$

does not depend on α because of $h(0) = 0$ and $h(1) = 1$ for all $h \in \mathcal{H}$. The same holds true for the estimated version

$$\hat{U}_i := \hat{\mathcal{T}}_\alpha(Y_\alpha, i) := \frac{\hat{F}_{Y_\alpha}(Y_{\alpha,i}) - \hat{F}_{Y_\alpha}(0)}{\hat{F}_{Y_\alpha}(1) - \hat{F}_{Y_\alpha}(0)},$$

since $I_{\{Y_{\alpha,j} \leq Y_{\alpha,i}\}} = I_{\{h(Y_{\alpha,j},\alpha) \leq h(Y_{\alpha,i},\alpha)\}}$ is independent of α.
As in (5.11) consider the transformation function Q corresponding to the pretransformed observations, that is

$$Q(\mathcal{T}_\alpha(Y)) = g(X) + \varepsilon.$$

Since the estimator $\hat{Q}$ of Q only depends on estimates $\hat{\mathcal{T}}_\alpha(Y_i)$ of the pretransformed observations and these in turn do not depend on α, the asymptotic expression of $\hat{Q} - Q$ does not change with α. The uniform consistency of $\hat{Q}'$ holds by the same argument. Later, it will be proven that

$$\begin{aligned}
\hat{\mathcal{T}}_\alpha(y) - \mathcal{T}_\alpha(y) &= \frac{1}{F_{Y_\alpha}(1) - F_{Y_\alpha}(0)}(\hat{F}_{Y_\alpha}(y) - \hat{F}_{Y_\alpha}(0) - F_{Y_\alpha}(y) + F_{Y_\alpha}(0)) \\
&\quad - \frac{F_{Y_\alpha}(y) - F_{Y_\alpha}(0)}{(F_{Y_\alpha}(1) - F_{Y_\alpha}(0))^2}(\hat{F}_{Y_\alpha}(1) - \hat{F}_{Y_\alpha}(0) - F_{Y_\alpha}(1) + F_{Y_\alpha}(0)) + o_p\big(n^{-\frac{1}{2}}\big) \\
&= \mathcal{O}_p\left(\frac{1}{\sqrt{n}}\right)
\end{aligned}$$

uniformly in $y \in \mathcal{Y}_w$ and $\alpha \in A$, so that

$$\begin{aligned}
\hat{h}(y) - h(y) &= \hat{Q}(\hat{\mathcal{T}}_\alpha(y)) - Q(\mathcal{T}_\alpha(y)) \\
&= \hat{Q}(\mathcal{T}_\alpha(y)) - Q(\mathcal{T}_\alpha(y)) + \hat{Q}'(\mathcal{T}_\alpha(y))(\hat{\mathcal{T}}_\alpha(y) - \mathcal{T}_\alpha(y)) + o_p\big(n^{-\frac{1}{2}}\big)
\end{aligned}$$

$$= \hat{Q}(\mathcal{T}_\alpha(y)) - Q(\mathcal{T}_\alpha(y)) + Q'(\mathcal{T}_\alpha(y))(\hat{\mathcal{T}}_\alpha(y) - \mathcal{T}_\alpha(y)) + o_p\big(n^{-\frac{1}{2}}\big).$$

$$= \frac{1}{n}\sum_{j=1}^{n} \psi(Z_j, \mathcal{T}_\alpha(y)) + o_p\big(n^{-\frac{1}{2}}\big).$$

Note that $\hat{\mathcal{T}}_\alpha(y) - \mathcal{T}_\alpha(y) = \mathcal{O}_p\big(\frac{1}{\sqrt{n}}\big)$ was used to ensure negligibility of some terms caused by the Taylor expansion or replacing $\hat{Q}'$ by Q', so that the remaining task consists not only in proving the expression above, but the stated order as well. If one shows

$$\sqrt{n}(\hat{F}_{Y_\alpha}(y) - F_{Y_\alpha}(y)) = \sqrt{n}(\hat{F}_S(h(y,\alpha)) - F_S(h(y,\alpha))) = \mathcal{O}_p(1)$$

uniformly in $y \in \mathbb{R}$ and $\alpha \in A$, both properties follow from the equality

$$\frac{\hat{a}}{\hat{b}} - \frac{a}{b} = \frac{1}{b}(\hat{a} - a) - \frac{a}{b^2}(\hat{b} - b) - \frac{\hat{b} - b}{\hat{b}b}\left(\hat{a} - a - \frac{a(\hat{b} - b)}{b}\right)$$

for arbitrary estimators $\hat{a}$ and $\hat{b}$. By standard arguments (see for example Van der Vaart and Wellner (1996)), it can be shown that the class

$$\mathcal{F}' = \{I_{(-\infty,t]}(\cdot) : t \in \mathbb{R}\}$$

is Donsker, so that the class

$$\mathcal{F} = \{I_{(-\infty,h(y,\alpha)]}(\cdot) : y \in \mathcal{Y}_w, \alpha \in A\} \subseteq \mathcal{F}'$$

is (as a subclass of a Donsker class) Donsker as well.

Proof of Weak Convergence: Convergence of the marginal distributions with the corresponding covariance function can be shown by the ordinary CLT. Furthermore, it has already been shown that

$$Z(y,\alpha) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n} \psi(Z_j, \mathcal{T}_\alpha(y)) + o_p(1).$$

For fixed h Colling and Van Keilegom (2019) proved that for any compact interval $\mathcal{U} \subseteq \mathcal{T}_S(\mathcal{S})$ the class

$$\tilde{\mathcal{F}} = \{z \mapsto \psi(z,u) : u \in \mathcal{U}\}$$

is Donsker. Although the proof there is carried out using (Y_i, X_i) (for fixed h) instead of $Z_i = (U_i, X_i)$ the statement stays valid since $\psi(Z_i, u)$ does not depend on h (see Section 1.4). Thus, when choosing $\mathcal{U}_0$ from Section 5.7.1 the class

$$\mathcal{F} = \{z \mapsto \psi(z, \mathcal{T}_\alpha(y)) : y \in \mathcal{Y}_w, \alpha \in A\} \subseteq \{z \mapsto \psi(z,u) : u \in \mathcal{U}_0\}$$

is (as a subclass of a Donsker class) Donsker as well.

The remaining conditions for ψ in assumption (**A7**) follow from the results of Colling and Van Keilegom (2019) since ψ does not depend on α. $\square$

5.8.1 Proof of Theorem 5.2.2

For ease of presentation define

$$M_n(c_1, c_2, \theta^t)^t = \sum_{j=1}^{n} w(Y_j)(\hat{h}(Y_j)c_1 + c_2 - \Lambda_\theta(Y_j))^2, \tag{5.66}$$

such that $T_n = \min_{\gamma\in\Upsilon} M_n(\gamma)$. Let $\tilde{\gamma} = (\tilde{c}_1, \tilde{c}_2, \tilde{\theta})$ denote the minimizer of M_n and γ_0 be the vector such that

$$h_0(y)c_{1,0} + c_{2,0} = \Lambda_{\theta_0}(y) \tag{5.67}$$

(see (**A5**)), that is, $c_{2,0} = \Lambda_{\theta_0}(0)$ and $c_{1,0} = \Lambda_{\theta_0}(1) - \Lambda_{\theta_0}(0)$. It already has been shown that

$$h(\cdot)c_1 + c_2 = \Lambda_{\theta_0}(\cdot) + n^{-\frac{1}{2}}r(\cdot)$$

for $c_2 = \Lambda_{\theta_0}(0) + n^{-\frac{1}{2}}r(0)$ and $c_1 = \Lambda_{\theta_0}(1) - \Lambda_{\theta_0}(0) + n^{-\frac{1}{2}}(r(1) - r(0))$, so that

$$\begin{aligned}
h(y)c_{1,0} + c_{2,0} - \Lambda_{\theta_0}(y) &= h(y)(c_{1,0} - c_1) + c_{2,0} - c_2 + n^{-\frac{1}{2}}r(y)\\
&= n^{-\frac{1}{2}}r(y) - n^{-\frac{1}{2}}h(y)(r(1) - r(0)) - n^{-\frac{1}{2}}r(0)\\
&= n^{-\frac{1}{2}}\big(r(y) - r(0) - h_0(y)(r(1) - r(0))\big) + o\big(n^{-\frac{1}{2}}\big)\\
&= n^{-\frac{1}{2}}c_{1,0}r_0(y) + o\big(n^{-\frac{1}{2}}\big) \qquad (5.68)\\
&= \mathcal{O}\big(n^{-\frac{1}{2}}\big)
\end{aligned}$$

uniformly for $y \in \mathcal{Y}_w$.

To get a rough idea of how the proof is structured, the main steps are sketched in the following:

Step 1: Reduce the problem of minimizing M_n to that of the quadratic function (h_0 was defined in (5.16))

$$\begin{aligned}
q(z) := &\sum_{k=1}^{n} w(h_0^{-1}(S_k))(\hat{h}(Y_k)c_{1,0} + c_{2,0} - \Lambda_{\theta_0}(Y_k))^2\\
&+ 2z^t \sum_{k=1}^{n} w(h_0^{-1}(S_k))(\hat{h}(Y_k)c_{1,0} + c_{2,0} - \Lambda_{\theta_0}(Y_k))R(S_k) + nz^t\Gamma z
\end{aligned}$$

and obtain the corresponding solution

$$\begin{aligned}
z_0 &= -\Gamma^{-1}\frac{1}{n}\sum_{k=1}^{n} w(h_0^{-1}(S_k))(\hat{h}(Y_k)c_{1,0} + c_{2,0} - \Lambda_{\theta_0}(Y_k))R(S_k)\\
&= -\frac{c_{1,0}}{n}\Gamma^{-1}\sum_{k=1}^{n}\varphi(Z_k) - \frac{c_{1,0}}{\sqrt{n}}\Gamma^{-1}\beta + o_p\big(n^{-\frac{1}{2}}\big)
\end{aligned}$$

with

$$\beta = E[w(h_0^{-1}(S_1))r_0(h_0^{-1}(S_1))R(S_1)]. \tag{5.69}$$

Step 2: Insert z_0 in q to obtain

$$T_n = c_{1,0}^2 \Bigg(\frac{1}{n^2} \sum_{k=1}^n \sum_{i=1}^n \sum_{j=1}^n w(h_0^{-1}(S_k))\psi(Z_i, U_k)\psi(Z_j, U_k) \\ + \frac{2}{n^{3/2}} \sum_{k=1}^n \sum_{j=1}^n w(Y_k) r_0(Y_k)\psi(Z_j, U_k) - \frac{1}{n} \sum_{j=1}^n \sum_{k=1}^n \varphi(Z_j)^t \Gamma^{-1} \varphi(Z_k) \\ - \frac{2}{\sqrt{n}} \sum_{k=1}^n \varphi(Z_k)^t \Gamma^{-1} \beta + E[w(h_0^{-1}(S)) r_0(h_0^{-1}(S))^2] - \beta^t \Gamma^{-1} \beta \Bigg) + o_p(1)$$

and rewrite this expression to obtain a sum of U-statistics with kernels ζ and $\tilde{\zeta}$.

Step 3: Show that the operator K from the assertion is symmetric as well as continuous and apply the results of Witting and Müller-Funk (1995) followed by those of Lee (1990) to derive the limit distribution.

Before going into the details of each step, some previous thoughts have to be done. Remember equation (5.15)

$$h(\cdot) = \frac{\Lambda_{\theta_0}(\cdot) - \Lambda_{\theta_0}(0) + n^{-\frac{1}{2}}(r(y) - r(0))}{\Lambda_{\theta_0}(1) - \Lambda_{\theta_0}(0) + n^{-\frac{1}{2}}(r(1) - r(0))} = h_0(\cdot) + n^{-\frac{1}{2}} r_0(\cdot) + \mathcal{O}\left(\frac{1}{n}\right)$$

uniformly on compact sets, where

$$h_0(\cdot) = \frac{\Lambda_{\theta_0}(\cdot) - \Lambda_{\theta_0}(0)}{\Lambda_{\theta_0}(1) - \Lambda_{\theta_0}(0)} \quad \text{and} \quad r_0(\cdot) = \frac{r(\cdot) - r(0) - h_0(\cdot)(r(1) - r(0))}{\Lambda_{\theta_0}(1) - \Lambda_{\theta_0}(0)}.$$

Note that (due to the identification constraints) one has $h(0) = 0$ and $h(1) = 1$. Moreover, since h_0' is bounded away from zero on compact sets one has for an appropriate $\eta > 0$ and a random variable $\tilde{S}$

$$\begin{aligned} |Y - h_0^{-1}(S)| &= \left|h_0^{-1}(h_0(h^{-1}(S))) - h_0^{-1}(S)\right| \\ &= \underbrace{\left|\frac{1}{h_0'(\tilde{S})}\right|}_{<\eta} \left|h_0(h^{-1}(S)) - S\right| \\ &\leq \eta \frac{|r_0(h^{-1}(S))|}{\sqrt{n}} + \mathcal{O}\left(\frac{1}{n}\right), \end{aligned}$$

so that Y converges uniformly to $h_0^{-1}(S)$ on compact sets. Furthermore, assumption (**A7**) guarantees equation (5.13).

Step 1: Reduce the problem of minimizing T_n to that of the quadratic function

$$q(z) := \sum_{k=1}^n w(h_0^{-1}(S_k))(\hat{h}(Y_k)c_{1,0} + c_{2,0} - \Lambda_{\theta_0}(Y_k))^2 \\ + 2z^t \sum_{k=1}^n w(h_0^{-1}(S_k))(\hat{h}(Y_k)c_{1,0} + c_{2,0} - \Lambda_{\theta_0}(Y_k))R(S_k) + nz^t \Gamma z$$

Remember that $\tilde{\gamma}$ denotes the minimizer of M_n and γ_0 is the vector such that

$$h_0(y)c_{1,0} + c_{2,0} = \Lambda_{\theta_0}(y).$$

For all γ in a compact set that does not contain γ_0 and an appropriate $\varepsilon > 0$ one has

$$\begin{aligned}\frac{1}{n}\sum_{k=1}^{n} w(Y_k)(\hat{h}(Y_k)c_1 + c_2 - \Lambda_\theta(Y_k))^2 &= \frac{1}{n}\sum_{k=1}^{n} w(Y_k)(h(Y_k)c_1 + c_2 - \Lambda_\theta(Y_k))^2 + o_p(1)\\ &= \underbrace{E\big[w(Y)(h(Y)c_1 + c_2 - \Lambda_\theta(Y))^2\big]}_{\geq\varepsilon} + o_p(1)\end{aligned}$$

uniformly in $\gamma \in \Upsilon$, where the last equality can be shown by the same arguments as later in Theorem 5.2.6. Therefore,

$$\tilde{\gamma} - \gamma_0 = o_p(1). \tag{5.70}$$

Assumption (**A7**) ensures

$$\sup_{y\in\mathcal{Y}_w} |\hat{h}(y) - h(y)| = \mathcal{O}_p\Big(\frac{1}{\sqrt{n}}\Big)$$

and by the definition of $\tilde{\gamma}$ as the minimizer in (5.17), it holds that

$$\sum_{k=1}^{n} w(Y_k)(\hat{h}(Y_k)\tilde{c}_1 + \tilde{c}_2 - \Lambda_{\tilde{\theta}}(Y_k)) \begin{pmatrix} \hat{h}(Y_k) \\ 1 \\ -\dot{\Lambda}_{\tilde{\theta}}(Y_k)^t \end{pmatrix} = 0.$$

Consequently, a Taylor expansion with respect to γ yields

$$\begin{aligned}&\sum_{k=1}^{n} w(Y_k)(\hat{h}(Y_k)c_{1,0} + c_{2,0} - \Lambda_{\theta_0}(Y_k))^2\\ &= \sum_{k=1}^{n} w(Y_k)(\hat{h}(Y_k)\tilde{c}_1 + \tilde{c}_2 - \Lambda_{\tilde{\theta}}(Y_k))^2\\ &\quad + \sqrt{n}(\gamma_0 - \tilde{\gamma})^t \frac{2}{\sqrt{n}}\sum_{k=1}^{n} w(Y_k)(\hat{h}(Y_k)\tilde{c}_1 + \tilde{c}_2 - \Lambda_{\tilde{\theta}}(Y_k)) \begin{pmatrix} \hat{h}(Y_k) \\ 1 \\ -\dot{\Lambda}_{\tilde{\theta}}(Y_k)^t \end{pmatrix}\\ &\quad + \sqrt{n}(\tilde{\gamma} - \gamma_0)^t \frac{1}{n}\sum_{k=1}^{n} w(Y_k)\\ &\quad\begin{pmatrix} \hat{h}(Y_k)^2 & \hat{h}(Y_k) & -\hat{h}(Y_k)\dot{\Lambda}_{\tilde{\theta}}(Y_k) \\ \hat{h}(Y_k) & 1 & -\dot{\Lambda}_{\tilde{\theta}}(Y_k) \\ -\hat{h}(Y_k)\dot{\Lambda}_{\tilde{\theta}}(Y_k)^t & -\dot{\Lambda}_{\tilde{\theta}}(Y_k)^t & \dot{\Lambda}_{\tilde{\theta}}(Y_k)^t\dot{\Lambda}_{\tilde{\theta}}(Y_k) - \ddot{\Lambda}_{\tilde{\theta}}(Y_k)(\hat{h}(Y_k)\tilde{c}_1 + \tilde{c}_2 - \Lambda_{\tilde{\theta}}(Y_k)) \end{pmatrix}\\ &\quad\sqrt{n}(\tilde{\gamma} - \gamma_0) + o_p(n||\tilde{\gamma} - \gamma_0||^2)\\ &= \sum_{k=1}^{n} w(Y_k)(\hat{h}(Y_k)\tilde{c}_1 + \tilde{c}_2 - \Lambda_{\tilde{\theta}}(Y_k))^2\end{aligned}$$

$$+\sqrt{n}(\tilde{\gamma}-\gamma_0)^t\frac{1}{n}\sum_{k=1}^n w(h_0^{-1}(S_k))R(S_k)R(S_k)^t\sqrt{n}(\tilde{\gamma}-\gamma_0)+o_p(n||\tilde{\gamma}-\gamma_0||^2)$$

$$=\sum_{k=1}^n w(Y_k)(\hat{h}(Y_k)\tilde{c}_1+\tilde{c}_2-\Lambda_{\tilde{\theta}}(Y_k))^2+\sqrt{n}(\tilde{\gamma}-\gamma_0)^t\Gamma\sqrt{n}(\tilde{\gamma}-\gamma_0)+o_p(n||\tilde{\gamma}-\gamma_0||^2).$$

Since Γ was assumed to be positive definite and

$$\begin{aligned}
&\sum_{k=1}^n w(Y_k)(\hat{h}(Y_k)\tilde{c}_1+\tilde{c}_2-\Lambda_{\tilde{\theta}}(Y_k))^2\\
&\quad\leq\sum_{k=1}^n w(Y_k)(\hat{h}(Y_k)c_{1,0}+c_{2,0}-\Lambda_{\theta_0}(Y_k))^2\\
&\quad=\sum_{k=1}^n w(Y_k)\Big(c_{1,0}(\hat{h}(Y_k)-h(Y_k))+h(Y_k)c_{1,0}+c_{2,0}-\Lambda_{\theta_0}(Y_k)\Big)^2\\
&\quad\overset{(5.68)}{=}c_{1,0}^2\sum_{k=1}^n w(Y_k)\Big(\hat{h}(Y_k)-h(Y_k)+n^{-\frac{1}{2}}r_0(h_0^{-1}(S_k))\Big)^2+o_p(1)\\
&\quad=\mathcal{O}_p(1),
\end{aligned}$$

one has $||\tilde{\gamma}-\gamma_0||=\mathcal{O}_p\Big(\frac{1}{\sqrt{n}}\Big)$. This in turn can be used to write

$$\begin{aligned}
&\sum_{k=1}^n w(Y_k)(\hat{h}(Y_k)\tilde{c}_1+\tilde{c}_2-\Lambda_{\tilde{\theta}}(Y_k))^2\\
&\quad=\sum_{k=1}^n w(Y_k)(\hat{h}(Y_k)c_{1,0}+c_{2,0}-\Lambda_{\theta_0}(Y_k))^2+\sqrt{n}(\tilde{\gamma}-\gamma_0)^t\Gamma\sqrt{n}(\tilde{\gamma}-\gamma_0)\\
&\qquad+\sqrt{n}(\tilde{\gamma}-\gamma_0)^t\frac{2}{\sqrt{n}}\sum_{k=1}^n w(h_0^{-1}(S_k))(\hat{h}(Y_k)c_{1,0}+c_{2,0}-\Lambda_{\theta_0}(Y_k))R(S_k)+o_p(1)\\
&\quad=q(\tilde{\gamma}-\gamma_0)+o_p(1) \qquad (5.71)
\end{aligned}$$

again using a Taylor expansion with respect to γ. For all values γ with $\gamma-\gamma_0=\mathcal{O}_p\Big(\frac{1}{\sqrt{n}}\Big)$ these calculations can be done analogously with all $o_p(1)$-terms holding uniformly in γ (again similar reasoning to 5.2.6). Minimizing q leads to

$$\begin{aligned}
D_zq(z_0)&=2\sum_{k=1}^n w(h_0^{-1}(S_k))(\hat{h}(Y_k)c_{1,0}+c_{2,0}-\Lambda_{\theta_0}(Y_k))R(S_k)^t+2nz_0^t\Gamma\\
&=0\\
\Leftrightarrow\qquad z_0&=-\Gamma^{-1}\frac{1}{n}\sum_{k=1}^n w(h_0^{-1}(S_k))(\hat{h}(Y_k)c_{1,0}+c_{2,0}-\Lambda_{\theta_0}(Y_k))R(S_k).
\end{aligned}$$

Hence,

$$z_0:=\arg\min_z q(z)=\mathcal{O}_p\Big(\frac{1}{\sqrt{n}}\Big),$$

which means that $\gamma_0 + z_0$ can be plugged into equation (5.71) as well. In total,

$$\begin{aligned}
T_n &= \sum_{k=1}^n w(Y_k)(\hat{h}(Y_k)\tilde{c}_1 + \tilde{c}_2 - \Lambda_{\tilde{\theta}}(Y_k))^2 \\
&= q(\tilde{\gamma} - \gamma_0) + o_p(1) \\
&\geq \min_z q(z) + o_p(1) \\
&= q(z_0) + o_p(1) \\
&= \sum_{k=1}^n w(Y_k)\big(\hat{h}(Y_k)\big(c_{1,0} + z_{0,1}\big) + c_{2,0} + z_{0,2} - \Lambda_{\theta_0 + z_{0,3}}(Y_k)\big)^2 + o_p(1) \\
&\geq T_n + o_p(1).
\end{aligned}$$

Therefore, it is sufficient to consider $\min_z q(z)$ instead of T_n.

Step 2: Write z_0 as

$$\begin{aligned}
z_0 &= -\frac{c_{1,0}}{n}\sum_{k=1}^n w(h_0^{-1}(S_k))(\hat{h}(Y_k) - h(Y_k))\Gamma^{-1}R(S_k) \\
&\quad - \frac{c_{1,0}}{n^{\frac{3}{2}}}\sum_{k=1}^n w(h_0^{-1}(S_k))r_0(h_0^{-1}(S_k))\Gamma^{-1}R(S_k) + o_p\big(n^{-\frac{1}{2}}\big) \\
&= -c_{1,0}\Gamma^{-1}\frac{1}{n^2}\sum_{k=1}^n\sum_{j=1}^n w(h_0^{-1}(S_k))\psi(Z_j, U_k)R(S_k) - \frac{c_{1,0}}{\sqrt{n}}\Gamma^{-1}\beta + o_p\big(n^{-\frac{1}{2}}\big) \\
&= -\frac{c_{1,0}}{n}\Gamma^{-1}\sum_{k=1}^n \varphi(Z_k) - \frac{c_{1,0}}{\sqrt{n}}\Gamma^{-1}\beta + o_p\big(n^{-\frac{1}{2}}\big) \\
&= \mathcal{O}_p\Big(\frac{1}{\sqrt{n}}\Big),
\end{aligned}$$

where the second last equation can be shown by

$$E\Bigg[\Bigg\|\frac{1}{n}\sum_{j=1}^n\Bigg(\varphi(Z_j) - \frac{1}{n}\sum_{k=1}^n w(h_0^{-1}(S_k))\psi(Z_j, U_k)R(S_k)\Bigg)\Bigg\|^2\Bigg] = \mathcal{O}\Big(\frac{1}{n^2}\Big)$$

(similar calculations will be done later). Then, inserting z_0 leads to

$$\begin{aligned}
T_n &= q(z_0) + o_p(1) \\
&= \sum_{k=1}^n w(Y_k)(\hat{h}(Y_k)c_{1,0} + c_{2,0} - \Lambda_{\theta_0}(Y_k))^2 - nz_0^t\Gamma z_0 + o_p(1) \\
&= c_{1,0}^2\sum_{k=1}^n w(Y_k)(\hat{h}(Y_k) - h(Y_k))^2 + \frac{2c_{1,0}^2}{\sqrt{n}}\sum_{k=1}^n w(Y_k)r_0(Y_k)(\hat{h}(Y_k) - h(Y_k)) \\
&\quad + \frac{c_{1,0}^2}{n}\sum_{k=1}^n w(Y_k)r_0(Y_k)^2 - nz_0^t\Gamma z_0 + o_p(1)
\end{aligned}$$

$$
\begin{aligned}
&= c_{1,0}^2 \sum_{k=1}^n w(h_0^{-1}(S_k))(\hat{h}(Y_k) - h(Y_k))^2 \\
&\quad + \frac{2c_{1,0}^2}{\sqrt{n}} \sum_{k=1}^n w(h_0^{-1}(S_k)) r_0(h_0^{-1}(S_k))(\hat{h}(Y_k) - h(Y_k)) - \frac{c_{1,0}^2}{n} \sum_{j=1}^n \sum_{k=1}^n \varphi(Z_j)^t \Gamma^{-1} \varphi(Z_k) \\
&\quad - \frac{2c_{1,0}^2}{\sqrt{n}} \Gamma^{-1} \sum_{k=1}^n \varphi(Z_k)^t \beta + c_{1,0}^2 E[w(h_0^{-1}(S)) r_0(h_0^{-1}(S))^2] - c_{1,0}^2 \beta^t \Gamma^{-1} \beta + o_p(1) \\
&= A + B + C + D + E + F + o_p(1).
\end{aligned}
$$

Asymptotic treatment of A:
Due to assumption (**A7**) one has

$$
\hat{h}(y) - h(y) = \frac{1}{n} \sum_{j=1}^n \psi(Z_j, \mathcal{T}(y)) + o_p\big(n^{-\frac{1}{2}}\big) = \mathcal{O}_p\Big(\frac{1}{\sqrt{n}}\Big)
$$

uniformly in y. Therefore, with $U_i = \mathcal{T}(Y_i)$ independent of h

$$
\begin{aligned}
A &= c_{1,0}^2 \sum_{k=1}^n w(h_0^{-1}(S_k))(\hat{h}(Y_k) - h(Y_k))^2 \\
&= \frac{c_{1,0}^2}{n^2} \sum_{k=1}^n \sum_{i=1}^n \sum_{j=1}^n w(h_0^{-1}(S_k)) \psi(Z_i, U_k) \psi(Z_j, U_k) + o_p(1) \\
&= \frac{c_{1,0}^2}{n^2} \sum_{k=1}^n w(h_0^{-1}(S_k)) \psi(Z_k, U_k)^2 + \frac{2c_{1,0}^2}{n^2} \sum_{k=1}^n \sum_{\substack{i=1 \\ i \neq k}}^n w(h_0^{-1}(S_k)) \psi(Z_k, U_k) \psi(Z_i, U_k) \\
&\quad + \frac{c_{1,0}^2}{n^2} \sum_{k=1}^n \sum_{\substack{i=1 \\ i \neq k}}^n w(h_0^{-1}(S_k)) \psi(Z_i, U_k)^2 + \frac{c_{1,0}^2}{n^2} \sum_{k=1}^n \sum_{\substack{i=1 \\ i \neq k}}^n \sum_{\substack{j=1 \\ j \neq i,k}}^n w(h_0^{-1}(S_k)) \psi(Z_i, U_k) \psi(Z_j, U_k) \\
&\quad + o_p(1).
\end{aligned}
$$

Note that

$$
\operatorname{Var}\left(\frac{2c_{1,0}^2}{n^2} \sum_{k=1}^n \sum_{\substack{i=1 \\ i \neq k}}^n w(h_0^{-1}(S_k)) \psi(Z_k, U_k) \psi(Z_i, U_k) \right) \overset{n\to\infty}{\longrightarrow} 0
$$

as well as

$$
\operatorname{Var}\left(\frac{c_{1,0}^2}{n^2} \sum_{k=1}^n \sum_{\substack{i=1 \\ i \neq k}}^n w(h_0^{-1}(S_k)) \psi(Z_i, U_k)^2 \right) \overset{n\to\infty}{\longrightarrow} 0
$$

and

$$
E[w(h_0^{-1}(S_1)) \psi(Z_1, U_1) \psi(Z_2, U_1)] = E\big[[w(h_0^{-1}(S_1)) \psi(Z_1, U_1) \underbrace{E[\psi(Z_2, U_1) | U_1]}_{=0}\big] = 0, \tag{5.72}
$$

so that

$$A = c_{1,0}^2 E[w(h_0^{-1}(S_1))\psi(Z_2,U_1)^2] + \frac{c_{1,0}^2}{n^2}\sum_{k=1}^n \sum_{\substack{i=1\\ i\neq k}}^n \sum_{\substack{j=1\\ j\neq i,k}}^n w(h_0^{-1}(S_k))\psi(Z_i,U_k)\psi(Z_j,U_k) + o_p(1)$$

$$= c_{1,0}^2 E[w(h_0^{-1}(S_1))\psi(Z_2,U_1)^2] + \frac{2c_{1,0}^2}{n^2}\sum_{k=1}^n \sum_{i=k+1}^n \sum_{j=i+1}^n \tilde{\psi}_A(Z_i,Z_j,Z_k) + o_p(1),$$

where the (symmetric) kernel $\tilde{\psi}_A$ is defined as

$$\tilde{\psi}_A(z_1,z_2,z_3) = w(h_0^{-1}(S_1))\psi(z_2,U_1)\psi(z_3,U_1) + w(h_0^{-1}(S_2))\psi(z_1,U_2)\psi(z_3,U_2)$$
$$+ w(h_0^{-1}(S_3))\psi(z_1,U_3)\psi(z_2,U_3).$$

As in equation (5.72)

$$\psi_{A,1}(z) := E[\tilde{\psi}_A(Z_1,Z_2,Z_3)|Z_1 = z] \equiv 0$$

and

$$\begin{aligned}\psi_{A,2}(z_1,z_2) &:= E[\tilde{\psi}_A(Z_1,Z_2,Z_3)|Z_1 = z_1, Z_2 = z_2]\\ &= w(h_0^{-1}(s_1))\psi(z_2,u_1)\underbrace{E[\psi(Z_3,U_1)|Z_1 = z_1, Z_2 = z_2]}_{=0}\\ &\quad + w(h_0^{-1}(s_2))\psi(z_1,u_2)\underbrace{E[\psi(Z_3,U_2)|Z_1 = z_1, Z_2 = z_2]}_{=0}\\ &\quad + E[w(h_0^{-1}(S_3))\psi(Z_1,U_3)\psi(Z_2,U_3)|Z_1 = z_1, Z_2 = z_2]\\ &= E[w(h_0^{-1}(S_3))\psi(Z_1,U_3)\psi(Z_2,U_3)|Z_1 = z_1, Z_2 = z_2].\end{aligned}$$

For

$$\psi_{A,3}(z_1,z_2,z_3) := \tilde{\psi}_A(z_1,z_2,z_3) - \psi_{A,2}(z_1,z_2) - \psi_{A,2}(z_1,z_3) - \psi_{A,2}(z_2,z_3),$$

the Hoeffding decomposition (see Hoeffding (1948)) leads to

$$\frac{1}{\binom{n}{3}}\sum_{k=1}^n \sum_{i=k+1}^n \sum_{j=i+1}^n \tilde{\psi}_A(Z_i,Z_j,Z_k) = \frac{3}{\binom{n}{2}}\sum_{k=1}^n \sum_{i=k+1}^n \psi_{A,2}(Z_k,Z_i)$$
$$+ \frac{1}{\binom{n}{3}}\sum_{k=1}^n \sum_{i=k+1}^n \sum_{j=i+1}^n \psi_{A,3}(Z_i,Z_j,Z_k).$$

Since $\psi_{A,3}$ is a degenerated kernel of order 2, the variance of the last term can be written as

$$\begin{aligned}\operatorname{Var}\left(\frac{1}{\binom{n}{3}}\sum_{k=1}^n \sum_{i=k+1}^n \sum_{j=i+1}^n \psi_{A,3}(Z_i,Z_j,Z_k)\right) &= \frac{1}{\binom{n}{3}^2}\sum_{k=1}^n \sum_{i=k+1}^n \sum_{j=i+1}^n E\big[\psi_{A,3}(Z_i,Z_j,Z_k)^2\big]\\ &= \mathcal{O}\left(\frac{1}{n^3}\right).\end{aligned}$$

Hence,

$$A = c_{1,0}^2 E[w(h_0^{-1}(S_1))\psi(Z_2, U_1)^2] + nU_{A,n} + o_p(1),$$

where $U_{A,n}$ is a U-statistic with kernel $\psi_A := c_{1,0}^2 \psi_{A,2}$.

Asymptotic treatment of B:
One can easily see

$$\begin{aligned}
B &= \frac{2c_{1,0}^2}{\sqrt{n}} \sum_{k=1}^n w(h_0^{-1}(S_k)) r_0(h_0^{-1}(S_k))(\hat{h}(Y_k) - h(Y_k)) \\
&= \frac{c_{1,0}^2 \sqrt{n}}{\binom{n}{2}} \sum_{k=1}^n \sum_{i=k+1}^n (w(h_0^{-1}(S_k)) r_0(h_0^{-1}(S_k))\psi(Z_i, U_k) + w(h_0^{-1}(S_i)) r_0(h_0^{-1}(S_i))\psi(Z_k, U_i)) \\
&\quad + o_p(1).
\end{aligned}$$

Apply the Hoeffding decomposition to obtain

$$B = \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_B(Z_i) + o_p(1) = \sqrt{n} U_{B,n} + o_p(1),$$

where $U_{B,n}$ is a U-statistic with kernel

$$\psi_B(z) = 2c_{1,0}^2 E[w(h_0^{-1}(S_2)) r_0(h_0^{-1}(S_2))\psi(Z_1, U_2) | Z_1 = z].$$

Asymptotic treatment of C:
Term C can be written in the following way

$$\begin{aligned}
C &= -\frac{c_{1,0}^2}{n} \sum_{j=1}^n \sum_{k=1}^n \varphi(Z_j)^t \Gamma^{-1} \varphi(Z_k) \\
&= -\frac{c_{1,0}^2}{n} \sum_{k=1}^n \varphi(Z_k)^t \Gamma^{-1} \varphi(Z_k) - \frac{c_{1,0}^2}{n} \sum_{j=1}^n \sum_{\substack{k=1 \\ k \neq j}}^n \varphi(Z_j)^t \Gamma^{-1} \varphi(Z_k) \\
&= -c_{1,0}^2 E[\varphi(Z_1)^t \Gamma^{-1} \varphi(Z_1)] + nU_{C,n} + o_p(1),
\end{aligned}$$

where $U_{C,n}$ is a U-statistic with kernel $\psi_C(z_1, z_2) = -c_{1,0}^2 \varphi(z_1)^t \Gamma^{-1} \varphi(z_2)$.

Putting things together:
It already has been shown that

$$\begin{aligned}
T_n &= nU_{A,n} + \sqrt{n} U_{B,n} + nU_{C,n} + \sqrt{n} U_{D,n} + c_{1,0}^2 E[w(h_0^{-1}(S_1))\psi(Z_2, U_1)^2] \\
&\quad - c_{1,0}^2 E[\varphi(Z_1)^t \Gamma^{-1} \varphi(Z_1)] + c_{1,0}^2 E[w(h_0^{-1}(S)) r_0(h_0^{-1}(S))^2] - c_{1,0}^2 \beta^t \Gamma^{-1} \beta + o_p(1),
\end{aligned}$$

where $U_{D,n}$ is a U-statistic with kernel $\psi_D(z) = -2c_{1,0}^2 \Gamma^{-1} \varphi(z)^t \beta$.
For calculating the expectation of T_n, recall definitions (5.23) as well as (5.69) and note

$$\begin{aligned}
E[w(h_0^{-1}(S))\bar{r}(S)^2] &= E\left[w(h_0^{-1}(S)) \left(r_0(h_0^{-1}(S)) - \beta^t \Gamma^{-1} R(S)\right)^2\right] \\
&= E\left[w(h_0^{-1}(S)) r_0(h_0^{-1}(S))^2\right]
\end{aligned}$$

$$
\begin{aligned}
&- 2\beta^t\Gamma^{-1}E\left[w(h_0^{-1}(S))r_0(h_0^{-1}(S))R(S)\right] + \beta^t\Gamma^{-1}\beta^t \\
&= E\left[w(h_0^{-1}(S))r_0(h_0^{-1}(S))^2\right] - \beta^t\Gamma^{-1}\beta^t,
\end{aligned}
$$

so that

$$E + F = c_{1,0}^2 E[w(h_0^{-1}(S))\bar{r}(S)^2].$$

Consider

$$
\begin{aligned}
b_A &:= c_{1,0}^2 E\left[w(h_0^{-1}(S_2))\left(\psi(Z_1,U_2) + \bar{r}(S_2) - \varphi(Z_1)^t\Gamma^{-1}R(S_2)\right)^2\right] \\
&= c_{1,0}^2 E[\varphi(Z_1)^t\Gamma^{-1}\varphi(Z_1)] - 2c_{1,0}^2 E\left[\varphi(Z_1)^t\Gamma^{-1}R(S_2)(\psi(Z_1,U_2) + \bar{r}(S_2))w(h_0^{-1}(S_2))\right] \\
&\quad + c_{1,0}^2 E[(\psi(Z_1,U_2) + \bar{r}(S_2))^2 w(h_0^{-1}(S_2))] \\
&= c_{1,0}^2 E[\varphi(Z_1)^t\Gamma^{-1}\varphi(Z_1)] - 2c_{1,0}^2 E\left[\varphi(Z_1)^t\Gamma^{-1}\underbrace{E\left[\psi(Z_1,U_2)w(h_0^{-1}(S_2))R(S_2)\middle| Z_1\right]}_{=\varphi(Z_1)}\right] \\
&\quad - 2c_{1,0}^2 \underbrace{E[\varphi(Z_1)^t]}_{=0}\Gamma^{-1}E\left[\bar{r}(S_2)w(h_0^{-1}(S_2))R(S_2)\right] + c_{1,0}^2 E[\psi(Z_1,U_2)^2 w(h_0^{-1}(S_2))] \\
&\quad + 2c_{1,0}^2 \underbrace{E[\psi(Z_1,U_2)\bar{r}(S_2)w(h_0^{-1}(S_2))]}_{=0} + c_{1,0}^2 E[\bar{r}(S_2)^2 w(h_0^{-1}(S_2))] \\
&= c_{1,0}^2 E[\psi(Z_1,U_2)^2 w(h_0^{-1}(S_2))] + c_{1,0}^2 E[\bar{r}(S_2)^2 w(h_0^{-1}(S_2))] - c_{1,0}^2 E[\varphi(Z_1)^t\Gamma^{-1}\varphi(Z_1)],
\end{aligned}
$$

so that b_A is an expression for the asymptotic expectation of T_n.
Define a kernel ζ_A as

$$
\begin{aligned}
\zeta_A(z_1,z_2) &:= c_{1,0}^2 E\Bigg[w(h_0^{-1}(S_3))\left(\psi(Z_1,U_3) + \frac{1}{\sqrt{n}}r_0(h_0^{-1}(S_3)) - \varphi(Z_1)^t\Gamma^{-1}R(S_3)\right) \\
&\qquad \left(\psi(Z_2,U_3) + \frac{1}{\sqrt{n}}r_0(h_0^{-1}(S_3)) - \varphi(Z_2)^t\Gamma^{-1}R(S_3)\right)\Bigg| Z_1 = z_1, Z_2 = z_2\Bigg] \\
&\qquad - \frac{c_{1,0}^2}{n}E[w(h_0^{-1}(S))r_0(h_0^{-1}(S))^2] \\
&= c_{1,0}^2\varphi(z_1)^t\Gamma^{-1}\varphi(z_2) + E[w(h_0^{-1}(S_3))\psi(Z_1,U_3)\psi(Z_2,U_3)|Z_1 = z_1, Z_2 = z_2] \\
&\quad - c_{1,0}^2\varphi(z_1)^t\Gamma^{-1}\varphi(z_2) - c_{1,0}^2\varphi(z_2)^t\Gamma^{-1}\varphi(z_1) - \frac{c_{1,0}^2}{\sqrt{n}}\varphi(z_1)^t\Gamma^{-1}\beta - \frac{c_{1,0}^2}{\sqrt{n}}\varphi(z_2)^t\Gamma^{-1}\beta \\
&\quad + \frac{c_{1,0}^2}{\sqrt{n}}E[w(h_0^{-1}(S_3))r_0(h_0^{-1}(S_3))\psi(Z_1,U_3)|Z_1 = z_1] \\
&\quad + \frac{c_{1,0}^2}{\sqrt{n}}E[w(h_0^{-1}(S_3))r_0(h_0^{-1}(S_3))\psi(Z_2,U_3)|Z_2 = z_2] \\
&= \psi_A(z_1,z_2) + \frac{1}{2\sqrt{n}}(\psi_B(z_1) + \psi_B(z_2)) + \psi_C(z_1,z_2) + \frac{1}{2\sqrt{n}}(\psi_D(z_1) + \psi_D(z_2)).
\end{aligned}
$$

By similar calculations one has

$$c_{1,0}^2 \zeta(z_1, z_2) = \psi_A(z_1, z_2) + \psi_C(z_1, z_2)$$

as well as

$$c_{1,0}^2 \tilde{\zeta}(z) = \psi_B(z) + \psi_D(z),$$

so that

$$\frac{n}{\binom{n}{2}} \sum_{i=1}^{n} \sum_{j=i+1}^{n} \zeta_A(Z_i, Z_j) = nU_{A,n} = nU_n + \sqrt{n} V_{n,0} \tag{5.73}$$

and thus

$$T_n = c_{1,0}^2 (nU_n + b + \sqrt{n} V_{n,0} + E[w(h_0^{-1}(S_1)) \bar{r}(S_1)^2]).$$

Here, U_n and $V_{n,0}$ are U-statistics with kernel ζ and $\tilde{\zeta}$, respectively, and

$$b = E[w(h_0^{-1}(S_2)) \psi(Z_1, U_2)^2] - 2E[\varphi(Z_1)^t \Gamma^{-1} \varphi(Z_1)] = E[\zeta(Z_1, Z_1)] \tag{5.74}$$

In the next step these different representations will be used to derive the asymptotic law via some version of Mercer's theorem for positive semi-definite kernels.

Step 3: Show the stated weak convergence results.
One part of this step's reasoning consists in applying a result from functional analysis that ensures that ζ can be written as a sum of weighted orthonormal functions (see for example Witting and Müller-Funk (1995, p. 141)). Hence, positive semi-definiteness of ζ needs to be shown. This in turn directly follows from the representation $\zeta(z_1, z_2) = E[I(z_1) I(z_2)]$ (see Remark 5.2.3) with

$$I(z) := (\Lambda_{\theta_0}(1) - \Lambda_{\theta_0}(0))(w(h_0^{-1}(S_1)))^{1/2} \left(\psi(z, U_1) - \varphi(z)^t \Gamma^{-1} R(S_1) \right).$$

Referring to Witting and Müller-Funk (1995) ζ can be written as

$$\zeta(z_1, z_2) = \sum_{j \in \mathbb{N} \setminus \{0\}} \lambda_j \rho_j(z_1) \rho_j(z_2) \tag{5.75}$$

for an orthonormal basis $(\rho_j)_{j \in \mathbb{N}}$, where the convergence is meant in $\mathcal{L}^2$-sense. Although the asymptotic distributions of nU_n and $\sqrt{n} V_{n,0}$ in equation (5.73) can be obtained by applying Theorem 1 of Lee (1990, p. 79) and the central limit theorem, some arguments still have to be added since both summands in general depend on each other, that is, the asymptotic distribution can not be mimicked by simply adding two independent random variables, which follow the same distributions as the U statistics. This problem can be solved by some minor adjustments in the proof of Theorem 1 of Lee (1990, p. 79). There, to prove the statement

$$nU_n \xrightarrow{\mathcal{D}} \sum_{j \in \mathbb{N} \setminus \{0\}} \lambda_j (V_j^2 - 1),$$

the indices of the sum in (5.75), that exceed a natural number $K \in \mathbb{N}$ are cut off. Then, the approximation $\sum_{k=1}^{K} \lambda_k \rho_k(Z_i) \rho_k(Z_j) \approx \zeta(Z_i, Z_j)$ (for large K) is used to obtain

$$nU_n = \frac{1}{n-1} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \zeta(Z_i, Z_j) \approx \sum_{k=1}^{K} \lambda_k (V_{k,n}^2 - v_{k,n}),$$

where $V_{k,n} := n^{-1/2}\sum_{i=1}^n \rho_k(Z_i)$ and $v_{k,n} := n^{-1}\sum_{i=1}^n \rho_k^2(Z_i) = 1 + o_p(1)$ by the law of large numbers and the orthonormality of the eigenfunctions.

Now to obtain convergence of $nU_n + n^{1/2}V_{0,n}$, note that applying the multivariate central limit theorem, $(V_{0,n}, V_{1,n}, \dots, V_{K,n})^t$ converges in distribution to $(V_0, V_1, \dots, V_K)^t$ as defined in Theorem 5.2.2, for each K. Thus, the continuous mapping theorem yields

$$\sum_{k=1}^K \lambda_k(V_{k,n}^2 - v_{k,n}) + V_{0,n} \xrightarrow{\mathcal{D}} \sum_{k=1}^K \lambda_k(V_k^2 - 1) + V_0$$

for each K. Proceeding as in the proof of Theorem 1 of Lee (1990, p. 79) by letting $K \to \infty$, one obtains $\sum_{k=1}^\infty \lambda_k(V_k^2 - 1) + V_0$ as the limit of $nU_n + n^{1/2}V_{0,n}$. Note further that (5.75) especially leads to

$$\sum_{k=1}^\infty \lambda_k = \int \sum_{k=1}^\infty \lambda_k \rho_k(z_1)^2 \, dP^Z(z_1) = E[\zeta(Z_1, Z_1)] = b,$$

such that $nU_n + b + n^{1/2}V_{0,n}$ converges to $\sum_{k=1}^\infty \lambda_k V_k^2 + V_0$, which completes the proof of Theorem 5.2.2. □

5.8.2 Proof of Theorem 5.2.6

Consistency of the test proposed in (5.9) can be shown rather easily. Recall that h is a strictly increasing transformation function with $h(0) = 0$ and $h(1) = 1$, which can not be linearly transformed into a function $\Lambda_\theta \in \{\Lambda_\theta : \theta \in \Theta\}$. Since $\hat{h}$ is a uniformly consistent estimator of h, one has

$$\begin{aligned}\frac{1}{n}T_n &= \min_{c_1,c_2,\theta} \frac{1}{n}\sum_{k=1}^n w(Y_k)(\hat{h}(Y_k)c_1 + c_2 - \Lambda_\theta(Y_k))^2 \\ &= \min_{c_1,c_2,\theta} \frac{1}{n}\sum_{k=1}^n w(Y_k)(h(Y_k)c_1 + c_2 - \Lambda_\theta(Y_k))^2 + o_p(1).\end{aligned}$$

Note that the functions

$$f_\gamma(y) = w(y)(h(y)c_1 + c_2 - \Lambda_\theta(y))^2$$

are bounded, the parameter set $C_1 \times C_2 \times \Theta$ is compact and for every y the map

$$\gamma \mapsto w(y)(h(y)c_1 + c_2 - \Lambda_\theta(y))^2$$

is continuous. Hence, following Lemma 6.1 of Wellner (2005) the class $\mathcal{F} = \{f_\gamma : \gamma \in C_1 \times C_2 \times \Theta\}$ is Glivenko-Cantelli so that

$$\frac{1}{n}T_n = \min_{c_1,c_2,\theta} E[w(Y)(h(Y)c_1 + c_2 - \Lambda_\theta(Y))^2] + o_p(1) = c + o_p(1)$$

for some constant $c \in \mathbb{R}$. Moreover, the dominated convergence theorem implies continuity of the map

$$\gamma \mapsto E[w(Y)(h(Y)c_1 + c_2 - \Lambda_\theta(Y))^2].$$

Since the set over which the minimum is calculated is compact, one has $c > 0$. This in turn leads to

$$P(T_n > \varepsilon) = P\Big(\underbrace{\frac{T_n}{n}}_{\to c} > \underbrace{\frac{\varepsilon}{n}}_{\to 0}\Big) \overset{n\to\infty}{\longrightarrow} 1 \quad \text{for all } \varepsilon > 0.$$

Therefore, any test of the form (5.9) is consistent. □

5.8.3 Proof of Theorem 5.3.1

Again, the notations (5.29) and (5.66) will be used. Let $\tilde{\gamma}$ be the minimizer of M_n. Moreover, define

$$\tilde{M}_n(c_1, c_2, \theta^t)^t = \sum_{j=1}^{n} w(Y_j)(h(Y_j)c_1 + c_2 - \Lambda_\theta(Y_j))^2. \tag{5.76}$$

As before, one has

$$D_\gamma M_n(\gamma)\Big|_{\gamma=\tilde{\gamma}} = 0 \quad \text{and} \quad \frac{1}{n}\,\mathrm{Hess}\, M_n(\gamma_0) = \Gamma' + o_p(1). \tag{5.77}$$

The structure of the proof is as follows: At the beginning, it will be shown that

$$T_n = M_n(\tilde{\gamma}) = \sum_{k=1}^{n} w(Y_k)(\hat{h}(Y_k)c_{1,0} + c_{2,0} - \Lambda_{\theta_0}(Y_k))^2 + \mathcal{O}_p(\sqrt{n}). \tag{5.78}$$

From this, it will be deduced $\tilde{\gamma} - \gamma_0 = \mathcal{O}_p\big(n^{-\frac{1}{4}}\big)$, which in turn can be used to prove even

$$T_n = \sum_{k=1}^{n} w(Y_k)(\hat{h}(Y_k)c_{1,0} + c_{2,0} - \Lambda_{\theta_0}(Y_k))^2 + o_p(\sqrt{n}). \tag{5.79}$$

At last, the asymptotic distribution of $n^{-\frac{1}{2}}T_n$ is derived via (5.79).
First, one has similar to before $\tilde{\gamma} - \gamma_0 = o_p(1)$ (see (5.70)). Then, since Υ is compact,

$$\begin{aligned}
&M_n(\gamma) - \tilde{M}_n(\gamma)\\
&= \sum_{k=1}^{n} w(Y_k)(\hat{h}(Y_k)c_1 + c_2 - \Lambda_\theta(Y_k))^2 - \sum_{k=1}^{n} w(Y_k)(h(Y_k)c_1 + c_2 - \Lambda_\theta(Y_k))^2\\
&= \underbrace{c_1\sum_{k=1}^{n} w(Y_k)(\hat{h}(Y_k) - h(Y_k))^2}_{=\mathcal{O}_p(1)} + 2c_1\sum_{k=1}^{n} w(Y_k)\underbrace{(\hat{h}(Y_k) - h(Y_k))}_{=\mathcal{O}_p\left(\frac{1}{\sqrt{n}}\right)}\underbrace{(h(Y_k)c_1 + c_2 - \Lambda_\theta(Y_k))}_{=\mathcal{O}_p(1)}\\
&= \mathcal{O}_p(\sqrt{n})
\end{aligned} \tag{5.80}$$

uniformly in c_1, c_2, θ. Now, consider (h is fixed) the function class

$$\mathcal{F} = \{y \mapsto f_\gamma(y) = w(y)(h(y)c_1 + c_2 + \Lambda_\theta(y))^2 : \gamma \in \Upsilon\}$$

and the corresponding empirical process

$$Z_n(\gamma) = \sqrt{n}\Big(\frac{1}{n}\sum_{k=1}^{n} w(Y_k)(h(Y_k)c_1 + c_2 - \Lambda_\theta(Y_k))^2 - E\big[w(Y)(h(Y)c_1 + c_2 - \Lambda_\theta(Y))^2\big]\Big)$$

$$= \frac{1}{\sqrt{n}}(\tilde{M}_n(\gamma) - \underbrace{E[\tilde{M}_n(\gamma)]}_{=nM(\gamma)}).$$

Convergence of the marginal distributions can be shown by the Central Limit Theorem. Note that for $R(y) = (h(y), 1, -\dot{\Lambda}_{\theta_0}(y))^t$, all $\gamma_1, \gamma_2 \in \Upsilon$ and an appropriate γ^* between γ_1 and γ_2, $C \in \mathbb{R}$

$$\begin{aligned}
&\sup_{y \in \operatorname{supp}(w)} |f_{\gamma_1}(y) - f_{\gamma_2}(y)| \\
&= \sup_{y \in \operatorname{supp}(w)} \big|D_\gamma f_{\gamma^*}(y)(\gamma_1 - \gamma_2)\big| \\
&= ||\gamma_1 - \gamma_2|| \sup_{y \in \operatorname{supp}(w)} \big|\big|D_\gamma f_{\gamma^*}(y)\big|\big| \\
&\leq C||\gamma_1 - \gamma_2||.
\end{aligned}$$

Hence, due to the compactness of Υ the class $\mathcal{F}$ is Donsker so that $(Z_n)_{\gamma \in \Upsilon}$ converges weakly to some centred Gaussian process. Consequently,

$$\sup_{\gamma \in \Upsilon} |\tilde{M}_n(\gamma) - nM(\gamma)| = \mathcal{O}_p(n^{1/2}). \tag{5.81}$$

This in turn can be used to write

$$\begin{aligned}
M_n(\tilde{\gamma}) &\overset{(5.80)}{=} \inf_{\gamma \in \Upsilon} \tilde{M}_n(\gamma) + \mathcal{O}_p(\sqrt{n}) \\
&\leq \sup_{\gamma \in \Upsilon} |\tilde{M}_n(\gamma) - nM(\gamma)| + \inf_{\gamma \in \Upsilon} nM(\gamma) + \mathcal{O}_p(\sqrt{n}) \\
&\overset{(5.81)}{=} nM(\gamma_0) + \mathcal{O}_p(\sqrt{n}).
\end{aligned}$$

On the contrary,

$$\begin{aligned}
M_n(\tilde{\gamma}) &\overset{(5.80)}{=} \inf_{\gamma \in \Upsilon} \tilde{M}_n(\gamma) + \mathcal{O}_p(\sqrt{n}) \\
&\geq -\sup_{\gamma \in \Upsilon} |\tilde{M}_n(\gamma) - nM(\gamma)| + \inf_{\gamma \in \Upsilon} nM(\gamma) + \mathcal{O}_p(\sqrt{n}) \\
&\overset{(5.81)}{=} nM(\gamma_0) + \mathcal{O}_p(\sqrt{n}).
\end{aligned}$$

Consequently,

$$M_n(\tilde{\gamma}) - nM(\gamma_0) = \mathcal{O}_p(\sqrt{n})$$

and thus (due to (5.81)),

$$M_n(\tilde{\gamma}) - \tilde{M}_n(\gamma_0) = M_n(\tilde{\gamma}) - nM(\gamma_0) + nM(\gamma_0) - \tilde{M}_n(\gamma_0) = \mathcal{O}_p(\sqrt{n}).$$

Now, write

$$\tilde{M}_n(\gamma_0) - M_n(\tilde{\gamma}) = \mathcal{O}_p(\sqrt{n})$$

and similar to the sections before for some γ^* between $\tilde{\gamma}$ and γ_0 (note that equation (5.78) is obtained in the second row)

$$\tilde{M}_n(\gamma_0) - M_n(\tilde{\gamma})$$

$$
\begin{aligned}
&= M_n(\gamma_0) - M_n(\tilde{\gamma}) + \mathcal{O}_p(\sqrt{n}) \\
&= \underbrace{D_\gamma M_n(\gamma)\Big|_{\gamma=\tilde{\gamma}}}_{\overset{(5.77)}{=}0}(\gamma_0 - \tilde{\gamma}) + (\gamma_0 - \tilde{\gamma})^t \,\mathrm{Hess}\, M_n(\gamma^*)(\gamma_0 - \tilde{\gamma}) + \mathcal{O}_p(\sqrt{n}) \\
&= \sqrt{n}\, n^{\frac{1}{4}}(\gamma_0 - \tilde{\gamma})^t \frac{1}{n} \,\mathrm{Hess}\, M_n(\gamma_0) n^{\frac{1}{4}}(\gamma_0 - \tilde{\gamma}) + o_p(n||\gamma_0 - \tilde{\gamma}||^2) + \mathcal{O}_p(\sqrt{n}) \\
&\overset{(5.77)}{=} \sqrt{n}\, n^{\frac{1}{4}}(\gamma_0 - \tilde{\gamma})^t \Gamma' n^{\frac{1}{4}}(\gamma_0 - \tilde{\gamma}) + \mathcal{O}_p(\sqrt{n}),
\end{aligned}
$$

so that due to the positive definiteness of Γ' one has $\tilde{\gamma} - \gamma_0 = \mathcal{O}_p\big(n^{-\frac{1}{4}}\big)$.
Similar to the proof of Theorem 5.2.2, define

$$
q(z) := M_n(\gamma_0) + D_\gamma M_n(\gamma)\Big|_{\gamma=\gamma_0} z + n z^t \Gamma' z.
$$

Then, for all $\bar{\gamma}$ with $\bar{\gamma} - \gamma_0 = \mathcal{O}_p\big(n^{-\frac{1}{4}}\big)$ one has

$$
\begin{aligned}
M_n(\bar{\gamma}) &= M_n(\gamma_0) + D_\gamma M_n(\gamma)\Big|_{\gamma=\gamma_0}(\bar{\gamma} - \gamma_0) + (\bar{\gamma} - \gamma_0)^t \,\mathrm{Hess}\, M_n(\gamma_0)(\bar{\gamma} - \gamma_0) + o_p(\sqrt{n}) \\
&= q(\bar{\gamma} - \gamma_0) + o_p(\sqrt{n}). \qquad (5.82)
\end{aligned}
$$

The minimizer z_0 of q can be written as

$$
\begin{aligned}
z_0 &= -\Gamma'^{-1} \frac{1}{n}\Big(D_\gamma M_n(\gamma)\Big|_{\gamma=\gamma_0}\Big)^t \\
&= -\Gamma'^{-1} \frac{1}{n} \sum_{k=1}^n w(Y_k)(\hat{h}(Y_k)c_{1,0} + c_{2,0} - \Lambda_{\theta_0}(Y_k)) \begin{pmatrix} \hat{h}(Y_k) \\ 1 \\ -\dot{\Lambda}_{\theta_0}(Y_k)^t \end{pmatrix} \\
&= -\Gamma'^{-1} \underbrace{\frac{1}{n} \sum_{k=1}^n w(Y_k)(h(Y_k)c_{1,0} + c_{2,0} - \Lambda_{\theta_0}(Y_k))R(Y_k)}_{E[\cdot]=0} + \mathcal{O}_p\Big(\frac{1}{\sqrt{n}}\Big) \\
&= \mathcal{O}_p\Big(\frac{1}{\sqrt{n}}\Big),
\end{aligned}
$$

that is, (5.82) yields $M_n(\gamma_0 + z_0) = q(z_0) + o_p(\sqrt{n})$. Note that the last equality above holds because γ_0 is the minimizer of $M(\gamma)$ and hence $D_\gamma M(\gamma)|_{\gamma=\gamma_0} = 0$, that is,

$$
E[w(Y_1)(h(Y_1)c_{1,0} + c_{2,0} - \Lambda_{\theta_0}(Y_1))R(Y_1)] = 0.
$$

So far, it has been proven

$$
T_n = q(\tilde{\gamma} - \gamma_0) + o_p(\sqrt{n}) \geq q(z_0) + o_p(\sqrt{n}) = M_n(\gamma_0 + z_0) + o_p(\sqrt{n}) \geq T_n + o_p(\sqrt{n}),
$$

so that to prove (5.32) it is sufficient to minimize q and replace T_n by $q(z_0)$ in equation (5.32). Therefore, it remains to look at $q(z_0)$ in detail and derive the corresponding asymptotic distribution. Inserting z_0 into q leads to

$$
q(z_0) = M_n(\gamma_0) - n z_0^t \Gamma' z_0
$$

$$= \underbrace{c_{1,0}^2 \sum_{k=1}^n w(Y_k)(\hat{h}(Y_k) - h(Y_k))^2}_{=\mathcal{O}_p(1)} + \tilde{M}_n(\gamma_0)$$

$$+ 2c_{1,0} \sum_{k=1}^n w(Y_k) \quad \underbrace{(\hat{h}(Y_k) - h(Y_k))}_{=\frac{1}{n}\sum_{j=1}^n \psi(Z_j,U_k) + o_p(\frac{1}{\sqrt{n}})} \quad (h(Y_k)c_{1,0} + c_{2,0} - \Lambda_{\theta_0}(Y_k)) + o_p(\sqrt{n})$$

$$= \tilde{M}_n(\gamma_0) + \frac{2c_{1,0}}{n} \sum_{j=1}^n \sum_{k=1}^n w(Y_k)\psi(Z_j, U_k)\tilde{R}_k + o_p(\sqrt{n}).$$

In total,

$$\begin{aligned} \frac{T_n - nM(\gamma_0)}{\sqrt{n}} &= \frac{q(z_0) - nM(\gamma_0)}{\sqrt{n}} + o_p(1) \\ &= Z_n(\gamma_0) + \frac{2\sqrt{n}}{n^2} \sum_{j=1}^n \sum_{\substack{k=1 \\ k\neq j}}^n w(Y_k)\psi(Z_j, U_k)\tilde{R}_k + o_p(1). \end{aligned}$$

The last term can be treated as before by applying the Hoeffding decomposition so that

$$\frac{T_n - nM(\gamma_0)}{\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{k=1}^n \big(f_{\gamma_0}(S_k) - E[f_{\gamma_0}(S)] + \delta(Z_k)\big) + o_p\Big(\frac{1}{\sqrt{n}}\Big)$$

with the centred kernel δ from the assertion. The assertion itself now follows from the ordinary Central Limit Theorem. □

5.8.4 Proof of Lemma 5.3.2

For all $n, m \in \mathbb{N}, m \le n$ the random variables $H_{m,1}, ..., H_{m,q}$ are independent and identically distributed, so that

$$\begin{aligned} P\big(\max_{\nu\in\{1,...,q\}} |H_{m,\nu}| > \varepsilon\big) &= 1 - P\big(\max_{\nu\in\{1,...,q\}} |H_{m,\nu}| \le \varepsilon\big) \\ &= 1 - \prod_{\nu=1}^q P(|H_{m,\nu}| \le \varepsilon) \\ &= 1 - P(|H_{m,1}| \le \varepsilon)^q \\ &= 1 - \big(1 - P(|H_{m,1}| > \varepsilon)\big)^q. \end{aligned}$$

Due to $\big(1 - \frac{c}{q}\big)^q \overset{q\to\infty}{\longrightarrow} \exp(-c)$ for all $c \in \mathbb{R}$, one has

$$P\big(\max_{\nu\in\{1,...,q\}} |H_{m,\nu}| > \varepsilon\big) = o(1) \quad \Leftrightarrow \quad P(|H_{m,1}| > \varepsilon) = o\Big(\frac{1}{q}\Big). \tag{5.83}$$

Remember $q = \lceil \frac{n}{m} - 1 \rceil$ and define

$$\tilde{m}_n := \min\left\{k \in \{1, ..., n\} : \Big\lceil \frac{n}{k} - 1 \Big\rceil \sup_{n\ge l\ge k} P(|H_{l,1}| > \varepsilon) \le \varepsilon\right\}.$$

Let $\xi \in (0,1)$ be arbitrarily small. Then,

$$\left\lceil \frac{n}{\lceil \xi n \rceil} - 1 \right\rceil \sup_{n \geq l \geq \lceil \xi n \rceil} P(|H_{l,1}| > \varepsilon) \to 0$$

for $n \to \infty$. Consequently, $\frac{\tilde{m}_n}{n} \leq \xi$ for all $n \geq N$ and a sufficiently large $N \in \mathbb{N}$. Since $\xi \in (0,1)$ can be chosen arbitrarily small, one has $\frac{\tilde{m}_n}{n} \to 0$ for $n \to \infty$, so that

$$m_n := \lceil \sqrt{n \tilde{m}_n} \rceil \geq \tilde{m}_n$$

is an intermediate sequence. Moreover,

$$\begin{aligned}
\left\lceil \frac{n}{m_n} - 1 \right\rceil P(|H_{m_n,1}| > \varepsilon) &\leq \left\lceil \frac{n}{m_n} - 1 \right\rceil \sup_{l \geq \tilde{m}_n} P(|H_{l,1}| > \varepsilon) \\
&= \underbrace{\frac{\lceil \frac{n}{m_n} - 1 \rceil}{\lceil \frac{n}{\tilde{m}_n} - 1 \rceil}}_{\to 0} \underbrace{\left\lceil \frac{n}{\tilde{m}_n} - 1 \right\rceil \sup_{l \geq \tilde{m}_n} P(|H_{\tilde{m}_n,1}| > \varepsilon)}_{\leq \varepsilon} \\
&\to 0
\end{aligned}$$

for $n \to \infty$. Hence, $\max_{\nu=1,\ldots,q} H_{m,\nu} = o_p(1)$ follows from (5.83), which together with Lemma 1.5.1 implies (5.34).

Proof of consistency: In the following, m is written instead of m_n. Note that

$$\begin{aligned}
&E\left[w(Y_1)(h(Y_1)c_{0,1} + c_{0,2} - \Lambda_{\theta_0}(Y_1)) \begin{pmatrix} h(Y_1) \\ 1 \\ -\dot{\Lambda}_{\theta_0}(Y_1)^t \end{pmatrix} \right] \\
&= D_\gamma E\big[w(Y_1)(h(Y_1)c_{0,1} + c_{0,2} - \Lambda_{\theta_0}(Y_1))^2\big]\Big|_{\gamma=\gamma_0} \\
&= 0, \qquad (5.84)
\end{aligned}$$

so that a Taylor expansion and (**A5'**) lead to

$$\begin{aligned}
&\frac{\sqrt{m}}{n} \sum_{i=1}^n w(Y_i)(\hat{h}(Y_i)\tilde{c}_1 + \tilde{c}_2 - \Lambda_{\tilde{\theta}}(Y_i))^2 \\
&= \frac{\sqrt{m}}{n} \sum_{i=1}^n w(Y_i)(h(Y_i)\tilde{c}_1 + \tilde{c}_2 - \Lambda_{\tilde{\theta}}(Y_i))^2 + \mathcal{O}_p\Big(\sqrt{\frac{m}{n}}\Big) \\
&= \frac{\sqrt{m}}{n} \sum_{i=1}^n w(Y_i)(h(Y_i)c_{0,1} + c_{0,2} - \Lambda_{\theta_0}(Y_i))^2 \\
&\quad + (\tilde{\gamma} - \gamma_0)^t \frac{2\sqrt{m}}{n} \underbrace{\sum_{i=1}^n w(Y_i)(h(Y_i)c_{0,1} + c_{0,2} - \Lambda_{\theta_0}(Y_i)) \begin{pmatrix} h(Y_i) \\ 1 \\ -\dot{\Lambda}_{\theta_0}(Y_i)^t \end{pmatrix}}_{=\mathcal{O}_p(\sqrt{n})}
\end{aligned}$$

$$
\begin{aligned}
&+\mathcal{O}_p\Big(\sqrt{m}||\tilde{\gamma}-\gamma_0||^2\Big)+\mathcal{O}_p\Big(\sqrt{\frac{m}{n}}\Big)\\
&=\frac{\sqrt{m}}{n}\sum_{i=1}^n w(Y_i)\tilde{R}_i^2+\mathcal{O}_p\Big(\sqrt{\frac{m}{n}}\Big)\\
&=\sqrt{m}E[w(Y_1)\tilde{R}_1^2]+\mathcal{O}_p\Big(\sqrt{\frac{m}{n}}\Big)
\end{aligned}
$$

with $\tilde{R}_1$ as in (5.31). Consequently, $\hat{\sigma}^2$ can be written as

$$
\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{q}\sum_{\nu=1}^q\Big(\frac{2\sqrt{m}}{n}\sum_{k=1}^n w(Y_k)\big(\hat{h}^{(\nu)}(Y_k)-\hat{h}(Y_k)\big)\big(\hat{h}(Y_k)\tilde{c}_1+\tilde{c}_2-\Lambda_{\tilde{\theta}}(Y_k)\big)\\
&+\frac{1}{\sqrt{m}}\sum_{j=(\nu-1)m+1}^{\nu m}\Big(w(Y_j)(\hat{h}(Y_j)\tilde{c}_1+\tilde{c}_2-\Lambda_{\tilde{\theta}}(Y_j))^2-E[w(Y_1)\tilde{R}_1^2]\Big)+\mathcal{O}_p\Big(\sqrt{\frac{m}{n}}\Big)\Big)^2\\
&=\frac{1}{q}\sum_{\nu=1}^q\Big(\frac{2\sqrt{m}}{n}\sum_{k=1}^n w(Y_k)\big(\hat{h}^{(\nu)}(Y_k)-h(Y_k)\big)\big(h(Y_k)\tilde{c}_1+\tilde{c}_2-\Lambda_{\tilde{\theta}}(Y_k)\big)\\
&+\frac{1}{\sqrt{m}}\sum_{j=(\nu-1)m+1}^{\nu m}\Big(w(Y_j)(\hat{h}(Y_j)\tilde{c}_1+\tilde{c}_2-\Lambda_{\tilde{\theta}}(Y_j))^2-E[w(Y_1)\tilde{R}_1^2]\Big)+\mathcal{O}_p\Big(\sqrt{\frac{m}{n}}\Big)\Big)^2.\\
&=\frac{1}{q}\sum_{\nu=1}^q\Big(\frac{2}{n}\sum_{k=1}^n\frac{1}{\sqrt{m}}\sum_{j=(\nu-1)m+1}^{\nu m} w(Y_k)\psi(Z_j,U_k)(h(Y_k)\tilde{c}_1+\tilde{c}_2-\Lambda_{\tilde{\theta}}(Y_k))\\
&+\frac{1}{\sqrt{m}}\sum_{j=(\nu-1)m+1}^{\nu m}\Big(w(Y_j)(\hat{h}(Y_j)\tilde{c}_1+\tilde{c}_2-\Lambda_{\tilde{\theta}}(Y_j))^2-E[w(Y_1)\tilde{R}_1^2]\Big)\\
&+\mathcal{O}_p\Big(\sqrt{\frac{m}{n}}+a_n\Big)\Big)^2
\end{aligned}
$$

with a_n from (5.34), where $\hat{h}$ was replaced by h in the second equality and equation (5.33) was applied to obtain the last equality. In the following, it will be shown that each of the occurring terms is of order $\mathcal{O}_p(1)$, so that the convergence rate can be shifted to the outside of the brackets by applying the Cauchy-Schwarz inequality. Hence,

$$
\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{q}\sum_{\nu=1}^q\Big(\frac{2}{n}\sum_{k=1}^n\frac{1}{\sqrt{m}}\sum_{j=(\nu-1)m+1}^{\nu m} w(Y_k)\psi(Z_j,U_k)(h(Y_k)\tilde{c}_1+\tilde{c}_2-\Lambda_{\tilde{\theta}}(Y_k))\\
&+\frac{1}{\sqrt{m}}\sum_{j=(\nu-1)m+1}^{\nu m}\Big(w(Y_j)(\hat{h}(Y_j)\tilde{c}_1+\tilde{c}_2-\Lambda_{\tilde{\theta}}(Y_j))^2-E[w(Y_1)\tilde{R}_1^2]\Big)\Big)^2\\
&+\mathcal{O}_p\Big(\sqrt{\frac{m}{n}}+a_n\Big)\\
&=\frac{1}{q}\sum_{\nu=1}^q\Big(\frac{1}{\sqrt{m}}\sum_{j=(\nu-1)m+1}^{\nu m}(A_{n,j}+B_{n,j}+C_{n,j}+D_j)\Big)^2+\mathcal{O}_p\Big(\sqrt{\frac{m}{n}}+a_n\Big)
\end{aligned}
$$

where

$$\begin{aligned}
A_{n,j} &= \frac{2}{n}\sum_{k=1}^{n} w(Y_k)\psi(Z_j,U_k)(h(Y_k)\tilde{c}_1+\tilde{c}_2-\Lambda_{\tilde{\theta}}(Y_k)-\tilde{R}_k)\\
&=: \frac{2}{n}\sum_{k=1}^{n} A_{n,j,k},\\
B_{n,j} &= \frac{1}{n}\sum_{k=1}^{n}(2w(Y_k)\psi(Z_j,U_k)\tilde{R}_k-\delta(Z_j))\\
&=: \frac{1}{n}\sum_{k=1}^{n} B_{n,j,k},\\
C_{n,j} &= w(Y_j)\big((\hat{h}(Y_j)\tilde{c}_1+\tilde{c}_2-\Lambda_{\tilde{\theta}}(Y_1))^2-\tilde{R}_j^2\big),\\
D_j &= \delta(Z_j)+w(Y_j)\tilde{R}_j^2-E\big[w(Y_j)\tilde{R}_1^2\big].
\end{aligned}$$

In the following, asymptotic negligibility of the terms corresponding to the $A_{n,j}, B_{n,j}$ and $C_{n,j}$ is proven and the convergence of the sum corresponding to D_j to σ^2 is developed. Write

$$A_{n,i,k} = w(Y_k)\psi(Z_i,U_k)(h(Y_k)(\tilde{c}_1-c_{1,0})+\tilde{c}_2-c_{2,0}-\Lambda_{\tilde{\theta}}(Y_k)+\Lambda_{\theta_0}(Y_k))$$

and define

$$\tilde{A}_{i,k} = w(Y_k)\psi(Z_i,U_k)\begin{pmatrix} h(Y_k)\\ 1\\ -\dot{\Lambda}_{\theta_0}(Y_k)^t\end{pmatrix}.$$

Then, a Taylor expansion with respect to θ yields

$$\begin{aligned}
&\frac{1}{q}\sum_{\nu=1}^{q}\Bigg(\frac{1}{\sqrt{m}}\sum_{i=(\nu-1)m+1}^{\nu m} A_{n,i}\Bigg)^2\\
&= \frac{2}{q}\sum_{\nu=1}^{q}\Bigg(\frac{2}{n\sqrt{m}}\sum_{i=(\nu-1)m+1}^{\nu m}\sum_{k=1}^{n} A_{n,i,k}\Bigg)^2\\
&= \frac{8}{qmn^2}\sum_{\nu=1}^{q}\Bigg(\sum_{i=(\nu-1)m+1}^{\nu m}(\tilde{\gamma}-\gamma_0)^t\sum_{k=1}^{n} w(Y_k)\psi(Z_i,U_k)\begin{pmatrix} h(Y_k)\\ 1\\ -\dot{\Lambda}_{\theta_0}(Y_k)^t\end{pmatrix}+\mathcal{O}_p(m\sqrt{n})\Bigg)^2\\
&= \frac{8}{qmn^2}\sum_{\nu=1}^{q}\Bigg((\tilde{\gamma}-\gamma_0)^t\sum_{i=(\nu-1)m+1}^{\nu m}\sum_{k=1}^{n}\tilde{A}_{i,k}+\mathcal{O}_p(m\sqrt{n})\Bigg)^2\\
&\le (\tilde{\gamma}-\gamma_0)^t\frac{16}{qmn^2}\sum_{\nu=1}^{q}\sum_{k=1}^{n}\sum_{l=1}^{n}\sum_{i=(\nu-1)m+1}^{\nu m}\sum_{j=(\nu-1)m+1}^{\nu m}\tilde{A}_{i,k}\tilde{A}_{j,l}^t(\tilde{\gamma}-\gamma_0)+\mathcal{O}_p\Big(\frac{m}{n}\Big).
\end{aligned}$$

Note that ("·" is meant componentwise) due to (**A7'**) one has

$$E[(\tilde{A}_{i,k}\tilde{A}_{j,l}^t)\cdot(\tilde{A}_{s,u}\tilde{A}_{t,v}^t)] \neq 0$$

only if none of the indices i, j, s, t is occurring only once in (i, k, j, l, s, u, t, v). Hence,

$$E\left[\left(\frac{16}{qmn^2}\sum_{\nu=1}^{q}\sum_{k=1}^{n}\sum_{l=1}^{n}\sum_{i=(\nu-1)m+1}^{\nu m}\sum_{j=(\nu-1)m+1}^{\nu m}\tilde{A}_{i,k}\tilde{A}_{j,l}^t\right)^2\right] < \infty \tag{5.85}$$

and thus

$$\frac{1}{q}\sum_{\nu=1}^{q}\left(\frac{1}{\sqrt{m}}\sum_{j=(\nu-1)m+1}^{\nu m}A_{n,j}\right)^2 = \mathcal{O}_p\left(\frac{m}{n}+n^{-\frac{1}{2}}\right)$$

due to $\tilde{\gamma}-\gamma_0 = \mathcal{O}_p\big(n^{-\frac{1}{4}}\big)$. The $C_{n,j}$ can be treated similarly, since

$$\begin{aligned}
C_{n,j} &= w(Y_j)\big((\hat{h}(Y_j)\tilde{c}_1+\tilde{c}_2-\Lambda_{\tilde{\theta}}(Y_j))^2-(h(Y_j)\tilde{c}_1+\tilde{c}_2-\Lambda_{\tilde{\theta}}(Y_j))^2\big)\\
&\quad + w(Y_j)\big((h(Y_j)\tilde{c}_1+\tilde{c}_2-\Lambda_{\tilde{\theta}}(Y_j))^2-\tilde{R}_j^2\big)\\
&= C_{n,j}^{(1)}+C_{n,j}^{(2)}.
\end{aligned}$$

Therefore, $\sup_{y\in\mathrm{supp}(w)}|\hat{h}(y)-h(y)| = \mathcal{O}_p\big(n^{-\frac{1}{2}}\big)$ and a Taylor expansion with respect to γ lead to

$$\begin{aligned}
&\frac{1}{q}\sum_{r=1}^{q}\left(\frac{1}{\sqrt{m}}\sum_{i=(\nu-1)m+1}^{\nu m}C_{n,i}\right)^2\\
&\leq \frac{2}{q}\sum_{\nu=1}^{q}\left(\frac{1}{\sqrt{m}}\sum_{i=(\nu-1)m+1}^{\nu m}C_{n,i}^{(1)}\right)^2+\frac{2}{q}\sum_{\nu=1}^{q}\left(\frac{1}{\sqrt{m}}\sum_{i=(\nu-1)m+1}^{\nu m}C_{n,i}^{(2)}\right)^2\\
&= \frac{2}{q}\sum_{\nu=1}^{q}\left(\frac{1}{\sqrt{m}}\sum_{i=(\nu-1)m+1}^{\nu m}w(Y_i)\big(\tilde{c}_1^2(\hat{h}(Y_i)-h(Y_i))^2\right.\\
&\quad\left.+2\tilde{c}_1(\hat{h}(Y_i)-h(Y_i))(h(Y_i)\tilde{c}_1+\tilde{c}_2-\Lambda_{\tilde{\theta}}(Y_i))\big)\right)^2+\frac{2}{q}\sum_{\nu=1}^{q}\left(\frac{1}{\sqrt{m}}\sum_{i=(\nu-1)m+1}^{\nu m}C_{n,i}^{(2)}\right)^2\\
&= \frac{2}{q}\sum_{\nu=1}^{q}\left(\frac{1}{\sqrt{m}}\sum_{i=(\nu-1)m+1}^{\nu m}w(Y_j)\big((h(Y_i)\tilde{c}_1+\tilde{c}_2-\Lambda_{\tilde{\theta}}(Y_i))^2-\tilde{R}_i^2\big)\right)^2+\mathcal{O}_p\left(\frac{m}{n}\right)\\
&= \frac{2}{q}\sum_{\nu=1}^{q}\left((\tilde{\gamma}-\gamma_0)^t\frac{1}{\sqrt{m}}\sum_{i=(\nu-1)m+1}^{\nu m}\tilde{R}_i\begin{pmatrix}h(Y_i)\\1\\-\dot{\Lambda}_{\theta_0}(Y_i)^t\end{pmatrix}\right.\\
&\quad\left.+\mathcal{O}_p\big(\sqrt{m}||\tilde{\gamma}-\gamma_0||^2\big)\right)^2+\mathcal{O}_p\left(\frac{m}{n}\right)\\
&\leq (\tilde{\gamma}-\gamma_0)^t\frac{4}{q}\sum_{\nu}^{q}\frac{1}{m}\sum_{i=(\nu-1)m+1}^{\nu m}\sum_{j=(\nu-1)m+1}^{\nu m}\tilde{R}_i\tilde{R}_j\begin{pmatrix}h(Y_i)\\1\\-\dot{\Lambda}_{\theta_0}(Y_i)^t\end{pmatrix}\\
&\quad\begin{pmatrix}h(Y_j) & 1 & -\dot{\Lambda}_{\theta_0}(Y_j)\end{pmatrix}(\tilde{\gamma}-\gamma_0)+\mathcal{O}_p\left(\frac{m}{n}\right)
\end{aligned}$$

$$= \mathcal{O}_p\left(\frac{m}{n} + n^{-\frac{1}{2}}\right),$$

where the last equality follows by applying (5.84) and a similar reasoning to obtaining (5.85).

Because the data are independent and identically distributed, the term corresponding to the $B_{n,j}$ can be written as

$$\frac{n}{m}E\left[\frac{1}{q}\sum_{\nu=1}^{q}\left(\frac{1}{\sqrt{m}}\sum_{j=(\nu-1)m+1}^{\nu m}B_{n,j}\right)^2\right]$$

$$= \frac{n}{m^2}E\left[\left(\sum_{j=1}^{m}B_{n,j}\right)^2\right]$$

$$= \frac{1}{m^2 n}\sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{n}\sum_{l=1}^{n}\underbrace{E[B_{n,i,k}B_{n,j,l}]}_{=0,\text{ if } i\notin\{j,k,l\} \text{ or } j\notin\{i,k,l\}}$$

$$= \frac{1}{m^2 n}\sum_{i=1}^{m}\sum_{k=1}^{n}\sum_{l=1}^{n}E[B_{n,i,k}B_{n,i,l}] + \frac{1}{m^2 n}\sum_{i=1}^{m}\sum_{\substack{j=1\\ j\neq i}}^{m}E[B_{n,i,i}B_{n,j,j}]$$

$$+ \frac{1}{m^2 n}\sum_{i=1}^{m}\sum_{\substack{j=1\\ j\neq i}}^{m}E[B_{n,j,i}B_{n,i,j}]$$

$$= \frac{(n-1)(n-2)}{mn}E[B_{n,1,2}B_{n,1,3}] + o(1)$$

$$= \frac{(n-1)(n-2)}{mn}\big(E[4w(Y_2)\psi(Z_1,U_2)\tilde{R}_2 w(Y_3)\psi(Z_1,U_3)\tilde{R}_3] - 2E[\delta(Z_1)w(Y_2)\psi(Z_1,U_2)\tilde{R}_2]$$

$$- 2E[\delta(Z_1)w(Y_3)\psi(Z_1,U_3)\tilde{R}_3] + E[\delta(Z_1)^2]\big) + o(1)$$

$$= \frac{(n-1)(n-2)}{mn}\big(E\big[2w(Y_2)\psi(Z_1,U_2)\tilde{R}_2 E[2w(Y_3)\psi(Z_1,U_3)\tilde{R}_3|Z_1,Z_2]\big]$$

$$- E[\delta(Z_1)E[2w(Y_2)\psi(Z_1,U_2)\tilde{R}_2|Z_1]] - E[\delta(Z_1)E[2w(Y_3)\psi(Z_1,U_3)\tilde{R}_3|Z_1]] + E[\delta(Z_1)^2]\big)$$

$$+ o(1)$$

$$= \frac{(n-1)(n-2)}{mn}\big(E\big[2w(Y_2)\psi(Z_1,U_2)\tilde{R}_2\delta(Z_1)\big] - E[\delta(Z_1)^2]\big) + o(1)$$

$$= o(1)$$

using iterated expectation and the definition of δ from Theorem 5.3.1. Thus,

$$\frac{1}{q}\sum_{\nu=1}^{q}\left(\frac{1}{\sqrt{m}}\sum_{j=(\nu-1)m+1}^{\nu m}B_{n,j}\right)^2 = o_p\left(\frac{m}{n}\right).$$

It remains to show

$$\frac{1}{q}\sum_{\nu=1}^{q}\left(\frac{1}{\sqrt{m}}\sum_{j=(\nu-1)m+1}^{\nu m}D_j\right)^2 = \sigma^2 + \mathcal{O}_p\left(\sqrt{\frac{m}{n}} + n^{-\frac{1}{4}}\right).$$

Note that the $D_j, j = 1, ..., qm$, are i.i.d. and centred with finite fourth moments, so that

$$E\Bigg[\frac{1}{q}\sum_{\nu=1}^{q}\Bigg(\frac{1}{\sqrt{m}}\sum_{j=(\nu-1)m+1}^{\nu m} D_j\Bigg)^2\Bigg] = E\Bigg[\Bigg(\frac{1}{\sqrt{m}}\sum_{j=1}^{m} D_j\Bigg)^2\Bigg] = E[D_1^2] = \sigma^2.$$

It holds that

$$\operatorname{Var}\Bigg(\frac{1}{q}\sum_{\nu=1}^{q}\Bigg(\frac{1}{\sqrt{m}}\sum_{j=(\nu-1)m+1}^{\nu m} D_j\Bigg)^2\Bigg) = \frac{1}{q}\operatorname{Var}\Bigg(\Bigg(\frac{1}{\sqrt{m}}\sum_{j=1}^{m} D_j\Bigg)^2\Bigg)$$

and

$$E\Bigg[\Bigg(\frac{1}{\sqrt{m}}\sum_{j=1}^{m} D_j\Bigg)^4\Bigg] = \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m}\sum_{l=1}^{m} E[D_iD_jD_kD_l] = 3\sigma^4,$$

that is

$$\frac{1}{q}\sum_{\nu=1}^{q}\Bigg(\frac{1}{\sqrt{m}}\sum_{j=(\nu-1)m+1}^{\nu m} D_j\Bigg)^2 - \sigma^2 = \mathcal{O}_p\Bigg(\frac{1}{\sqrt{q}}\Bigg) = \mathcal{O}_p\Bigg(\sqrt{\frac{m}{n}}\Bigg).$$

Finally, the Cauchy-Schwarz inequality can be applied to obtain

$$\begin{aligned}
&|\hat{\sigma}^2 - \sigma^2| \\
&= \Bigg|\frac{1}{q}\sum_{\nu=1}^{q}\Bigg(\frac{1}{\sqrt{m}}\sum_{j=(\nu-1)m+1}^{\nu m}(A_{n,j}+B_{n,j}+C_{n,j}+D_j)\Bigg)^2 - \sigma^2\Bigg| + \mathcal{O}_p\Bigg(\sqrt{\frac{m}{n}} + a_n\Bigg) \\
&\le \frac{1}{q}\sum_{\nu=1}^{q}\Bigg(\frac{1}{\sqrt{m}}\sum_{j=(\nu-1)m+1}^{\nu m}(A_{n,j}+B_{n,j}+C_{n,j})\Bigg)^2 \\
&\quad + \Bigg|\frac{1}{q}\sum_{\nu=1}^{q}\Bigg(\frac{1}{\sqrt{m}}\sum_{j=(\nu-1)m+1}^{\nu m} D_j\Bigg)^2 - \sigma^2\Bigg| \\
&\quad + \Bigg|\frac{2}{q}\sum_{\nu=1}^{q}\Bigg(\frac{1}{\sqrt{m}}\sum_{j=(\nu-1)m+1}^{\nu m}(A_{n,j}+B_{n,j}+C_{n,j})\Bigg)\Bigg(\frac{1}{\sqrt{m}}\sum_{j=(\nu-1)m+1}^{\nu m} D_j\Bigg)\Bigg| \\
&\quad + \mathcal{O}_p\Bigg(\sqrt{\frac{m}{n}} + a_n\Bigg) \\
&\le 2\sqrt{\frac{1}{q}\sum_{\nu=1}^{q}\Bigg(\frac{1}{\sqrt{m}}\sum_{j=(\nu-1)m+1}^{\nu m}(A_{n,j}+B_{n,j}+C_{n,j})\Bigg)^2}\sqrt{\frac{1}{q}\sum_{\nu=1}^{q}\Bigg(\frac{1}{\sqrt{m}}\sum_{j=(\nu-1)m+1}^{\nu m} D_j\Bigg)^2} \\
&\quad + \mathcal{O}_p\Bigg(\sqrt{\frac{m}{n}} + a_n + n^{-\frac{1}{2}}\Bigg) \\
&= \mathcal{O}_p\Bigg(\sqrt{\frac{m}{n}} + a_n + n^{-\frac{1}{4}}\Bigg).
\end{aligned}$$

The optimal convergence rate directly follows from this equation. □

5.8.5 Proof of Lemma 5.4.4

Only the second assertion is shown since the first one can be concluded similarly. The proof uses similar techniques as Hansen (2008). First, for the deviation terms $R_i = \hat{\varepsilon}_i(\tilde{\theta}) - \varepsilon_i(\theta_0)$

and appropriate $u_i^*, i=1,...,n$, a Taylor expansion leads to

$$\frac{1}{na_n}\sum_{i=1}^n f_\xi\Big(\frac{u-\varepsilon_i(\theta_0)-R_i}{a_n}\Big)=\frac{1}{na_n}\sum_{i=1}^n\Bigg(\sum_{j=0}^{r-1}\frac{(-R_i)^j}{a_n^j j!}f_\xi^{(j)}\Big(\frac{u-\varepsilon_i(\theta_0)}{a_n}\Big)+\frac{(-R_i)^r}{a_n^r r!}f_\xi^{(r)}(u_i^*)\Bigg). \tag{5.86}$$

For appropriate $\tilde{\theta}_y$ between $\tilde{\theta}$ and θ_0 the R_i can be split into

$$\begin{aligned}
R_i &= \hat{\varepsilon}_i(\tilde{\theta})-\varepsilon_i(\theta_0)\\
&=\frac{\Lambda_{\tilde{\theta}}(Y_i)-\Lambda_{\tilde{\theta}}(0)}{\Lambda_{\tilde{\theta}}(1)-\Lambda_{\tilde{\theta}}(0)}-\hat{g}(X_i)-\varepsilon_i(\theta_0)\\
&=\frac{\Lambda_{\tilde{\theta}}(Y_i)-\Lambda_{\tilde{\theta}}(0)}{\Lambda_{\tilde{\theta}}(1)-\Lambda_{\tilde{\theta}}(0)}-\frac{\Lambda_{\theta_0}(Y_i)-\Lambda_{\theta_0}(0)}{\Lambda_{\theta_0}(1)-\Lambda_{\theta_0}(0)}+g_{\theta_0}(X_i)-\hat{g}(X_i)\\
&=\frac{\Lambda_{\tilde{\theta}}(Y_i)-\Lambda_{\tilde{\theta}}(0)}{\Lambda_{\tilde{\theta}}(1)-\Lambda_{\tilde{\theta}}(0)}-\frac{\Lambda_{\tilde{\theta}}(Y_i)-\Lambda_{\tilde{\theta}}(0)}{\Lambda_{\theta_0}(1)-\Lambda_{\theta_0}(0)}+\frac{\Lambda_{\tilde{\theta}}(Y_i)-\Lambda_{\tilde{\theta}}(0)}{\Lambda_{\theta_0}(1)-\Lambda_{\theta_0}(0)}-\frac{\Lambda_{\theta_0}(Y_i)-\Lambda_{\theta_0}(0)}{\Lambda_{\theta_0}(1)-\Lambda_{\theta_0}(0)}\\
&\quad+g_{\theta_0}(X_i)-\hat{g}(X_i)\\
&=\frac{1}{(\Lambda_{\tilde{\theta}}(1)-\Lambda_{\tilde{\theta}}(0))(\Lambda_{\theta_0}(1)-\Lambda_{\theta_0}(0))}(\Lambda_{\tilde{\theta}}(Y_i)-\Lambda_{\tilde{\theta}}(0))(\Lambda_{\theta_0}(1)-\Lambda_{\tilde{\theta}}(1)+\Lambda_{\tilde{\theta}}(0)-\Lambda_{\theta_0}(0))\\
&\quad+\frac{\Lambda_{\tilde{\theta}}(Y_i)-\Lambda_{\theta_0}(Y_i)+\Lambda_{\theta_0}(0)-\Lambda_{\tilde{\theta}}(0)}{\Lambda_{\theta_0}(1)-\Lambda_{\theta_0}(0)}+g_{\theta_0}(X_i)-\hat{g}(X_i)\\
&=\frac{1}{(\Lambda_{\tilde{\theta}}(1)-\Lambda_{\tilde{\theta}}(0))(\Lambda_{\theta_0}(1)-\Lambda_{\theta_0}(0))}\Big(\Lambda_{\theta_0}(Y_i)-\Lambda_{\theta_0}(0)+\big(\dot{\Lambda}_{\theta_0}(Y_i)^t-\dot{\Lambda}_{\theta_0}(0)^t\big)(\tilde{\theta}-\theta_0)\\
&\qquad+\frac{1}{2}(\tilde{\theta}-\theta_0)^t\big(\operatorname{Hess}\Lambda_{\tilde{\theta}_{Y_i}}(Y_i)-\operatorname{Hess}\Lambda_{\tilde{\theta}_0}(0)\big)(\tilde{\theta}-\theta_0)\Big)\\
&\qquad\Big(-\dot{\Lambda}_{\theta_0}(1)^t(\tilde{\theta}-\theta_0)-\frac{1}{2}(\tilde{\theta}-\theta_0)^t\operatorname{Hess}\Lambda_{\tilde{\theta}_1}(1)(\tilde{\theta}-\theta_0)\\
&\qquad+\dot{\Lambda}_{\theta_0}(0)^t(\tilde{\theta}-\theta_0)+\frac{1}{2}(\tilde{\theta}-\theta_0)^t\operatorname{Hess}\Lambda_{\tilde{\theta}_0}(0)(\tilde{\theta}-\theta_0)\Big)\\
&\quad+\frac{1}{\Lambda_{\theta_0}(1)-\Lambda_{\theta_0}(0)}\Big(\dot{\Lambda}_{\theta_0}(Y_i)^t(\tilde{\theta}-\theta_0)+\frac{1}{2}(\tilde{\theta}-\theta_0)^t\operatorname{Hess}\Lambda_{\tilde{\theta}_{Y_i}}(Y_i)(\tilde{\theta}-\theta_0)\\
&\qquad-\dot{\Lambda}_{\theta_0}(0)^t(\tilde{\theta}-\theta_0)-\frac{1}{2}(\tilde{\theta}-\theta_0)^t\operatorname{Hess}\Lambda_{\tilde{\theta}_0}(0)(\tilde{\theta}-\theta_0)\Big)+g_{\theta_0}(X_i)-\hat{g}(X_i)\\
&=\tilde{R}_i+g_{\theta_0}(X_i)-\hat{g}(X_i).
\end{aligned}\tag{5.87}$$

Therefore,

$$\begin{aligned}
\frac{1}{na_n}\sum_{i=1}^n\Big|\frac{R_i^j}{a_n^j j!}f_\xi^{(j)}\Big(\frac{u-\varepsilon_i(\theta_0)}{a_n}\Big)\Big| &\leq \frac{C}{na_n}\sum_{i=1}^n\Big|\frac{\tilde{R}_i^j}{a_n^j j!}f_\xi^{(j)}\Big(\frac{u-\varepsilon_i(\theta_0)}{a_n}\Big)\Big|\\
&\quad+\frac{C}{na_n}\sum_{i=1}^n\Big|\frac{(\hat{g}(X_i)-g_{\theta_0}(X_i))^j}{a_n^j j!}f_\xi^{(j)}\Big(\frac{u-\varepsilon_i(\theta_0)}{a_n}\Big)\Big|
\end{aligned}$$

for all $j=1,...,r-1$ and

$$\frac{1}{na_n}\sum_{i=1}^n\Big|\frac{R_i^r}{a_n^r r!}f_\xi^{(r)}(u_i^*)\Big|\leq\frac{C}{na_n}\sum_{i=1}^n\Big|\frac{\tilde{R}_i^r}{a_n^r r!}f_\xi^{(r)}(u_i^*)\Big|+\frac{C}{na_n}\sum_{i=1}^n\Big|\frac{(\hat{g}(X_i)-g_{\theta_0}(X_i))^r}{a_n^r r!}f_\xi^{(r)}(u_i^*)\Big|$$

for some sufficiently large constant $C > 0$, so that it suffices to treat the cases $R_i^{(1)} = \hat{g}(X_i) - g_{\theta_0}(X_i)$ and $R_i^{(2)} = \tilde{R}_i$ separately.
When inserting $R_i^{(1)}$ in equation (5.86) negligibility of the last summand directly follows from (5.42) and the boundedness of $f_\xi^{(r)}$. Thanks to Hansen (2008), to prove

$$\sup_{u\in\mathcal{K}} \frac{1}{na_n}\sum_{i=1}^n \left| f_\xi^{(j)}\left(\frac{u-\varepsilon_i(\theta_0)}{a_n}\right)\right| \leq C + o_p(1)$$

for all $j = 0, ..., r-1$ and some constant $C > 0$, it suffices to show uniform (with respect to u) boundedness of the expectation. Hence, due to

$$\begin{aligned}
E\left[\frac{1}{na_n}\sum_{i=1}^n \left| f_\xi^{(j)}\left(\frac{u-\varepsilon_i(\theta_0)}{a_n}\right)\right|\right] &= E\left[\frac{1}{a_n}\left| f_\xi^{(j)}\left(\frac{u-\varepsilon(\theta_0)}{a_n}\right)\right|\right] \\
&= \int \frac{1}{a_n}\left| f_\xi^{(j)}\left(\frac{u-e}{a_n}\right)\right| f_{\varepsilon(\theta_0)}(e)\, de \\
&= \int |f_\xi^{(j)}(e)| f_{\varepsilon(\theta_0)}(u-a_n e)\, de \\
&\leq C
\end{aligned}$$

for some constant $C > 0$ (see (5.61)) and thus

$$\begin{aligned}
&\frac{1}{na_n}\sum_{i=1}^n \left| \frac{(\hat{g}(X_i)-g_{\theta_0}(X_i))^j}{a_n^j j!} f_\xi^{(j)}\left(\frac{u-\varepsilon_i(\theta_0)}{a_n}\right)\right| \\
&\leq \underbrace{\frac{\max_{k=1,\ldots,n} |\hat{g}(X_k)-g_{\theta_0}(X_k)|^j}{a_n^j j!}}_{=o_p(1)} \underbrace{\frac{1}{na_n}\sum_{i=1}^n \left| f_\xi^{(j)}\left(\frac{u-\varepsilon_i(\theta_0)}{a_n}\right)\right|}_{=\mathcal{O}_p(1)} \\
&= o_p(1)
\end{aligned}$$

for all $j = 1, ..., r-1$ one has

$$\frac{1}{na_n}\sum_{i=1}^n f_\xi\left(\frac{u-\varepsilon_i(\theta_0)+\hat{g}(X_i)-g_{\theta_0}(X_i)}{a_n}\right) = \frac{1}{na_n}\sum_{i=1}^n f_\xi\left(\frac{u-\varepsilon_i(\theta_0)}{a_n}\right) + o_p(1)$$

uniformly on compact sets. $\tilde{R}_i$ can be written as

$$\begin{aligned}
\tilde{R}_i = \mathcal{O}_p(||\tilde{\theta}-\theta_0||) &+ \mathcal{O}_p(||\tilde{\theta}-\theta_0||)\Big(\Lambda_{\theta_0}(Y_i) + \dot{\Lambda}_{\theta_0}(Y_i)^t(\tilde{\theta}-\theta_0) \\
&+ \frac{1}{2}(\tilde{\theta}-\theta_0)^t\big(\operatorname{Hess}\Lambda_{\tilde{\theta}_{Y_i,0}}(Y_i) - \operatorname{Hess}\Lambda_{\tilde{\theta}_{Y_i,0}}(0)\big)(\tilde{\theta}-\theta_0)\Big) \\
&+ \frac{1}{\Lambda_{\theta_0}(1)-\Lambda_{\theta_0}(0)}\Big(\dot{\Lambda}_{\theta_0}(Y_i)^t(\tilde{\theta}-\theta_0) + \frac{1}{2}(\tilde{\theta}-\theta_0)^t \operatorname{Hess}\Lambda_{\tilde{\theta}_{Y_i}}(Y_i)(\tilde{\theta}-\theta_0)\Big),
\end{aligned}$$

where the $\mathcal{O}_p$-terms are independent of i. When inserting $\tilde{R}_i$ in equation (5.86), one has for any $\delta > 0$

$$\frac{1}{na_n}\sum_{i=1}^n \left| \frac{\tilde{R}_i^j}{a_n^j j!} f_\xi^{(j)}\left(\frac{u-\varepsilon_i(\theta_0)}{a_n}\right)\right|$$

$$\leq \mathcal{O}_p\bigg(\frac{||\tilde{\theta}-\theta_0||^j}{a_n^j}\frac{1}{na_n}\sum_{i=1}^n\bigg|f_\xi^{(j)}\Big(\frac{u-\varepsilon_i(\theta_0)}{a_n}\Big)\bigg|\Big(1+|\Lambda_{\theta_0}(Y_i)|^j+||\dot{\Lambda}_{\theta_0}(Y_i)||^j$$

$$+\,||\tilde{\theta}-\theta_0||^j\sup_{||\theta-\theta_0||<\delta}||\operatorname{Hess}\Lambda_\theta(Y_i)||^j\Big)\bigg)$$

for all $j = 1, ..., r-1$. By assumptions (5.61) and (5.62) the expected value of the sum

$$\frac{1}{na_n}\sum_{i=1}^n\bigg|f_\xi^{(j)}\Big(\frac{u-\varepsilon_i(\theta_0)}{a_n}\Big)\bigg|\Big(1+|\Lambda_{\theta_0}(Y_i)|^j+||\dot{\Lambda}_{\theta_0}(Y_i)||^j+||\tilde{\theta}-\theta_0||^j\sup_{||\theta-\theta_0||<\delta}||\operatorname{Hess}\Lambda_\theta(Y_i)||^j\Big)$$

can be bounded by some constant $C > 0$, so that

$$\frac{1}{na_n}\sum_{i=1}^n\bigg|\frac{\tilde{R}_i^j}{a_n^j j!}f_\xi^{(j)}\Big(\frac{u-\varepsilon_i(\theta_0)}{a_n}\Big)\bigg|$$

$$\leq \mathcal{O}_p\bigg(\frac{||\tilde{\theta}-\theta_0||^{2j}}{a_n^j}\frac{1}{na_n}\sum_{i=1}^n\bigg|f_\xi^{(j)}\Big(\frac{u-\varepsilon_i(\theta_0)}{a_n}\Big)\bigg|\sup_{||\theta-\theta_0||<\delta}||\operatorname{Hess}\Lambda_\theta(Y_i)||^j\bigg)+o_p(1)$$

$$= o_p(1)$$

by (5.42) and (5.63). The remaining term can be treated similarly by applying (5.64) and (5.65) to obtain

$$\frac{1}{na_n}\sum_{i=1}^n\bigg|\frac{\tilde{R}_i^r}{a_n^r r!}f_\xi^{(r)}(u_i^*)\bigg|$$

$$\leq \mathcal{O}_p\bigg(\frac{||\tilde{\theta}-\theta_0||^r}{a_n^{r+1}}\frac{1}{n}\sum_{i=1}^n\Big(|\Lambda_{\theta_0}(Y_i)|^r+||\dot{\Lambda}_{\theta_0}(Y_i)||^r+||\tilde{\theta}-\theta_0||^r\sup_{||\theta-\theta_0||<\delta}||\operatorname{Hess}\Lambda_\theta(Y_i)||^r\Big)\bigg)$$

$$= o_p(1).$$

□

5.8.6 Proof of Lemma 5.4.6

Note that conditional on $(Y_1, X_1), ..., (Y_n, X_n)$ the random variables $(Y_1^*, X_1^*), ..., (Y_m^*, X_m^*)$ are independent as well as identically distributed. Moreover, after conditioning on the original data, the assumptions (**B1**)–(**B10**) are valid with probability converging to one, so that due to Remark 5.4.3 the same reasoning as in the paper of Colling and Van Keilegom (2019) can be applied to obtain (5.41).
For notational convenience the conditional distribution of $(Y_1^*, X_1^*), ..., (Y_m^*, X_m^*)$ conditional on $(Y_1, X_1), ..., (Y_n, X_n)$ is written as P^* and the expectation with respect to P^* is written as E^*. From a notational point of view, P^* and E^* replace P_2^1 and $E[\cdot|\omega]$ in the following if misunderstandings can be excluded. Let $F_{Y^*|X^*}$ denote the conditional distribution function of Y_1^* conditioned on X_1^* (and $(Y_1, X_1), ..., (Y_n, Y_n)$). To verify (**A7***) ψ^* has to be examined further and to define ψ^* some further notations are needed. Let v be the weighting function from assumption (**B7**) and define

$$s_1^*(u,x)=\int_0^u\frac{\frac{\partial F_{Y^*|X^*}(y|x)}{\partial y}}{\frac{\partial F_{Y^*|X^*}(y|x)}{\partial x_1}}\,dy,\quad \tilde{v}_1^*(u_0,x)=\frac{v(x)}{s_1^*(u_0,x)},\quad \tilde{v}_2^*(u_0,x)=\frac{v(x)s_1^*(u_0,x)}{s_1^*(1,x)^2}$$

and (for $\tilde v^* = \tilde v_1^*, \tilde v_2^*$)

$$
\begin{aligned}
\delta_j^{*\tilde v^*}(u_0,u) = &\int_{\max(0,U_j^*)}^{\max(u,U_j^*)} \Big(\tilde v^*(u_0,X_j^*)D_{p,0}^*(r,X_j^*) - \frac{\partial}{\partial x_1}\big(\tilde v^*(u_0,x)D_{p,1}^*(r,x)\big)\Big|_{x=X_j^*}\Big)\,dr\\
&+\int_0^u \Big(\tilde v^*(u_0,X_j^*)D_{f,0}^*(r,X_j^*) - \frac{\partial}{\partial x_1}\big(\tilde v^*(u_0,X_j^*)D_{f,1}^*(r,x)\big)\Big|_{x=X_j^*}\Big)\,dr\\
&+(I_{\{U_j^*\le u\}} - I_{\{U_j^*\le 0\}})\tilde v^*(u_0,X_j^*)D_{p,u}^*(U_j^*,X_j^*)\\
&+\int_0^u \Big(\frac{I_{\{U_j^*\le u\}} - I_{\{U_j^*\le 0\}}}{F_{U^*}(1)-F_{U^*}(0)} - r\Big)\\
&\int_{\mathcal X}\bigg(\Big(\tilde v^*(u_0,x)D_{p,0}^*(r,x) + \frac{\partial}{\partial x_1}\big(\tilde v^*(u_0,x)D_{p,1}^*(r,x)\big)\Big)\\
&f_{U^*,X^*}(r,x) + \tilde v^*(u_0,x)D_{p,u}^*(r,x)\frac{\partial}{\partial r}f_{U^*,X^*}(r,x)\bigg)\,dx\,dr\\
&-\Big(\frac{I_{\{U_j^*\le 1\}} - I_{\{U_j^*\le 0\}}}{F_{U^*}(1)-F_{U^*}(0)} - 1\Big)\int_0^u r\int_{\mathcal X}\Big(\tilde v^*(u_0,x)D_{p,0}^*(r,x)\\
&-\tilde v^*(u_0,x)\frac{\partial}{\partial r}D_{p,u}^*(r,x) + \frac{\partial}{\partial x_1}\big(\tilde v^*(u_0,x)D_{p,1}^*(r,x)\big)\Big)f_{U^*,X^*}(r,x)\,dx\,dr\\
&-\Big(\frac{\hat F_{U^*}(1)-\hat F_{U^*}(0)}{F_{U^*}(1)-F_{U^*}(0)} - 1\Big)u\int_{\mathcal X}\tilde v^*(u_0,x)D_{p,u}^*(u,x)f_{U^*,X^*}(u,x)\,dx,
\end{aligned}
$$

where $D_{p,0}^*(u,x),...,D_{f,1}^*(u,x)$ are defined as

$$
\begin{aligned}
D_{p,0}^*(u,x) &= \frac{\Phi^*(u,x)\frac{\partial}{\partial x_1}f_{X^*}(x)}{\Phi_i^*(u,x)^2 f_{X^*}(x)^2}, &(5.88)\\
D_{p,u}^*(u,x) &= \frac{1}{f_{X^*}(x)\Phi_i^*(u,x)},\\
D_{p,1}^*(u,x) &= \frac{-\Phi_u^*(u,x)}{f_{X^*}(x)\Phi_i^*(u,x)^2}\\
D_{f,0}^*(u,x) &= \frac{-\Phi_u^*(u,x)\Phi^*(u,x)\frac{\partial}{\partial x_1}f_{X^*}(x)}{\Phi_i^*(u,x)^2 f_{X^*}(x)^2}\\
D_{f,1}^*(u,x) &= \frac{\Phi_u^*(u,x)\Phi^*(u,x)}{\Phi_i^*(u,x)^2 f_{X^*}(x)}, &(5.89)
\end{aligned}
$$

where $\Phi^*(u|x) = F_{U^*|X^*}(u,x) = \frac{p^*(u,x)}{f_{X^*}(x)}, \Phi_u^*(u|x) = \frac{\partial}{\partial u}\Phi^*(u,x)$, $\Phi_i^*(u|x) = \frac{\partial}{\partial x_1}\Phi^*(u,x)$ (compare (1.4)) for U^* as in (5.40) and f_{X^*} are analogously defined as in Section 1.1, but based on $(Y_i^*, X_i^*), i = 1,...,m$, from Algorithm 5.4.1. Then, ψ^* is defined as

$$
\begin{aligned}
&\psi^*(Z_j^*,u)\\
&= \delta_j^{*\tilde v_1^*}(1,u) - \delta_j^{*\tilde v_2^*}(u,1) + \frac{Q^{*\prime}(u)}{F_{U^*}(1)-F_{U^*}(0)}\big(I_{\{U_j^*\le u\}} - I_{\{U_j^*\le 0\}} - F_{U^*}(u) + F_{U^*}(0)\big)\\
&\quad - Q^{*\prime}(u)\frac{F_{U^*}(u)-F_{U^*}(0)}{(F_{U^*}(1)-F_{U^*}(0))^2}\big(I_{\{U_j^*\le 1\}} - I_{\{U_j^*\le 0\}} - F_{U^*}(1) + F_{U^*}(0)\big). \qquad (5.90)
\end{aligned}
$$

Condition (**A7***) for ψ^* is implied by the same reasoning as in the paper of Colling and Van Keilegom (2019). Note that the first part of Remark 5.4.3 ensures that v can be used as the weighting function for the bootstrap data as well.
To prove (**A8***), an auxiliary lemma is shown in the following. Thanks to the expressions above for $\psi^*, D^*_{p,0}(u,x), ..., D^*_{f,1}(u,x)$, equation (5.58) and (5.60) will be a direct consequence of Lemma 5.8.2, while (5.59) follows from expression (5.90) and boundedness of $\delta_j^{*\tilde{v}_1^*}, \delta_j^{*\tilde{v}_2^*}$ on compact sets, so that ψ^* fulfils (**A8***) then.

Lemma 5.8.2 *Let* $\mathcal{C} \subseteq \big(-\frac{F_Y(0)}{F_Y(1)-F_Y(0)}, \frac{1-F_Y(0)}{F_Y(1)-F_Y(0)}\big)$ *be a compact set and define*

$$F_{U^*,X^*}(t,x) = P^*(U^* \leq t, X^* \leq x), \quad \text{and} \quad p^*(t,x) = \frac{\partial^{d_X}}{\partial x_1 ... \partial x_{d_X}} F_{U^*,X^*}(t,x).$$

*Let i be as in (**B3**). Under the assumptions of Lemma 5.4.6, one has*

$$\begin{aligned}
\sup_{x \in \mathrm{supp}(v)} |f_{X^*}(x) - f(x)| &= o_p(1), \\
\sup_{x \in \mathrm{supp}(v)} \left| \frac{\partial}{\partial x_1} f_{X^*}(x) - \frac{\partial}{\partial x_1} f(x) \right| &= o_p(1), \\
\sup_{x \in \mathrm{supp}(v)} \left| \frac{\partial^2}{\partial x_1^2} f_{X^*}(x) - \frac{\partial^2}{\partial x_1^2} f(x) \right| &= o_p(1), \\
\sup_{u \in \mathcal{C}, x \in \mathrm{supp}(v)} |p^*(u,x) - p(u,x)| &= o_p(1), \\
\sup_{u \in \mathcal{C}, x \in \mathrm{supp}(v)} \left| \frac{\partial}{\partial u} p^*(u,x) - \frac{\partial}{\partial u} p(u,x) \right| &= o_p(1), \qquad (5.91) \\
\sup_{u \in \mathcal{C}, x \in \mathrm{supp}(v)} \left| \frac{\partial}{\partial x_1} p^*(u,x) - \frac{\partial}{\partial x_1} p(u,x) \right| &= o_p(1), \\
\sup_{u \in \mathcal{C}} |F_{U^*}(u) - F_U(u)| &= o_p(1), \\
\sup_{z \in \mathcal{R}^{d_X+1}} |F_{Z^*}(u) - F_Z(z)| &= o_p(1).
\end{aligned}$$

Here, p is defined as in 1.3, but for $f_{U,X}$ *instead of* $f_{Y,X}$*, and* f_{X^*} *is defined as the density of* X_1^* *from Algorithm 5.4.1 (conditional on* $(Y_i, X_i), i = 1, ..., n$*).*

Proof: Most of the proof will be quite similar to proving uniform convergence rates for kernel estimates in the unconditional case (see Lemma 4.2.12), especially the results of Hansen (2008) are applied several times. While doing so, note that due to (5.39) and (5.44) kernel estimates like

$$\hat{f}_{\varepsilon(\theta_0)}(e) = \frac{1}{na_n} \sum_{i=1}^n f_\xi \bigg(\frac{e - \varepsilon(\theta_0)}{a_n} \bigg)$$

converge uniformly in $e \in \mathbb{R}$ to their expectation (see Theorem 4 of Hansen (2008)).
First, appropriate expressions for f_{X^*} and p^* have to be found. While for f_{X^*} the kernel estimator

$$f_{X^*}(x) = \frac{1}{nb_n^{d_X}} \sum_{i=1}^n f_W \bigg(\frac{x - X_i}{b_n} \bigg)$$

is obtained, p^* can be expressed for any $j \in \{1, ..., m\}$ as

$$
\begin{aligned}
& p^*(\mathcal{T}^*((h^*)^{-1}(u)), x) \\
&= \frac{\partial^{d_X}}{\partial x_1 ... \partial x_{d_X}} F_{U^*,X^*}(\mathcal{T}^*((h^*)^{-1}(u)), x) \\
&= \frac{\partial^{d_X}}{\partial x_1 ... \partial x_{d_X}} P^*(S_j^* \leq u, X_j^* \leq x) \\
&= \frac{\partial^{d_X}}{\partial x_1 ... \partial x_{d_X}} \int_{(-\infty,x]} P^*(S_j^* \leq u | X_j^* = z) f_{X^*}(z)\, dz \\
&= \frac{\partial^{d_X}}{\partial x_1 ... \partial x_{d_X}} \int_{(-\infty,x]} P^*\bigg(\xi \leq \frac{u - \hat{g}(X_j^*) - \hat{\varepsilon}_{j_\varepsilon^*}(\tilde{\theta}) + \frac{1}{n}\sum_{l=1}^n \hat{\varepsilon}_l(\tilde{\theta})}{a_n} \bigg| X^* = z\bigg) f_{X^*}(z)\, dz \\
&= \frac{\partial^{d_X}}{\partial x_1 ... \partial x_{d_X}} \int_{(-\infty,x]} \frac{1}{n}\sum_{k=1}^n P^*\bigg(\xi \leq \frac{u - \hat{g}(z) - \hat{\varepsilon}_k(\tilde{\theta}) + \frac{1}{n}\sum_{l=1}^n \hat{\varepsilon}_l(\tilde{\theta})}{a_n}\bigg) f_{X^*}(z)\, dz \\
&= \frac{\partial^{d_X}}{\partial x_1 ... \partial x_{d_X}} \int_{(-\infty,x]} \frac{1}{n}\sum_{k=1}^n F_\xi\bigg(\frac{u - \hat{g}(z) - \hat{\varepsilon}_k(\tilde{\theta}) + \frac{1}{n}\sum_{l=1}^n \hat{\varepsilon}_l(\tilde{\theta})}{a_n}\bigg) f_{X^*}(z)\, dz \\
&= \frac{1}{n}\sum_{k=1}^n F_\xi\bigg(\frac{u - \hat{g}(x) - \hat{\varepsilon}_k(\tilde{\theta}) + \frac{1}{n}\sum_{l=1}^n \hat{\varepsilon}_l(\tilde{\theta})}{a_n}\bigg) f_{X^*}(x),
\end{aligned}
$$

where F_ξ denotes the cumulative distribution function of ξ. See Section 4.2.2 for

$$
\sup_{x \in \text{supp}(v)} |f_{X^*}(x) - f_X(x)| = \mathcal{O}_p\bigg(b_n^2 + \sqrt{\frac{\log(n)}{n b_n^{d_X}}}\bigg) = o_p(1)
$$

(note that f_W is a kernel of order 2). The assertion for $\frac{\partial}{\partial x_1} f_{X^*}$ and $\frac{\partial^2}{\partial x_1^2} f_{X^*}$ follows similarly by applying for example

$$
\begin{aligned}
E\bigg[\frac{\partial^2}{\partial v_i^2} f_{X^*}(v)\bigg] &= E\bigg[\frac{1}{b_n^{d_X+2}} \frac{\partial^2}{\partial z_i^2} f_W(z)\Big|_{z=\frac{v-X}{b_n}}\bigg] \\
&= \frac{1}{b_n^{d_X+2}} \int \frac{\partial^2}{\partial z_i^2} f_W(z)\Big|_{z=\frac{v-x}{b_n}} f_X(x)\, dx \\
&= \frac{1}{b_n^2} \int \frac{\partial^2}{\partial x_1^2} f_W(x) f_X(v - b_n x)\, dx \\
&= \underbrace{\bigg[\frac{1}{b_n^2} \frac{\partial}{\partial x_1} f_W(x) f_X(x - b_n v)\bigg]_{-\infty}^{\infty}}_{=0} + \frac{1}{b_n} \int \frac{\partial}{\partial x_1} f_W(x) \frac{\partial}{\partial z_i} f_X(z)\big|_{z=v-b_n x}\, dx \\
&= \underbrace{\bigg[\frac{1}{b_n} f_W(x) \frac{\partial}{\partial z_i} f_X(z)\big|_{z=v-b_n x}\bigg]_{-\infty}^{\infty}}_{=0} + \int f_W(x) \frac{\partial^2}{\partial z_i^2} f_X(z)\big|_{z=v-b_n x}\, dx \\
&= \frac{\partial^2}{\partial v_i^2} f_X(v) + \mathcal{O}(b_n^2).
\end{aligned}
$$

where $(-\infty, x] = \times_{i=1}^{d_X}(-\infty, x_i]$. From now on, only $\frac{\partial}{\partial x_1}p^*$ is considered, since the other terms can be treated analogously. Recall that for all $k \in \{1, ..., n\}$ and appropriate $\tilde{\theta}_y$ between $\tilde{\theta}$ and θ_0 it holds that (see (5.87))

$$\begin{aligned}
&\hat{\varepsilon}_k(\tilde{\theta})\\
&= \varepsilon_k(\theta_0) + \frac{1}{(\Lambda_{\tilde{\theta}}(1) - \Lambda_{\tilde{\theta}}(0))(\Lambda_{\theta_0}(1) - \Lambda_{\theta_0}(0))}\Big(\Lambda_{\theta_0}(Y_k) - \Lambda_{\theta_0}(0)\\
&\quad + \big(\dot{\Lambda}_{\theta_0}(Y_k)^t - \dot{\Lambda}_{\theta_0}(0)^t\big)(\tilde{\theta} - \theta_0) + \frac{1}{2}(\tilde{\theta} - \theta_0)^t\big(\operatorname{Hess}\Lambda_{\tilde{\theta}_{Y_k}}(Y_k) - \operatorname{Hess}\Lambda_{\tilde{\theta}_0}(0)\big)(\tilde{\theta} - \theta_0)\Big)\\
&\quad \Big(- \dot{\Lambda}_{\theta_0}(1)^t(\tilde{\theta} - \theta_0) - \frac{1}{2}(\tilde{\theta} - \theta_0)^t \operatorname{Hess}\Lambda_{\tilde{\theta}_1}(1)(\tilde{\theta} - \theta_0)\\
&\quad + \dot{\Lambda}_{\theta_0}(0)^t(\tilde{\theta} - \theta_0) + \frac{1}{2}(\tilde{\theta} - \theta_0)^t \operatorname{Hess}\Lambda_{\tilde{\theta}_0}(0)(\tilde{\theta} - \theta_0)\Big)\\
&\quad + \frac{1}{\Lambda_{\theta_0}(1) - \Lambda_{\theta_0}(0)}\Big(\dot{\Lambda}_{\theta_0}(Y_k)^t(\tilde{\theta} - \theta_0) + \frac{1}{2}(\tilde{\theta} - \theta_0)^t \operatorname{Hess}\Lambda_{\tilde{\theta}_{Y_k}}(Y_k)(\tilde{\theta} - \theta_0)\\
&\quad - \dot{\Lambda}_{\theta_0}(0)^t(\tilde{\theta} - \theta_0) - \frac{1}{2}(\tilde{\theta} - \theta_0)^t \operatorname{Hess}\Lambda_{\tilde{\theta}_0}(0)(\tilde{\theta} - \theta_0)\Big) + g(X_k) - \hat{g}(X_k).
\end{aligned}$$

Thus, (if $||\tilde{\theta} - \theta_0|| \leq \delta$)

$$\begin{aligned}
\left|\frac{1}{n}\sum_{l=1}^n \hat{\varepsilon}_l(\tilde{\theta})\right| &\leq \left|\frac{1}{n}\sum_{l=1}^n \varepsilon_l(\theta_0)\right| + \mathcal{O}_p(||\tilde{\theta} - \theta_0||)\frac{1}{n}\sum_{l=1}^n \Big(1 + |\Lambda_{\theta_0}(Y_l)| + ||\dot{\Lambda}_{\theta_0}(Y_l)||\\
&\quad + ||\tilde{\theta} - \theta_0|| \sup_{||\theta - \theta_0|| < \delta} ||\operatorname{Hess}\Lambda_\theta(Y_l)||\Big) + \max_{k=1,\dots,n}|\hat{g}(X_k) - g(X_k)|\\
&= \mathcal{O}_p\left(\frac{1}{\sqrt{n}}\right) + \mathcal{O}_p\big(||\tilde{\theta} - \theta_0||\big) + \mathcal{O}_p\Big(\max_{k=1,\dots,n}|\hat{g}(X_k) - g(X_k)|\Big)\\
&= o_p(1).
\end{aligned}$$

Since $P(||\tilde{\theta} - \theta_0|| \leq \delta) \to 1$, this can be used together with Lemma 5.4.4 to obtain

$$\begin{aligned}
&\frac{\partial}{\partial x_1}p^*(\mathcal{T}^*((h^*)^{-1}(u)), x)\\
&= \frac{1}{n}\sum_{k=1}^n \frac{\partial}{\partial x_1}F_\xi\Big(\frac{u - \hat{g}(x) - \hat{\varepsilon}_k(\tilde{\theta}) + \frac{1}{n}\sum_{l=1}^n \hat{\varepsilon}_l(\tilde{\theta})}{a_n}\Big)f_{X^*}(x)\\
&= \frac{1}{n}\sum_{k=1}^n \Big[- \frac{f_{X^*}(x)}{a_n}f_\xi\Big(\frac{u - \hat{g}(x) - \hat{\varepsilon}_k(\tilde{\theta}) + \frac{1}{n}\sum_{l=1}^n \hat{\varepsilon}_l(\tilde{\theta})}{a_n}\Big)\frac{\partial}{\partial x_1}\hat{g}(x)\\
&\quad + F_\xi\Big(\frac{u - \hat{g}(x) - \hat{\varepsilon}_k(\tilde{\theta}) + \frac{1}{n}\sum_{l=1}^n \hat{\varepsilon}_l(\tilde{\theta})}{a_n}\Big)\frac{\partial}{\partial x_1}f_{X^*}(x)\Big]\\
&= -\frac{f_{X^*}(x)\frac{\partial}{\partial x_1}\hat{g}(x)}{na_n}\sum_{k=1}^n f_\xi\Big(\frac{u - \hat{g}(x) + \frac{1}{n}\sum_{l=1}^n \hat{\varepsilon}_l(\tilde{\theta}) - \varepsilon_k(\theta_0)}{a_n}\Big)\\
&\quad + \frac{\frac{\partial}{\partial x_1}f_{X^*}(x)}{n}\sum_{k=1}^n F_\xi\Big(\frac{u - \hat{g}(x) + \frac{1}{n}\sum_{l=1}^n \hat{\varepsilon}_l(\tilde{\theta}) - \varepsilon_k(\theta_0)}{a_n}\Big) + o_p(1)
\end{aligned}$$

$$= -f_{\varepsilon(\theta_0)}\Big(u - \hat{g}(x) + \frac{1}{n}\sum_{l=1}^{n}\hat{\varepsilon}_l(\tilde{\theta})\Big) f_{X^*}(x)\frac{\partial}{\partial x_1}\hat{g}(x)$$

$$+ F_{\varepsilon(\theta_0)}\Big(u - \hat{g}(x) + \frac{1}{n}\sum_{l=1}^{n}\hat{\varepsilon}_l(\tilde{\theta})\Big)\frac{\partial}{\partial x_1} f_{X^*}(x) + o_p(1)$$

$$= -f_{\varepsilon(\theta_0)}(u - g(x)) f_X(x)\frac{\partial}{\partial x_1} g(x) + F_{\varepsilon(\theta_0)}(u - g(x))\frac{\partial}{\partial x_1} f_X(x) + o_p(1)$$

$$= \frac{\partial}{\partial x_1} p(\mathcal{T}((h)^{-1}(u)), x) + o_p(1) \tag{5.92}$$

uniformly with respect to $x \in \mathrm{supp}(v)$ and with respect to u belonging to some compact set $\mathcal{K}$, where the third from last equality again follows from Theorem 4 of Hansen (2008) as in Section 4.2.2. A similar reasoning for $\frac{\partial}{\partial u} p^*$ results in

$$\frac{\partial}{\partial u} p^*(\mathcal{T}^*((h^*)^{-1}(u)), x) = \frac{\partial}{\partial u} p(\mathcal{T}((h)^{-1}(u)), x) \tag{5.93}$$

uniformly in $(u, x) \in \mathcal{K} \times \mathrm{supp}(v)$. Later, it will be shown that

$$\mathcal{T}^*((h^*)^{-1}(u)) - \mathcal{T}(h^{-1}(u)) = o_p(1)$$

as well as

$$\frac{\partial}{\partial u}\mathcal{T}^*((h^*)^{-1}(u)) - \frac{\partial}{\partial u}\mathcal{T}(h^{-1}(u)) = o_p(1) \tag{5.94}$$

uniformly on compact sets. Hence, after possibly adjusting the set of admissible values for u, (5.93) leads to

$$\frac{\partial}{\partial u} p^*(u, x) = \frac{\frac{\partial}{\partial u} p^*(\mathcal{T}^*((h^*)^{-1}(t)), x)\big|_{t = h^*((\mathcal{T}^*)^{-1}(u))}}{\frac{\partial}{\partial u}\mathcal{T}^*((h^*)^{-1}(t))\big|_{t = h^*((\mathcal{T}^*)^{-1}(u))}}$$

$$= \frac{\frac{\partial}{\partial u} p(\mathcal{T}(h^{-1}(t)), x)\big|_{t = h^*((\mathcal{T}^*)^{-1}(u))}}{\frac{\partial}{\partial u}\mathcal{T}^*((h^*)^{-1}(t))\big|_{t = h^*((\mathcal{T}^*)^{-1}(u))}} + o_p(1)$$

$$= \frac{\frac{\partial}{\partial u} p(\mathcal{T}(h^{-1}(t)), x)\big|_{t = h(\mathcal{T}^{-1}(u))}}{\frac{\partial}{\partial u}\mathcal{T}(h^{-1}(t))\big|_{t = h(\mathcal{T}^{-1}(u))}} + o_p(1)$$

$$= \frac{\partial}{\partial u} p(u, x) + o_p(1)$$

as well as

$$\frac{\partial}{\partial x_1} p^*(u, x) = \frac{\partial}{\partial x_1} p(u, x) + o_p(1)$$

uniformly on $(u, x) \in \mathcal{C} \times \mathrm{supp}(v)$. Indeed, equation (5.94) can be shown very similar to (5.92), because

$$\mathcal{T}^*((h^*)^{-1}(u)) = \frac{F_{S^*}(u) - F_{S^*}(0)}{F_{S^*}(1) - F_{S^*}(0)}$$

as well as

$$F_{S^*}(u) = P^*(S_j^* \leq u)$$

$$
\begin{aligned}
&= P^*\Big(\hat{g}\big(X^*_{j^*_X} + b_n W_j\big) + \hat{\varepsilon}_{j^*_\varepsilon}(\tilde{\theta}) - \frac{1}{n}\sum_{l=1}^n \hat{\varepsilon}_l(\tilde{\theta}) + a_n \xi_j \le u\Big) \\
&= \frac{1}{n}\sum_{i=1}^n \frac{1}{n}\sum_{k=1}^n \int F_\xi\Big(\frac{u - \hat{g}(X_i + b_n w) - \hat{\varepsilon}_k(\tilde{\theta}) + \frac{1}{n}\sum_{l=1}^n \hat{\varepsilon}_l(\tilde{\theta})}{a_n}\Big) f_W(w)\, dw
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial}{\partial u} F_{S^*}(u) &= \int \frac{1}{n}\sum_{i=1}^n \frac{1}{n}\sum_{k=1}^n \frac{\partial}{\partial u} F_\xi\Big(\frac{u - \hat{g}(X_i + b_n w) - \hat{\varepsilon}_k(\tilde{\theta}) + \frac{1}{n}\sum_{l=1}^n \hat{\varepsilon}_l(\tilde{\theta})}{a_n}\Big) f_W(w)\, dw \\
&= \int \frac{1}{n}\sum_{i=1}^n \frac{1}{n a_n}\sum_{k=1}^n f_\xi\Big(\frac{u - \hat{g}(X_i + b_n w) + \frac{1}{n}\sum_{l=1}^n \hat{\varepsilon}_l(\tilde{\theta}) - \hat{\varepsilon}_k(\tilde{\theta})}{a_n}\Big) f_W(w)\, dw \\
&= \int \frac{1}{n}\sum_{i=1}^n f_\varepsilon\Big(u - \hat{g}(X_i + b_n w) + \frac{1}{n}\sum_{l=1}^n \hat{\varepsilon}_l(\tilde{\theta})\Big) f_W(w)\, dw + o_p(1) \\
&= \frac{1}{n}\sum_{i=1}^n f_{\varepsilon(\theta_0)}(u - g(X_i)) + o_p(1) \\
&= E[f_{\varepsilon(\theta_0)}(u - g(X_1))] + o_p(1) \\
&= \frac{\partial}{\partial u} F_S(u) + o_p(1)
\end{aligned}
$$

uniformly on compact sets, where the third equation follows as in (5.92) and the last equation follows from the dominated convergence theorem. The uniform convergence of F_{U^*} follows by the same arguments. For the treatment of F_{Z^*}, equation (5.91) can be used together with the fact that $\frac{\partial}{\partial u} p^*(u,x)$ is a density:
For an arbitrary compact set $\mathcal{K}$ equation (5.91) implies the uniform convergence

$$
\sup_{(u,x)\in\mathcal{K}} \Big|\frac{\partial}{\partial u} p^*(u,x) - \frac{\partial}{\partial u} p(u,x)\Big| = o_p(1).
$$

Let $\varepsilon > 0$. As a probability measure on a polish space, P^Z is tight. Thus, there exists a compact set $\mathcal{K} \subseteq \mathbb{R}$ such that $P^Z(\mathcal{K}) > 1 - \frac{\varepsilon}{2}$. Hence, for all $z \in \mathbb{R}$ it holds that

$$
\begin{aligned}
|F_{Z^*}(z) - F_Z(z)| &\le \int\int I_{\mathcal{K}}(t,x)\Big|\frac{\partial}{\partial t} p^*(t,x) - \frac{\partial}{\partial t} p(t,x)\Big|\, dt\, dx + P_2^1(\cdot, \mathcal{K}^c)^{Z^*} + P^Z(\mathcal{K}^c) \\
&\le \varepsilon + o_p(1).
\end{aligned}
$$

Since $\varepsilon > 0$ can be chosen arbitrarily small one has $\sup_{z\in\mathcal{K}} |F_{Z^*}(z) - F_Z(z)| = o_p(1)$. □

5.8.7 Proof of Theorem 5.4.7

Again, as in the proof of Lemma 5.4.6, the conditional distribution and conditional expectation of $(Y_1^*, X_1^*), ..., (Y_m^*, X_m^*)$ given $(Y_1, X_1), ..., (Y_n, X_n)$ are denoted by P^* and E^*, respectively. Consider $\omega \in A_n$ with A_n from (**A7***). The proof can be divided into two parts: First, the uniform convergence of some bootstrap components appearing in the asymptotic distribution of the bootstrap test statistic is proven and second, the assertion itself is shown by the convergence of the conditional distribution functions in probability. Referring to the definition of h^*, one has

(**A5***) With probability converging to one, minimizing the function $M^* : \Upsilon \to \mathbb{R}$,

$$\gamma \;= (c_1, c_2, \theta^t)^t \mapsto E^*\big[w(Y^*)(h^*(Y^*)c_1 + c_2 - \Lambda_\theta(Y^*))^2\big]$$
$$\Big(= E\big[w(Y^*)(h^*(Y^*)c_1 + c_2 - \Lambda_\theta(Y^*))^2|\omega\big]\Big)$$

leads to a unique solution

$$\gamma^* = (c^*_{1,0}, c^*_{2,0}, \theta^*_0) \quad \Big(= \big(\Lambda_{\tilde{\theta}}(1) - \Lambda_{\tilde{\theta}}(0), \Lambda_{\tilde{\theta}}(0), \tilde{\theta}\big)\Big)$$

in the interior of Υ.

Here, uniqueness follows due to $E^*[w(Y^*)(h^*(Y^*)(\Lambda_{\tilde{\theta}}(1) - \Lambda_{\tilde{\theta}}(0)) + \Lambda_{\tilde{\theta}}(0) - \Lambda_{\tilde{\theta}}(Y^*))^2] = 0$ from (**A5**). With the notations

$$R^*(s) := (s, 1, -\dot{\Lambda}_{\tilde{\theta}}((h^*)^{-1}(s)))^t,$$
$$\Gamma^* := E^*[w((h^*)^{-1}(S^*_1))R^*(S^*_1)R^*(S^*_1)^t],$$
$$\varphi^*(z) := E^*[w((h^*)^{-1}(S^*_2))\psi^*(Z^*_1, U^*_2)R^*(S^*_2) \mid Z^*_1 = z],$$

a function ζ^* can be defined as

$$\zeta^*(z_1, z_2) := E^*\Big[w((h^*)^{-1}(S^*_3))\big(\psi^*(Z^*_1, U^*_3) - \varphi^*(Z^*_1)^t(\Gamma^*)^{-1}R^*(S^*_3)\big)$$
$$\big(\psi^*(Z^*_2, U^*_3) - \varphi^*(Z^*_2)^t(\Gamma^*)^{-1}R^*(S^*_3)\big) \mid Z^*_1 = z_1, Z^*_2 = z_2\Big].$$

Moreover, define

$$T^*_n = {c^*_{1,0}}^2 \sum_{k=1}^{\infty} \lambda^*_k V_k^2 \quad \text{with} \quad c^*_{1,0} = \Lambda_{\tilde{\theta}}(1) - \Lambda_{\tilde{\theta}}(0) \tag{5.95}$$

and $b^* = E[\zeta^*(Z^*_1, Z^*_1)|\omega]$, where $V_1, V_2, ...$ are independent and identically standard normally distributed and $\lambda^*_1, \lambda^*_2, ...$ are the eigenvalues of the operator

$$K^*\rho(z_1) := \int \rho(z_2)\zeta^*(z_1, z_2)\, dF_{Z^*}(z_2).$$

One can proceed as in the proof of Theorem 5.2.2 to obtain

$$P_1\Bigg(\omega \in \Omega_1 : P^1_2\Bigg(\omega, \Bigg\{\limsup_{m\to\infty}\Bigg|T^*_{n,m} - \frac{1}{m-1}\sum_{i=1}^{m}\sum_{\substack{j=1\\ j\neq i}}^{m} \zeta^*(Z^*_i, Z^*_j) - b^*\Bigg| > 0\Bigg\}\Bigg) > 0\Bigg) = o(1) \tag{5.96}$$

as well as

$$P_1\Big(\omega \in \Omega_1 : \limsup_{m\to\infty}\sup_{t\in\mathbb{R}}\big|P^1_2(\omega, \{T^*_{n,m} \le t\}) - P^1_2(\omega, \{T^*_n \le t\})\big| = 0\Big) = 1 + o(1),$$

that is, $T^*_{n,m}$ converges in distribution to T^*_n for $m \to \infty$.

Convergence of the bootstrap components: Note that neither ψ depends on h nor ψ^* on h^* (conditional on $(Y_i, X_i), i = 1, ..., n$). In the following, the convergence in probability of $h^*, R^*, \Gamma^*, \varphi^*, \zeta^*, b^*$ to h_0 (the true transformation under H_0), $R(s)$ from (5.19), Γ from (5.20), φ from (5.21), ζ from (5.22) and b from (5.74) is shown.
One has $\tilde{\theta} = \theta_0 + o_p(1)$. (5.38) implies $h^* = h + o_p(1)$ uniformly on compact sets and thus $w((h^*)^{-1}(s))R^*(s) = w(h^{-1}(s))R(s) + o_p(1)$ uniformly in $s \in \mathbb{R}$. Further, there exists some $\tilde{C} > 0$, such that h^* is bijective on $\mathcal{Y}_w$ and $|h^*| < \tilde{C}$ as well as $|(h^*)^{-1}| < \tilde{C}$ on $\mathcal{Y}_w$ and $h^*(\mathcal{Y}_w)$, respectively, with probability converging to one, that is

$$
\begin{aligned}
&P_1\Big(\omega \in \Omega_1 : \ h^* \text{ bijective}, \quad |h^*(y)| \leq \tilde{C} \ \forall y \in \mathcal{Y}_w, \quad |(h^*)^{-1}(s)| \leq \tilde{C} \ \forall y \in h^*(\mathcal{Y}_w)\Big) \\
&= 1 + o(1).
\end{aligned}
$$

This in turn means that (see Remark 5.4.3 for a possible adjustment of h^*)

$$
P_1\Big(\omega \in \Omega_1 : w((h^*)^{-1}(s))||R^*(s)||^2 \leq C \ \forall s \in \mathbb{R}\Big) = 1 + o(1)
$$

for some sufficiently large $C > 0$. Let f_S, f_{S^*} denote the densities of S and S^*, respectively, conditioned on $(Y_i, X_i), i = 1, ..., n$. The dominated convergence theorem leads to (the inequality is meant componentwise)

$$
\begin{aligned}
&P_1\big(\omega \in \Omega_1 : |\Gamma^* - \Gamma| > \delta\big) \\
&= P_1\big(\omega \in \Omega_1 : \big|E[w((h^*)^{-1}(S_1^*))R^*(S_1^*)R^*(S_1^*)^t|\omega] - E[w(h^{-1}(S_1))R(S_1)R(S_1)^t]\big| > \delta\big) \\
&= P_1\Big(\omega \in \Omega_1 : \Big|\int w((h^*)^{-1}(s))R^*(s)R^*(s)^t f_{S^*}(s)\, ds - \int w(h^{-1}(s))R(s)R(s)^t f_S(s)\, ds\Big| > \delta\Big) \\
&\leq P_1\Big(\omega \in \Omega_1 : \Big|\int w((h^*)^{-1}(s))R^*(s)R^*(s)^t f_{S^*}(s)\, ds - \int w(h^{-1}(s))R(s)R(s)^t f_S(s)\, ds\Big| > \delta, \\
&\qquad w((h^*)^{-1}(s))||R^*(s)||^2 \leq C \ \forall s \in \mathbb{R}\Big) + P_1\big(w((h^*)^{-1}(s))||R^*(s)||^2 > C \text{ for some } s \in \mathbb{R}\big) \\
&= o(1)
\end{aligned}
$$

for all $\delta > 0$. Consequently,

$$
\Gamma^* = \Gamma + o_p(1)
$$

with respect to P_1. Due to part (5.59) of (**A8***), the map

$$
(Z_1^*, U_2^*) \mapsto w((h^*)^{-1}(S_2^*))\psi^*(Z_1^*, U_2^*)
$$

is bounded by some constant C uniformly over compact sets with probability converging to one. Together with the dominated convergence theorem, this leads to boundedness of φ^* as well as $\varphi^*(z) = \varphi(z) + o_p(1)$ and finally boundedness of ζ^* and

$$
\begin{aligned}
\zeta^*(z_1, z_2) &= \int w((h^*)^{-1}(s))\big(\psi^*(z_1, \mathcal{T}_{S^*}(s)) - \varphi^*(z_1)^t(\Gamma^*)^{-1}R^*(s)\big) \\
&\qquad \big(\psi^*(z_2, \mathcal{T}_{S^*}(s)) - \varphi^*(z_2)^t(\Gamma^*)^{-1}R^*(s)\big) f_{S^*}(s)\, ds \\
&= \int w(h^{-1}(s))\big(\psi(z_1, \mathcal{T}_S(s)) - \varphi(z_1)^t\Gamma^{-1}R(s)\big)\big(\psi(z_2, \mathcal{T}_S(s))
\end{aligned}
$$

$$- \varphi(z_2)^t\Gamma^{-1}R(s)\big)f_S(s)\,ds + o_p(1)$$

$$= \zeta(z_1, z_2) + o_p(1)$$

for all $z_1, z_2 \in \mathbb{R}^{d_X+1}$. Additionally, one has

$$P_1\bigg(\omega \in \Omega_1 : \sup_{z_1,z_2\in\mathbb{R}^{d_X+1}} |\zeta^*(z_1, z_2)| > C\bigg) \to 0$$

for some $C > 0$, so that for all $\delta > 0$

$$\begin{aligned}
&P_1\big(\omega \in \Omega_1 : |b^* - b| > \delta\big)\\
&= P_1\big(\omega \in \Omega_1\big|E[\zeta^*(Z_1^*, Z_1^*)|\omega] - E[\zeta(Z_1, Z_1)]\big| > \delta\big)\\
&\leq P_1\bigg(\omega \in \Omega_1 : \big|E[\zeta^*(Z_1^*, Z_1^*)|\omega] - E[\zeta(Z_1, Z_1)]\big| > \delta, \sup_{z_1,z_2\in\mathbb{R}^{d_X+1}} |\zeta^*(z_1, z_2)| \leq C\bigg)\\
&\quad + P_1\bigg(\omega \in \Omega_1 : \sup_{z_1,z_2\in\mathbb{R}^{d_X+1}} |\zeta^*(z_1, z_2)| > C\bigg)\\
&= o(1),
\end{aligned}$$

that is, $b^* = b + o_p(1)$ with respect to P_1. Now, all ingredients to prove the convergence of the distribution functions $F_{T_n^*}(t) = P^*(T_n^* \leq t)$ to $P(T \leq t)$ in probability have been presented.

Convergence of the distribution functions: Let $t \in \mathbb{R}, \varepsilon > 0$ be arbitrary and $\tilde{\varepsilon}, \delta_{\tilde{\varepsilon}} > 0, M_{\tilde{\varepsilon}} \in \mathbb{N}$ such that

$$P(T > t - 2\tilde{\varepsilon}c_{1,0}^2) - P(T > t + 2\tilde{\varepsilon}c_{1,0}^2) + 2\tilde{\varepsilon} \leq \frac{\varepsilon}{2},$$

$$\sup_{s\in\mathbb{R}}\bigg|P\bigg(c_{1,0}^2\bigg(\frac{m}{\binom{m}{2}}\sum_{i=1}^{m}\sum_{j=i+1}^{m}\zeta(Z_i, Z_j) + b\bigg) > s\bigg) - P(T > s)\bigg| \leq \tilde{\varepsilon} \quad \text{for all } m \geq M_{\tilde{\varepsilon}} \tag{5.97}$$

and

$$|c_{1,0}^* - c_{1,0}| + |b^* - b| \leq \delta_{\tilde{\varepsilon}} \quad \Rightarrow \quad \bigg|\frac{t}{{c_{1,0}^*}^2} - b^* - \frac{t}{c_{1,0}^2} + b\bigg| \leq \tilde{\varepsilon}.$$

For a moment consider m as fixed and define

$$\mathcal{K}_{\tilde{\varepsilon},m} := \Big\{(z_1, z_2) \in \mathbb{R}^{2(d+1)} : |\zeta^*(z_1, z_2) - \zeta(z_1, z_2)| \leq \frac{\tilde{\varepsilon}}{m}\Big\}.$$

Then, $P((Z_1^*, Z_2^*) \in \mathcal{K}_{\tilde{\varepsilon},m}) \overset{n\to\infty}{\longrightarrow} 1$. For all $m \geq M_\varepsilon$ one has

$$\begin{aligned}
&P^*\bigg({c_{1,0}^*}^2\bigg(\frac{m}{\binom{m}{2}}\sum_{i=1}^{m}\sum_{j=i+1}^{m}\zeta^*(Z_i^*, Z_j^*) + b^*\bigg) > t\bigg)\\
&\leq P^*\bigg({c_{1,0}^*}^2\bigg(\frac{m}{\binom{m}{2}}\sum_{i=1}^{m}\sum_{j=i+1}^{m}\zeta^*(Z_i^*, Z_j^*) + b^*\bigg) > t\bigg)I_{\big\{(Z_i^*,Z_j^*)\in\mathcal{K}_{\tilde{\varepsilon},m}\ \forall i\neq j\in\{1,\ldots,m\}\big\}}\\
&\quad + I_{\big\{\exists i\neq j\in\{1,\ldots,m\}:(Z_i^*,Z_j^*)\notin\mathcal{K}_{\tilde{\varepsilon},m}\big\}}
\end{aligned}$$

$$\le P^*\bigg(\frac{m}{\binom{m}{2}}\sum_{i=1}^m\sum_{j=i+1}^m \zeta(Z_i^*,Z_j^*) > \frac{t}{c_{1,0}^2} - b - 2\tilde\varepsilon\bigg) + I_{\big\{|c_{1,0}^*-c_{1,0}|+|b^*-b|>\delta_{\tilde\varepsilon}\big\}}$$

$$+ I_{\big\{\exists i\neq j\in\{1,\dots,m\}:(Z_i^*,Z_j^*)\notin\mathcal{K}_{\tilde\varepsilon,m}\big\}}$$

$$\le \int\cdots\int I_{\big\{\frac{m}{\binom{m}{2}}\sum_{i=1}^m\sum_{j=i+1}^m \zeta(z_i,z_j)>\frac{t}{c_{1,0}^2}-b-2\tilde\varepsilon\big\}}\,dF_{Z^*}(z_1)\dots dF_{Z^*}(dz_m)$$

$$+ I_{\big\{|c_{1,0}^*-c_{1,0}|+|b^*-b|>\delta_{\tilde\varepsilon}\big\}} + I_{\big\{\exists i\neq j\in\{1,\dots,m\}:(Z_i^*,Z_j^*)\notin\mathcal{K}_{\tilde\varepsilon,m}\big\}}$$

$$= \int\cdots\int I_{\big\{\frac{m}{\binom{m}{2}}\sum_{i=1}^m\sum_{j=i+1}^m \zeta(z_i,z_j)>\frac{t}{c_{1,0}^2}-b-2\tilde\varepsilon\big\}}\,dF_{Z}(z_1)\dots dF_{Z}(dz_m) + o_p(1)$$

$$+ I_{\big\{|c_{1,0}^*-c_{1,0}|+|b^*-b|>\delta_{\tilde\varepsilon}\big\}} + I_{\big\{\exists i\neq j\in\{1,\dots,m\}:(Z_i^*,Z_j^*)\notin\mathcal{K}_{\tilde\varepsilon,m}\big\}}$$

$$\le P(T > t - 2\tilde\varepsilon c_{1,0}^2) + \tilde\varepsilon + o_p(1).$$

Here, the second to last equality follows from approximating the indicator function by the sum of the indicator functions of disjoint cubes of dimension $m(d_X+1)$ (remember $F_{Z^*}(z_1) - F_{Z^*}(z_2) = F_Z(z_1) - F_Z(z_2) + o_p(1)$ uniformly in $z_1, z_2 \in \mathbb{R}^{d_X+1}$ by (5.58)). The same reasoning leads to

$$P^*\bigg({c_{1,0}^*}^2\bigg(\frac{m}{\binom{m}{2}}\sum_{i=1}^m\sum_{j=i+1}^m \zeta^*(Z_i^*,Z_j^*)+b^*\bigg) > t\bigg) \ge P(T > t + 2\tilde\varepsilon c_{1,0}^2) - \tilde\varepsilon + o_p(1)$$

for all $\varepsilon > 0$ and thus

$$\bigg|P^*\bigg({c_{1,0}^*}^2\bigg(\frac{m}{\binom{m}{2}}\sum_{i=1}^m\sum_{j=i+1}^m \zeta^*(Z_i^*,Z_j^*)+b^*\bigg) \le t\bigg) - F_T(t)\bigg| \le \frac{\varepsilon}{2} + o_p(1). \tag{5.98}$$

Let $M \in \mathbb{N}$ and define

$$A_{M,\varepsilon} := \bigg\{\omega\in\Omega_1 : \forall m \ge M : \bigg|P^*\bigg({c_{1,0}^*}^2\bigg(\frac{m}{\binom{m}{2}}\sum_{i=1}^m\sum_{j=i+1}^m \zeta^*(Z_i^*,Z_j^*)+b^*\bigg) > t\bigg) - P^*(T_n^* > t)\bigg| < \varepsilon\bigg\}$$

as well as

$$B_n := \bigg\{\omega\in\Omega_1 : \limsup_{m\to\infty}\bigg|P^*\bigg({c_{1,0}^*}^2\bigg(\frac{m}{\binom{m}{2}}\sum_{i=1}^m\sum_{j=i+1}^m \zeta^*(Z_i^*,Z_j^*)+b^*\bigg) > t\bigg) - P^*(T_n^* > t)\bigg| = 0\bigg\}.$$

Note that $P_1(B_n) \overset{n\to\infty}{\longrightarrow} 1$ because of (5.96) and $P_1(B_n \cap A_{M,\varepsilon}) \overset{M\to\infty}{\longrightarrow} P_1(B_n)$. Let $N \in \mathbb{N}$ such that $P_1(B_N) \ge 1-\varepsilon$ and let M_ε fulfil (5.97) and $P_1(B_N) \le P_1(B_N \cap A_{M_\varepsilon,\varepsilon}) + \varepsilon$. Then, one has

$$\limsup_{n\to\infty} P_1(|F_{T_n^*}(t) - F_T(t)| > 3\varepsilon)$$

$$= \limsup_{n\to\infty} P_1\bigg(\limsup_{m\to\infty}\bigg|P^*\bigg({c_{1,0}^*}^2\bigg(\frac{m}{\binom{m}{2}}\sum_{i=1}^m\sum_{j=i+1}^m \zeta^*(Z_i^*,Z_j^*)+b^*\bigg) > t\bigg) - P(T>t)\bigg| > 3\varepsilon, B_n\bigg)$$

$$= \limsup_{n\to\infty}\lim_{M\to\infty} P_1\bigg(\sup_{m\ge M}\bigg|P^*\bigg({c_{1,0}^*}^2\bigg(\frac{m}{\binom{m}{2}}\sum_{i=1}^m\sum_{j=i+1}^m \zeta^*(Z_i^*,Z_j^*)+b^*\bigg) > t\bigg) - P(T>t)\bigg| > 3\varepsilon,$$

$$B_n \cap A_{M,\varepsilon}\Big)$$

$$\leq \limsup_{n\to\infty} \lim_{M\to\infty} P_1\bigg(\sup_{m\geq M}\bigg|P^*\bigg({c^*_{1,0}}^2\bigg(\frac{m}{\binom{m}{2}}\sum_{i=1}^{m}\sum_{j=i+1}^{m}\zeta^*(Z_i^*,Z_j^*)+b^*\bigg)>t\bigg)-P(T>t)\bigg|>3\varepsilon,$$

$$B_N \cap A_{M_\varepsilon,\varepsilon}\bigg)+2\varepsilon$$

$$\leq \limsup_{n\to\infty} P_1\bigg(\bigg|P^*\bigg({c^*_{1,0}}^2\bigg(\frac{M_\varepsilon}{\binom{M_\varepsilon}{2}}\sum_{i=1}^{M_\varepsilon}\sum_{j=i+1}^{M_\varepsilon}\zeta^*(Z_i^*,Z_j^*)+b^*\bigg)>t\bigg)-P(T>t)\bigg|>\varepsilon, B_N \cap A_{M_\varepsilon,\varepsilon}\bigg)$$

$$+2\varepsilon$$

$$\leq \limsup_{n\to\infty} P_1\bigg(\bigg|P^*\bigg({c^*_{1,0}}^2\bigg(\frac{M_\varepsilon}{\binom{M_\varepsilon}{2}}\sum_{i=1}^{M_\varepsilon}\sum_{j=i+1}^{M_\varepsilon}\zeta^*(Z_i^*,Z_j^*)+b^*\bigg)>t\bigg)-P(T>t)\bigg|>\varepsilon\bigg)+2\varepsilon$$

$$\overset{(5.98)}{=} 2\varepsilon.$$

Since $\varepsilon > 0$ can be chosen arbitrarily small, one has

$$P_1\big(|F_{T_n^*}(t)-F_T(t)|>\varepsilon\big)=o(1) \quad \text{for all } t\in\mathbb{R}, \varepsilon>0.$$

In total, (5.36) was proven, that is,

$$P_1\Big(\omega\in\Omega_1 : \limsup_{m\to\infty}\big|P_2^1(\omega,\{T^*_{n,m}\leq q\})-P(T_n\leq q)\big|>\delta\Big)=o(1) \quad \text{for all } \delta>0.$$

It remains to deduce

$$P_1\Big(\omega\in\Omega_1 : \limsup_{m\to\infty}|q^*_\alpha-q_\alpha|>\delta\Big)=o(1) \quad \text{for all } \delta>0$$

from this. Let $\delta > 0$ be arbitrarily small and let q_α be the α-quantile of T and define

$$\varepsilon_n=\min\left(\frac{\alpha-P(T_n\leq q_\alpha-\delta)}{2},\frac{P(T_n\leq q_\alpha+\delta)-\alpha}{2}\right)\to\varepsilon$$

for some $\varepsilon > 0$ and $n \to \infty$. Then, if

$$\big|P_2^1(\omega,\{T^*_{n,m}\leq q_\alpha-\delta\})-P(T_n\leq q_\alpha-\delta)\big|\leq\varepsilon_n,$$

one has

$$P_2^1(\omega,\{T^*_{n,m}\leq q_\alpha-\delta\})<\alpha,$$

that is, $q^*_\alpha > q_\alpha - \delta$, for sufficiently big n. Analogously, $\big|P_2^1(\omega,\{T^*_{n,m}\leq q_\alpha+\delta\})-P(T_n\leq q_\alpha+\delta)\big|\leq\varepsilon_n$ implies $P_2^1(\omega,\{T^*_{n,m}\leq q_\alpha+\delta\})>\alpha$ and $q^*_\alpha<q_\alpha+\delta$ for sufficiently big n, so that in total

$$P_1\Big(\omega\in\Omega_1 : \limsup_{m\to\infty}|q^*_\alpha-q_\alpha|\leq\delta\Big)$$

$$\geq P_1\Big(\omega\in\Omega_1 : \limsup_{m\to\infty}\big|P_2^1(\omega,\{T^*_{n,m}\leq q_\alpha-\delta\})-P(T_n\leq q_\alpha-\delta)\big|\leq\varepsilon_n,$$

$$\limsup_{m\to\infty}\big|P_2^1(\omega,\{T^*_{n,m}\leq q_\alpha+\delta\})-P(T_n\leq q_\alpha+\delta)\big|\leq\varepsilon_n\Big)+o(1)$$

$$\geq P_1\bigg(\omega \in \Omega_1 : \limsup_{m\to\infty} \big|P_2^1(\omega, \{T^*_{n,m} \leq q_\alpha - \delta\}) - P(T_n \leq q_\alpha - \delta)\big| \leq \frac{\varepsilon}{2},$$

$$\limsup_{m\to\infty} \big|P_2^1(\omega, \{T^*_{n,m} \leq q_\alpha + \delta\}) - P(T_n \leq q_\alpha + \delta)\big| \leq \frac{\varepsilon}{2}\bigg) + o(1)$$

$$= 1 + o(1).$$

□

5.8.8 Proof of Lemma 5.4.8

Only equation (5.57) needs to be proven, since the remaining conditions follow as in the proof of Lemma 5.4.6 by the results of Colling and Van Keilegom (2019). In contrast to the proof of Lemma 5.4.6, one has $\tilde{\theta} - \theta_0 = \mathcal{O}_p\big(n^{-\frac{1}{4}}\big)$ (here and in the following, o_p- and $\mathcal{O}_p$-terms are with respect to P_1 and for $n \to \infty$) and the asymptotic behaviour of ψ^* can not be reduced to the convergence to ψ. Nevertheless, ψ^* can be expressed as in (5.90) with $D^*_{p,0}, ..., D^*_{f,1}$ as in (5.88)–(5.89). The main idea is to prove uniform convergence of $\frac{\partial}{\partial u}\Phi^*$ and $\frac{\partial}{\partial u}\Phi^*$ on $\mathcal{U}_0 \times \mathrm{supp}(v)$ to $\frac{\partial}{\partial u}\tilde{\Phi}$ and $\frac{\partial}{\partial x_1}\tilde{\Phi}$, respectively, while the remaining parts of $\delta_j^{*\tilde{v}_1^*}, \delta_j^{*\tilde{v}_2^*}$ and $Q^{*\prime}$ are bounded in probability.
Due to (5.45) it holds that $\mathcal{U}_0 \subseteq \big(-\frac{F_S^B(0)}{F_S^B(1)-F_S^B(0)}, \frac{1-F_S^B(0)}{F_S^B(1)-F_S^B(0)}\big)$. In the following, it is proven that under the assumptions of Lemma 5.4.8 one has

$$\sup_{t\in\mathbb{R}} |\mathcal{T}_{S^*}(t) - \mathcal{T}_S^B(t)| = o_p(1), \tag{5.99}$$

$$\sup_{u\in\mathcal{U}_0} |(\mathcal{T}_{S^*})^{-1}(u) - (\mathcal{T}_S^B)^{-1}(u)| = o_p(1), \tag{5.100}$$

$$\sup_{(u,x)\in\mathcal{U}_0,\times\,\mathrm{supp}(v)} |\Phi^*(u,x) - \tilde{\Phi}(u,x)| = o_p(1), \tag{5.101}$$

$$\sup_{(u,x)\in\mathcal{U}_0,\times\,\mathrm{supp}(v)} \left|\frac{\partial}{\partial u}\Phi^*(u,x) - \frac{\partial}{\partial u}\tilde{\Phi}(u,x)\right| = o_p(1), \tag{5.102}$$

$$\sup_{(u,x)\in\mathcal{U}_0,\times\,\mathrm{supp}(v)} \left|\frac{\partial}{\partial x_1}\Phi^*(u,x) - \frac{\partial}{\partial x_1}\tilde{\Phi}(u,x)\right| = o_p(1). \tag{5.103}$$

Due to Lemma 5.4.4 F_{S^*} can be written for appropriate $u^*_{i,w} \in \mathbb{R}, i = 1, ..., n,\ w \in \mathrm{supp}(f_W)$, as

$$
\begin{aligned}
&F_{S^*}(u)\\
&= P^*\bigg(\hat{g}(X_j^* + b_nW_j) + \hat{\varepsilon}_{j_\varepsilon^*}(\tilde{\theta}) - \frac{1}{n}\sum_{l=1}^n \hat{\varepsilon}_l(\tilde{\theta}) + a_n\xi_j \leq u\bigg)\\
&= \frac{1}{n^2}\sum_{i=1}^n\sum_{k=1}^n P^*\bigg(\hat{g}(X_i + b_nW_j) + \hat{\varepsilon}_k(\tilde{\theta}) - \frac{1}{n}\sum_{l=1}^n \hat{\varepsilon}_l(\tilde{\theta}) + a_n\xi_j \leq u\bigg)\\
&= \frac{1}{n^2}\sum_{i=1}^n\sum_{k=1}^n \int F_\xi\bigg(\frac{u - \hat{g}(X_i + b_nw) - \hat{\varepsilon}_k(\tilde{\theta}) + \frac{1}{n}\sum_{l=1}^n \hat{\varepsilon}_l(\tilde{\theta})}{a_n}\bigg) f_W(w)\, dw
\end{aligned}
$$

$$= \frac{1}{n^2}\sum_{i=1}^n\sum_{k=1}^n \int F_\xi\bigg(\frac{u - g_{\theta_0}(X_i + b_n w) - \hat{\varepsilon}_k(\tilde{\theta})}{a_n}\bigg) f_W(w)\,dw$$

$$+ \frac{1}{n^2}\sum_{i=1}^n\sum_{k=1}^n \int f_\xi(u_{i,w}^*) f_W(w)\bigg(\frac{\hat{g}(X_i + b_n w) - g_{\theta_0}(X_i + b_n w) + \frac{1}{n}\sum_{l=1}^n \hat{\varepsilon}_l(\tilde{\theta})}{a_n}\bigg)\,dw$$

$$= \frac{1}{n^2}\sum_{i=1}^n\sum_{k=1}^n \int F_\xi\bigg(\frac{u - g_{\theta_0}(X_i + b_n w) - \varepsilon_k(\theta_0)}{a_n}\bigg) f_W(w)\,dw + o_p(1).$$

As a distribution function F_ξ is bounded so that

$$\operatorname{Var}\left(\frac{1}{n^2}\sum_{i=1}^n\sum_{k=1}^n \int F_\xi\bigg(\frac{u - g_{\theta_0}(X_i + b_n w) - \varepsilon_k(\theta_0)}{a_n}\bigg) f_W(w)\,dw\right) \to 0,$$

that is

$$F_{S^*}(u) = E\bigg[\int F_\xi\bigg(\frac{u - g_{\theta_0}(X_1 + b_n w) - \varepsilon_2(\theta_0)}{a_n}\bigg) f_W(w)\,dw\bigg] + o_p(1)$$

$$= \int\int E\bigg[F_\xi\bigg(\frac{u - g_{\theta_0}(x + b_n w) - \varepsilon_2(\theta_0)}{a_n}\bigg)\bigg] f_W(w) f_X(x)\,dx\,dw + o_p(1)$$

$$= \int\int E\big[I_{\{u - g_{\theta_0}(x + b_n w) - \varepsilon_2(\theta_0) \geq 0\}}\big] f_W(w) f_X(x)\,dx\,dw + o_p(1)$$

$$= \int\int F_{\varepsilon(\theta_0)}\big(u - g_{\theta_0}(x + b_n w)\big) f_W(w) f_X(x)\,dw\,dx + o_p(1)$$

$$= \int\int F_{\varepsilon(\theta_0)}\big(u - g_{\theta_0}(x)\big) f_W(w) f_X(x)\,dw\,dx + o_p(1).$$

Since F_{S^*} and F_S^B are distribution functions, this leads to the uniform convergence

$$\sup_{t\in\mathbb{R}} |F_{S^*}(t) - F_S^B(t)| = o_p(1)$$

and thus to (5.99). To prove (5.100) write $F_{\varepsilon(\theta_0)}$ as

$$F_{\varepsilon(\theta_0)}(e) = P(\varepsilon(\theta_0) \leq e)$$

$$= P\big(g(X) + \varepsilon \leq h(h_{\theta_0}^{-1}(e + g_{\theta_0}(X)))\big)$$

$$= \int F_\varepsilon\big(h(h_{\theta_0}^{-1}(e + g_{\theta_0}(x))) - g(x)\big) f_X(x)\,dx,$$

which implies

$$F_S^B(u) = \int\int F_\varepsilon\big(h(h_{\theta_0}^{-1}(u - g_{\theta_0}(z) + g_{\theta_0}(x))) - g(x)\big) f_X(x) f_X(z)\,dx\,dz.$$

for

$$h_{\theta_0}^{-1}(u) = \left\{\begin{matrix} -\infty \\ \infty \end{matrix}\right\}, \quad \text{if} \quad h_{\theta_0}(y)\left\{\begin{matrix} > \\ < \end{matrix}\right\} u \quad \text{for all } y \in \mathbb{R}.$$

Since h and h_{θ_0} are strictly increasing and F_S^B is continuous, one has $F_S^B(u_1) < F_S^B(u_2)$ for all $u_1 < u_2 \in (\mathcal{T}_S^B)^{-1}(\mathcal{U}_0) \subseteq (\mathcal{T}_S^B)^{-1}\Big(\big(-\frac{F_S^B(0)}{F_S^B(1) - F_S^B(0)}, \frac{1 - F_S^B(0)}{F_S^B(1) - F_S^B(0)}\big)\Big)$. Especially, $(\mathcal{T}_S^B)^{-1}$

is strictly increasing on $\mathcal{U}_0$, that is, (5.100) follows from (5.99).
Finally, this can be used to obtain

$$\begin{aligned}
\Phi_i^*(u|x) &= \frac{\partial}{\partial x_1} P^*(U_j^* \leq u | X_j^* = x) \\
&= \frac{\partial}{\partial x_1} P^*\Big(\hat{g}(X_j^*) + \hat{\varepsilon}_{j_\varepsilon^*}(\tilde{\theta}) - \frac{1}{n}\sum_{l=1}^n \hat{\varepsilon}_l(\tilde{\theta}) + a_n\xi \leq \mathcal{T}_{S^*}^{-1}(u) \big| X_j^* = x\Big) \\
&= \frac{1}{n}\sum_{k=1}^n \frac{\partial}{\partial x_1} P^*\Big(\hat{g}(x) + \hat{\varepsilon}_k(\tilde{\theta}) - \frac{1}{n}\sum_{l=1}^n \hat{\varepsilon}_l(\tilde{\theta}) + a_n\xi \leq \mathcal{T}_{S^*}^{-1}(u)\Big) \\
&= \frac{1}{a_n n}\sum_{k=1}^n f_\xi\left(\frac{\mathcal{T}_{S^*}^{-1}(u) - \hat{g}(x) + \frac{1}{n}\sum_{l=1}^n \hat{\varepsilon}_l(\tilde{\theta}) - \hat{\varepsilon}_k(\tilde{\theta})}{a_n}\right)\frac{\partial}{\partial x_1}\hat{g}(x) \\
&= -\frac{\partial}{\partial x_1}\hat{g}(x)\frac{1}{na_n}\sum_{k=1}^n f_\xi\left(\frac{\mathcal{T}_{S^*}^{-1}(u) - \hat{g}(x) + \frac{1}{n}\sum_{l=1}^n \hat{\varepsilon}_l(\tilde{\theta}) - \varepsilon_k(\theta_0)}{a_n}\right) + o_p(1) \\
&= -\frac{\frac{\partial}{\partial x_1}\hat{g}(x)}{a_n}\int f_\xi\left(\frac{\mathcal{T}_{S^*}^{-1}(u) - \hat{g}(x) + \frac{1}{n}\sum_{l=1}^n \hat{\varepsilon}_l(\tilde{\theta}) - e}{a_n}\right) f_{\varepsilon(\theta_0)}(e)\,de + o_p(1) \\
&= -\frac{\partial}{\partial x_1}\hat{g}(x)\int f_\xi(e) f_{\varepsilon(\theta_0)}\Big(\mathcal{T}_{S^*}^{-1}(u) - \hat{g}(x) + \frac{1}{n}\sum_{l=1}^n \hat{\varepsilon}_l(\tilde{\theta}) - a_n e\Big)\,de + o_p(1) \\
&= -\frac{\partial}{\partial x_1} g_{\theta_0}(x)\int f_\xi(e) f_{\varepsilon(\theta_0)}\big((\mathcal{T}_S^B)^{-1}(u) - g_{\theta_0}(x)\big)\,de + o_p(1) \\
&= -\frac{\partial}{\partial x_1} g_{\theta_0}(x) f_{\varepsilon(\theta_0)}\big((\mathcal{T}_S^B)^{-1}(u) - g_{\theta_0}(x)\big) + o_p(1)
\end{aligned}$$

uniformly in $(u, x) \in \mathcal{U}_0 \times \mathrm{supp}(v)$, where the second last equality follows from the continuity of $f_{\varepsilon(\theta_0)}$. The bootstrap functions

$$\frac{\partial}{\partial u} F_{S^*}(u) = \frac{1}{n^2 a_n}\sum_{i=1}^n\sum_{k=1}^n \int f_\xi\left(\frac{u - \hat{g}(X_i + b_n w) - \hat{\varepsilon}_k(\tilde{\theta}) + \frac{1}{n}\sum_{l=1}^n \hat{\varepsilon}_l(\tilde{\theta})}{a_n}\right) f_W(w)\,dw,$$

Φ^* and $\frac{\partial}{\partial u}\Phi^*$ can be treated by similar arguments to obtain

$$\sup_{(u,x)\in\mathcal{U}_0\times\mathrm{supp}(v)} |\Phi^*(u|x) - \tilde{\Phi}(u|x)| + \left|\frac{\partial}{\partial x_1}\Phi^*(u|x) - \frac{\partial}{\partial x_1}\tilde{\Phi}(u|x)\right| = o_p(1).$$

Since $Q^* = \mathcal{T}_{S^*}^{-1}$ equations (5.100) and (5.88)–(5.89) lead to

$$\sup_{y\in\mathbb{R}, z\in\mathbb{R}^{d_X+1}} |w(y)\psi^*(z, \mathcal{T}^*(y))| = \mathcal{O}_p(1).$$

□

5.8.9 Proof of Theorem 5.4.9

Borrowing the notations for R^*, Γ^* and φ^* from the proof of Theorem 5.4.7 recall that ζ^* can be written as

$$\zeta^*(z_1, z_2) := E^*\Big[w((h^*)^{-1}(S_3^*))\big(\psi^*(Z_1^*, U_3^*) - \varphi^*(Z_1^*)^t(\Gamma^*)^{-1}R^*(S_3^*)\big)$$

$$\big(\psi^*(Z_2^*, U_3^*) - \varphi^*(Z_2^*)^t(\Gamma^*)^{-1}R^*(S_3^*)\big) \mid Z_1^* = z_1, Z_2^* = z_2\Big].$$

Similar to the proof of Theorem 5.4.7 assumption (**A5'**) leads to (**A5***) from the proof of Theorem 5.4.7. As before, equation (5.96) remains valid, that is,

$$P_1\bigg(\omega \in \Omega_1 : P_2^1\bigg(\omega, \bigg\{\limsup_{m\to\infty}\bigg|T_{n,m}^* - \frac{1}{m-1}\sum_{i=1}^m \sum_{\substack{j=1\\ j\neq i}}^m \zeta^*(Z_i^*, Z_j^*) - b^*\bigg| > 0\bigg\}\bigg) > 0\bigg) = o(1)$$

for $b^* = E[\zeta^*(Z_1^*, Z_1^*)|\omega]$. Since (**A7***) ensures that

$$P_1\Big(\omega \in \Omega_1 : \sup_{y\in\mathcal{K}, z\in\mathbb{R}^{d_X+1}} |w(y)\psi^*(z, \mathcal{T}^*(y))| > \delta\sqrt{n}\Big) = o(1) \quad \text{for all } \delta > 0$$

and all compact sets $\mathcal{K} \subseteq \mathbb{R}$, one has

$$P_1\Big(\omega \in \Omega_1 : \sup_{z_1, z_2\in\mathbb{R}^{d_X+1}} |\zeta^*(z_1, z_2)| > \delta n\Big) = o(1) \quad \text{for all } \delta > 0$$

(boundedness of R^* and Γ^* follows as in the proof of Theorem 5.4.7). The same reasoning as in the proof of Theorem 5.2.2 leads to

$$E\bigg[\bigg(\frac{1}{m-1}\sum_{i=1}^m \sum_{\substack{j=1\\ j\neq i}}^m \zeta^*(Z_i^*, Z_j^*)\bigg)^2 \bigg|\omega\bigg] = \frac{2m^2}{(m-1)^2} E\big[\zeta^*(Z_1^*, Z_2^*)^2\,|\omega\big],$$

so that

$$\limsup_{m\to\infty}\bigg|\frac{1}{m-1}\sum_{i=1}^m \sum_{\substack{j=1\\ j\neq i}}^m \zeta^*(Z_i^*, Z_j^*) + b^*\bigg| = o_p(n)$$

and thus $\limsup_{m\to\infty}|T_{n,m}^*| = o_p(n)$ due to (5.96). Referring to Theorem 5.3.1, one has

$$\frac{1}{n}T_n = M(\gamma_0) + o_p(1)$$

with $M(\gamma_0) > 0$. Consequently,

$$\begin{aligned}
&P_1\Big(\omega \in \Omega_1 : \limsup_{m\to\infty} P_2^1\big(\omega, \{T_n \leq T_{n,m}^*\}\big) > 0\Big) \\
&= P_1\bigg(\omega \in \Omega_1 : \limsup_{m\to\infty} P_2^1\bigg(\omega, \bigg\{\frac{1}{n}T_n \leq \frac{1}{n}T_{n,m}^*\bigg\}\bigg) > 0\bigg) \\
&\leq P_1\bigg(\omega \in \Omega_1 : \limsup_{m\to\infty} P_2^1\bigg(\omega, \bigg\{\frac{M(\gamma_0)}{2} \leq \frac{1}{n}T_{n,m}^*\bigg\}\bigg) > 0\bigg) + P_1\bigg(\omega \in \Omega_1 : \frac{T_n}{n} < \frac{M(\gamma_0)}{2}\bigg) \\
&= o(1).
\end{aligned}$$

In total, (5.37) was proven, that is,

$$P_1\Big(\omega \in \Omega_1 : \limsup_{m\to\infty} P_2^1(\omega, \{T_n \leq T_{n,m}^*\}) > \delta\Big) = o(1) \quad \text{for all } \delta > 0.$$

It remains to prove

$$P_1\Big(\omega \in \Omega_1 : T_n > \limsup_{m\to\infty} q_\alpha^*\Big) = 1 + o(1)$$

for all $\alpha \in (0,1)$. Let $\alpha \in (0,1)$. As was shown above, one has

$$P_1\left(\omega \in \Omega_1 : \limsup_{m\to\infty} P_2^1\left(\omega, \left\{\frac{nM(\gamma_0)}{2} \leq T^*_{n,m}\right\}\right) > 0\right) = o(1).$$

Therefore,

$$\begin{aligned} P_1\left(\omega \in \Omega_1 : T_n \leq \limsup_{m\to\infty} q^*_\alpha\right) &\leq P_1\left(\omega \in \Omega_1 : \frac{nM(\gamma_0)}{2} < \limsup_{m\to\infty} q^*_\alpha\right) \\ &\quad + P_1\left(\omega \in \Omega_1 : \frac{T_n}{n} \leq \frac{M(\gamma_0)}{2}\right) \\ &= o(1). \end{aligned}$$

□

5.9 Miscellaneous

Finally, some additional statements concerning the case of a finite parameter space as well as some thoughts about related issues and possible extensions are given.

5.9.1 Finite Transformation Parameter Sets

In this section, the case of a finite parameter set Θ is considered. Hence, when testing for

$$H_0 : h \in \left\{\frac{\Lambda_\theta - \Lambda_\theta(0)}{\Lambda_\theta(0) - \Lambda_\theta(0)} : \theta \in \Theta\right\}$$

the previous estimation of the transformation parameter now becomes a classification. Further, it is no longer possible to take any derivatives with respect to θ, but as will be seen in the following, this is no longer necessary as well.
Recall that T_n (and analogously the corresponding minimizer $\tilde{\theta}$) was defined as

$$T_n = \min_{c_1,c_2,\theta} \sum_{j=1}^n w(Y_j)(\hat{h}(Y_j)c_1 + c_2 - \Lambda_\theta(Y_j))^2.$$

If $d(\Lambda_{\theta_1}, \Lambda_{\theta_2}) > 0$ for all $\theta_1, \theta_2 \in \Theta$ with $\theta_1 \neq \theta_2$, one can show by the same reasoning as in the proof of Theorem 5.2.6 that even under the local alternative (5.14) the corresponding minimizer $\tilde{\theta}$ converges in probability to θ_0. Since convergence in this context means equality, the transformation parameter can be considered as known when calculating the asymptotic distribution as in Theorem 5.2.2. Thus, T_n can be written as

$$T_n = \min_{c_1,c_2} \sum_{j=1}^n w(Y_j)(\hat{h}(Y_j)c_1 + c_2 - \Lambda_{\theta_0}(Y_j))^2 + o_p(1).$$

Now one can proceed similarly to Section 5.2 except for the need of estimating the transformation parameter. This leads to

$$R(s) = (s,1)^t, \tag{5.104}$$

while all remaining components can be defined as in Section 5.2 (apart from using definition (5.104) instead of (5.19)). After inserting these quantities in Theorem 5.2.2 the asymptotic distribution in (5.25) again holds. Alternatively (although a bit imprecisely), setting all derivatives with respect to θ equal to zero and "inverting" only the upper left 2×2-matrix of the original Γ leads to the same result.

5.9.2 Testing in a Heteroscedastic Model

Recall the considered model and the notations of Chapter 4. Here, some thoughts about how to adapt the test statistic and the heteroscedastic estimators in (4.20) or (4.21) to obtain a testing procedure for either the null hypothesis of a parametric transformation or the precise hypotheses as in (5.28) in the heteroscedastic transformation model

$$h(Y) = g(X) + \sigma(X)\varepsilon$$

for some variance function $\sigma > 0$ are presented.
Apart from some other technical assumptions on the nonparametric estimator of h, the main condition used in this chapter was the linear expansion (5.13)

$$\hat{h}(y) - h(y) = \frac{1}{n}\sum_{i=1}^{n}\psi(Z_i, \mathcal{T}(y)) + o_p\Big(\frac{1}{\sqrt{n}}\Big).$$

In Chapter 4 it was shown that (under several conditions) the heteroscedastic estimator $\hat{h}$ of h fulfils (4.58), that is

$$\hat{h}(y) - h(y) = -h(y)\frac{1}{n}\sum_{i=1}^{n}\left(B\eta_i(y) + \int_{y_1}^{y}\frac{1}{\lambda(u)}\,du\ \psi_{\Gamma_2}(Y_i, X_i)\right) + o_p\left(\frac{1}{\sqrt{n}}\right)$$

(see Chapter 4 for details and the definition of the occurring components). As in the paper of Colling and Van Keilegom (2019), it should be possible to adjust the heteroscedastic estimator of h by first transforming the data with $\mathcal{T}$ and then applying the techniques of Chapter 4. The test statistic

$$T_n = \min_{\theta\in\Theta, c_1\in C_1, c_2\in C_2}\sum_{i=1}^{n} w(Y_i)(\hat{h}(Y_i)c_1 + c_2 - \Lambda_\theta(Y_i))^2.$$

for some appropriate compact sets $C_1 \subseteq \mathbb{R}^+, C_2 \subseteq \mathbb{R}$ would look as in the homoscedastic case then.
As mentioned in Chapter 4, equation (4.58) might not always be fulfilled and the nonparametric estimator in the heteroscedastic model may converge to the true h even under the null hypothesis with a rate slower than $\mathcal{O}_p(\frac{1}{\sqrt{n}})$ (see for example parts (ii) and (iii) of Theorem 4.2.6 or Remark 4.2.3). This slower convergence of the estimator might also induce some additional factor to standardize the test statistic. In the situation of part (iii) in Theorem 4.2.6, the process given by $\sqrt{nh_y^3}(\tilde{h}(y) - h(y))$ (for some bandwidth h_y and the estimator $\tilde{h}$ from equation (4.17)) converges weakly to some centred Gaussian process. It is conjectured that the resulting test statistic would consequently look like

$$T_n = \min_{\theta\in\Theta, c_1\in C_1, c_2\in C_2} h_y^3\sum_{i=1}^{n} w(Y_i)(\tilde{h}(Y_i)c_1 + c_2 - \Lambda_\theta(Y_i))^2.$$

Remark 5.9.1 *The convergence rate of the nonparametric estimator $\hat{h}(y)$ in the heteroscedastic model is worse for $y \in (-\infty, y_0)$ (see Theorem 4.2.11), where y_0 was defined in Remark 3.2.1. Note that this problem can be circumvented by choosing the weighting function such that its (compact) support is contained in (y_0, ∞).*

5.9.3 Basing the Weights on Pretransformed Data

Not only the nonparametric estimation of the transformation function can be based on the pretransformed data $\hat{U}_i = \hat{\mathcal{T}}(Y_i), i = 1, ..., n$, but the weighting as well, leading to the test statistic

$$T_n = \min_{\theta \in \Theta, c_1 \in C_1, c_2 \in C_2} \sum_{i=1}^{n} w(\hat{U}_i)(\hat{h}(Y_i)c_1 + c_2 - \Lambda_\theta(Y_i))^2$$
$$= \min_{\theta \in \Theta, c_1 \in C_1, c_2 \in C_2} \sum_{i=1}^{n} w(\hat{U}_i)(\hat{Q}(\hat{U}_i)c_1 + c_2 - \Lambda_\theta(Y_i))^2,$$

where $\hat{Q}$ was defined in (5.46) and $C_1 \subseteq \mathbb{R}^+, C_2 \subseteq \mathbb{R}$ are some appropriate compact sets. Due to

$$\min_{\theta \in \Theta, c_1 \in C_1, c_2 \in C_2} \sum_{i=1}^{n} w(\hat{U}_i)(\hat{h}(Y_i)c_1 + c_2 - \Lambda_\theta(Y_i))^2$$
$$= \min_{\theta \in \Theta, c_1 \in C_1, c_2 \in C_2} \left(1 + \mathcal{O}_p\left(\frac{1}{\sqrt{n}}\right)\right) \sum_{i=1}^{n} w(U_i)(\hat{h}(Y_i)c_1 + c_2 - \Lambda_\theta(Y_i))^2,$$

this would only have an influence on the asymptotic behaviour of T_n via the weighting term (that is, basically $w(h_0^{-1}(S))$ has to be replaced by $w(U)$ in (5.19)–(5.24)).
Similarly, one can construct a procedure for testing the null hypothesis of a parametric Q by applying the test statistic

$$T_n^Q = \min_{\theta \in \Theta, c_1 \in C_1, c_2 \in C_2} \sum_{i=1}^{n} w(\hat{U}_i)(\hat{Q}(\hat{U}_i)c_1 + c_2 - \Lambda_\theta(\hat{U}_i))^2.$$

It is conjectured that this adjustment does have an impact on the asymptotic distribution, since in contrast to before Q, only affects the test statistic via its estimator $\hat{Q}$ and no longer via the distribution of $Y_i, i = 1, ..., n$.

6

Conclusion and Outlook

Finally, some concluding remarks on this thesis as well as an outlook on possible future research topics are given in this chapter. Since more detailed discussions can be found in each of the previous chapters 2–5, the results of this thesis are viewed from a higher level perspective here.

In the author's opinion, the main contribution of this thesis consists rather in the introduction of new ideas and concepts in the context of transformation models and the provision of corresponding tools to exploit them than in the detailed examination of each of the considered models and approaches. The first hypothesis test in the literature for a parametric regression model in a nonparametric transformation model has been presented in Chapter 2. Chapter 3 has provided the first identification result for a nonparametric and heteroscedastic transformation model. Consequently, the first estimators for the transformation function in such models have been given in Chapter 4. In Chapter 5, relevant hypotheses have been considered for the first time in the literature in the context of nonparametric transformation models. Here, "relevant" means, that the null hypothesis is rejected if the transformation function is sufficiently close to the parametric class, which is tested for. Furthermore, a new bootstrap approach has been developed.

Nevertheless, note that none the research fields, which were treated in the chapters 2–5, is completed in the sense that each of them offers high potential for further adjustments. To mention only some examples, a bootstrap procedure would probably increase the performance of the test in Chapter 2 and extensions of the theory in Chapter 3 as suggested in Remark 3.2.6 as well as a further analysis of the performance of the nonparametric estimator in Chapter 4 would be desirable. Moreover, the implementation of a test for the relevant hypotheses in Chapter 5 and a corresponding simulation study would be worthwhile. See Sections 2.6, 3.3, 4.4 and 5.6 for more details on possible extensions and adjustments of the presented results. Furthermore, there are many links between the chapters. The nonparametric estimator of Chapter 4, e.g., can be used similarly to the theory of Chapters 2 and 5 to construct a test for a parametric regression or transformation function, respectively, in heteroscedastic transformation models.

Of course, there are various opportunities for future research activities, which have not been considered in this thesis in detail yet. In the remainder, two of such potential research topics are explained briefly.

Arguably two of the most important parametric estimators of the transformation parameter θ_0 in the semiparametric model (1.10) have been considered in Section 1.3. The so called "mean square distance from independence estimator" (MDE) and the "profile likelihood estimator" (PLE) were developed by Linton et al. (2008). Although it seems that in most cases the PLE performs better than the MDE, it is mentioned there that at least for small sample sizes none of the estimators seems to outperform the other. Nevertheless, the PLE is used much more often in the literature.
As was described in Section 1.3, both of the estimators use the fact that (under some identification constraints)

$$\varepsilon_\theta := \Lambda_\theta(Y) - E[\Lambda_\theta(Y)|X]$$

is independent from X if and only if $\theta = \theta_0$ for the true transformation parameter θ_0. Note that the same is true for

$$\delta_\theta := \Lambda_\theta(Y) - \mathrm{median}(\Lambda_\theta(Y)|X).$$

There are two main arguments for applying the conditional median instead of the conditional mean. First, estimation of the mean is much more sensitive to "inconvenient" data than the median, e.g., when the error terms are highly skewed. Second, since the transformation functions are assumed to be strictly increasing, it is

$$\mathrm{median}(\Lambda_\theta(Y)|X) = \Lambda_\theta(\mathrm{median}(Y|X)).$$

Therefore, the median has to be estimated only once for all $\theta \in \Theta$. Both arguments may induce robustness and reduce computation costs. (Linton et al., 2008) mentioned that the PLE typically has a smaller variance than the MDE whereas the MDE has a smaller bias. Applying the conditional median instead of the conditional mean may reduce the variance so that it no longer has to be the case that the PLE outperforms the MDE. It is worthwhile to examine if an estimator, which is similar to the MDE, but based on the conditional median instead of the conditional mean, may even outperform the PLE.
So far, finite dimensional models have been considered. Nevertheless, for various reasons it is desirable to apply transformation models to functional explanatory variables as well. Since the functional structure increases the complexity of the model severely, a starting point for treating functional data could consist in the linear model

$$h(Y) = \langle X, \beta \rangle + \varepsilon,$$

where X is a random variable attaining values in some Hilbert space (e.g. $L^2(\mathbb{R})$). As functional transformation models to the author's knowledge have not been considered yet, there are various possibilities for further research activities. For example, the transformation could be estimated parametrically or nonparametrically. Among others, Cardot, Ferraty, and Sarda (1999) and Cardot, Mas, and Sarda (2007) dealt with functional (linear) models in the case without transformations.

References

M. G. Akritas and I. Van Keilegom. Non-parametric estimation of the residual distribution. *Scandinavian Journal of Statistics*, 28(3):549–567, 2001.

J. S. Allison, M. Hušková, and S. G. Meintanis. Testing the adequacy of semiparametric transformation models. *TEST*, 27(1):70–94, 2018.

R. Bellman. *Stability theory of differential equations*. Dover Publications, 1953.

J. O. Berger and M. Delampady. Testing precise hypotheses. *Statistical Science*, 2(3): 317–335, 1987.

P. J. Bickel and K. A. Doksum. An analysis of transformations revisited. *Journal of the American Statistical Association*, 76:296–311, 1981.

H. J. Bierens. Consistent model specification tests. *Journal of Econometrics*, 20:105–134, 1982.

H. J. Bierens and D. K. Ginter. Integrated conditional moment testing of quantile regression models. *Empirical Economics*, 26:307–324, 2001.

G. E. P. Box and D. R. Cox. An analysis of transformations. *Journal of the Royal Statistical Society. Series B*, 26(2):211–252, 1964.

H. Cardot, F. Ferraty, and P. Sarda. Functional linear model. *Statistics & Probability Letters*, 45:11–22, 1999.

H. Cardot, A. Mas, and P. Sarda. Clt in functional linear regression models. *Probability Theory and Related Fields*, 138:325–361, 2007.

R. J. Carroll and D. Ruppert. Power transformations when fitting theoretical models to data. *Journal of the American Statistical Association*, 79(386):321–328, 1984.

P. Chaudhuri. Global nonparametric estimation of conditional quantile functions and their derivatives. *Journal of Multivariate Analysis*, 39:246–269, 1991.

S. Chen. Rank estimation of transformation models. *Econometrica*, 70:1683–1697, 2002.

X. Chen, O. Linton, and I. Van Keilegom. Estimation of semiparametric models when the criterion function is not smooth. *Econometrica*, 71(5):1591–1608, 2003.

S. C. Cheng, L. J. Wei, and Z. Ying. Analysis of transformation models with censored data. *Biometrika*, 82(4):835–845, 1995.

P.-A. Chiappori, I. Komunjer, and D. Kristensen. Nonparametric identification and estimation of transformation. *Journal of Econometrics*, 188(1):22–39, 2015.

B. Colling and I. Van Keilegom. Goodness-of-fit tests in semiparametric transformation models. *TEST*, 25(2):291–308, 2016.

B. Colling and I. Van Keilegom. Goodness-of-fit tests in semiparametric transformation models using the integrated regression function. *Journal of Multivariate Analysis*, 160: 10–30, 2017.

B. Colling and I. Van Keilegom. Estimation of a semiparametric transformation model: a novel approach based on least squares minimization. preprint KU Leuven available at https://limo.libis.be/primo-explore/search?vid=Lirias, 2018.

B. Colling and I. Van Keilegom. Estimation of fully nonparametric transformation models. Bernoulli, to appear, 2019.

P. De Jong. A central limit theorem for generalized quadratic forms. *Probability Theory and Related Fields*, 75:261–277, 1987.

H. Dette and S. Volgushev. Non-crossing non-parametric estimates of quantile curves. *Journal of the Royal Statistical Society*, 70(3):609–627, 2008.

H. Dette, K. Kokot, and S. Volgushev. Testing relevant hypotheses in functional time series via self-normalization. preprint available at https://arxiv.org/abs/1809.06092, 2018.

L. Devroye. On the almost everywhere convergence of nonparametric regression function estimates. *The Annals of Statistics*, 9(6):1310–1319, 1981.

I. Ekeland, J. J. Heckman, and L. Nesheim. Identification and estimation of hedonic models. *Journal of Political Economy*, 112(1):60–109, 2004.

J.C. Escanciano. A consistent diagnostic test for regression models using projections. *Econometric Theory*, 22:1030–1051, 2006.

O. Forster. *Analysis 2*, volume 4. Vieweg, 1999.

W. González-Manteiga and R. M. Crujeiras. An updated review of goodness-of-fit tests for regression models. *TEST*, 22:361–411, 2013.

T. Gørgens and J. L. Horowitz. Semiparametric estimation of a censored regression model with an unknown transformation of the dependent variable. *Journal of Econometrics*, 90:155–191, 1999.

T. H. Grönwall. Note on the derivatives with respect to a parameter of the solutions of a system of differential equations. *Annals of Mathematics*, 20(4):292–296, 1919.

M. Guhlich. *Tests auf Modellannahmen in der nichparametrischen Quantilsregression*. PhD thesis, Ruhr Universität Bochum, 2013. available at http://www-brs.ub.ruhr-uni-bochum.de/netahtml/HSS/Diss/GuhlichMatthias/diss.pdf.

P. Hall, R. C. L. Wolff, and Q. Yao. Methods for estimating a conditional distribution function. *Journal of the American Statistical Association*, 94(445):154–163, 1999.

B. E. Hansen. Nonparametric estimation of smooth conditional distributions. manuscript, available at https://pdfs.semanticscholar.org/8162/c193d16c1d1bafe12785d3e5afe080be3dd7.pdf, 2004.

B. E. Hansen. Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, 24(3):726–748, 2008.

B. E. Hansen. Lecture notes on nonparametrics, 2009. available at https://www.ssc.wisc.edu/ bhansen/718/NonParametrics1.pdf.

W. Härdle and E. Mammen. Nonparametric versus parametric regression fits. *The Annals of Statistics*, 21(4):1926–1947, 1993.

X. He and L.-X. Zhu. A lack-of-fit test for quantile regression. *Journal of the American Statistical Association*, 98(464):1013–1022, 2003.

C. Heuchenne, R. Samb, and I. Van Keilegom. Estimating the error distribution in semiparametric transformation models. *Electronic Journal of Statistics*, 9:2391–2419, 2015.

Z. Hlávka, M. Hušková, C. Kirch, and S.G. Meintanis. Fourier-type tests involving martingale difference processes. *Econometric Reviews*, 36(4):468–492, 2017.

W. Hoeffding. A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, 19(3):293–325, 1948.

J. L. Horowitz. Semiparametric estimation of a regression model with an unknown transformation of the dependent variable. *Econometrica*, 64(1):103–137, 1996.

J. L. Horowitz. *Semiparametric and nonparametric methods in econometrics*. Springer, 2009.

J. L. Horowitz and S. Lee. Nonparametric estimation of an additive quantile regression model. *Journal of the American Statistical Association*, 100(472):1238–1249, 2005.

J. L. Horowitz and S. Lee. Nonparametric instrumental variables estimation of a quantile regression model. *Econometrica*, 75(4):1191–1208, 2007.

J. L. Horowitz and S. Lee. Testing a parametric quantile-regression model with an endogenous explanatory variable against a nonparametric alternative. *Journal of Econometrics*, 152:141–152, 2009.

J. L. Horowitz and V. G. Spokoiny. An adaptive, rate-optimal test of linearity for median regression models. *Journal of the American Statistical Association*, 97(459):822–835, 2002.

M. Hušková, S. G. Meintanis, N. Neumeyer, and C. Pretorius. Independence tests in semiparametric transformation models. *South African Journal of Statistics*, 52:1–13, 2018.

M. Hušková, S. G. Meintanis, and C. Pretorius. Tests for validity of the semiparametric heteroskedastic transformation model. preprint available at https://arxiv.org/abs/1901.02744, 2019.

M. D. Jiménez Gamero, J. Muñoz García, and R. Pino Mejías. Testing goodness of fit for the distribution of errors in multivariate linear models. *Journal of Multivariate Analysis*, 95:301–322, 2005.

K. Jochmans. Pairwise-comparison estimation with non-parametric controls. *Econometrics Journal*, 16:340–372, 2013.

J. John and N. Draper. An alternative family of transformations. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 29(2):190–197, 1980.

M. C. Jones and A. Pewsey. Sinh-arcsinh distributions. *Biometrica*, 96(4):761–780, 2009.

S. Klaaßen, J. Kück, and M. Spindler. Transformation models in high dimensions. preprint available at https://arxiv.org/pdf/1712.07364.pdf, 2017.

N. Kloodt and N. Neumeyer. Specification tests in semiparametric transformation models - a multiplier bootstrap approach. preprint available at https://arxiv.org/pdf/1709.06855.pdf, 2019.

N. Kloodt, N. Neumeyer, and I. Van Keilegom. Specification testing in semi-parametric transformation models. preprint available at https://arxiv.org/pdf/1907.01223.pdf, 2019.

R. Koenker. *Quantile Regression*. Cambridge University Press, 2005.

H. L. Koul. *Weighted empirical processes in dynamic nonlinear models*. Springer, 2002.

D. Lakens. Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, 8(4):355–362, 2017.

P. Lavergne, S. Maistre, and V. Patilea. A significance test for covariates in nonparametric regression. *Electronic Journal of Statistics*, 9:643–678, 2015.

A. J. Lee. *U-Statistics: theory and practice*. Dekker, 1990.

O. Linton, R. Chen, N. Wang, and W. Härdle. An analysis of transformations for additive nonparametric regression. *Journal of the American Statistical Association*, 92:1512–1521, 1997.

O. Linton, S. Sperlich, and I. Van Keilegom. Estimation of a semiparametric transformation model. *The Annals of Statistics*, 36(2):686–718, 2008.

R. Liu. Bootstrap procedures under some non-iid. models. *The Annals of Statistics*, 16(4): 1696–1708, 1988.

L. Michaelis and M. L. Menten. The kinetics of invertin action. *FEBS Letters*, 587:2712–2720, 1913.

Y. Mu and X. He. Power transformation toward a linear regression quantile. *Journal of the American Statistical Association*, 102(477):269–279, 2007.

E. A. Nadaraya. On estimating regression. *Theory of Probability and Its Applications*, 9 (1):141–142, 1964.

N. Neumeyer, H. Noh, and I. Van Keilegom. Heteroscedastic semiparametric transformation models: estimation and testing for validity. *Statistica Sinica*, 26:925–954, 2016.

R Core Team. R: A language and environment for statistical computing, 2017. http://www.r-project.org/.

D. Ruppert, N. Cressie, and R. J. Carroll. A transformation/weighting model for estimating michaelis-menten parameters. *Biometrics*, 45(2):637–656, 1989.

B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, 1986.

J. M. Stanton. Galton, pearson, and the peas: A brief history of linear regression for statistics instructors. *Journal of Statistics Education*, 9(3), 2001.

C. J. Stone. Consistent nonparametric regression. *The Annals of Statistics*, 5(4):595–645, 1977.

W. Stute. Nonparametric model checks for regression. *The Annals of Statistics*, 25(2): 613–641, 1997.

A. Szydłowski. Testing a parametric transformation model versus a nonparametric alternative. preprint available at https://www.le.ac.uk/economics/research/RePEc/lec/leecon/dp17-15.pdf, 2017.

G. J. Van den Berg. *Handbook of Economics*, volume 5, chapter 55. North Holland, 2001.

A. W. Van der Vaart and J. A. Wellner. *Weak convergence and empirical processes*. Springer, 1996.

I. Van Keilegom, W. González-Manteiga, and C. Sánchez Sellero. Goodness-of-fit tests in parametric regression based on the estimation of the error distribution. *TEST*, 17: 401–415, 2008.

A. Vanhems and I. Van Keilegom. Semiparametric transformation model with endogeneity: a control function approach. *Econometric Theory*, 2019. to appear.

M. P. Wand and M. C. Jones. *Kernel Smoothing.* Chapman & Hall, 1995.

Q. Wang and X. Wang. Analysis of censored data under heteroscedastic transformation regression models with unknown transformation function. *The Canadian Journal of Statistics*, 46(2):233–245, 2018.

G. S. Watson. Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics*, 26 (4):359–372, 1964.

J. A. Wellner. Empirical processes: theory and applications, 2005. available at https://www.stat.washington.edu/jaw/RESEARCH/TALKS/Delft/emp-proc-delft-big.pdf.

H. Wen, X. Bu, and L. Zhang. The application of box-cox transformation in selecting functional form for hedonic price models. *Applied Mechanics and Materials*, 357-360: 2869–2875, 2013.

H. Witting and U. Müller-Funk. *Mathematical statistics II - Asymptotic statistics: parametric models and nonparametric functionals.* B. G. Teubner, 1995.

I.-K. Yeo and R. A. Johnson. A new family of power transformations to improve normality or symmetry. *Biometrika*, 98(4):954–959, 2000.

K. Yu and M. C. Jones. A comparison of local constant and local linear regression quantile estimators. *Computational Statistics and Data Analysis*, 25:159–166, 1997.

A. Zellner and N. S. Revankar. Generalized production functions. *Review of Economic Studies*, 36(2):241–250, 1969.

J. X. Zheng. A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics*, 75(2):263–289, 1996.

J. X. Zheng. A consistent nonparametric test of parametric regression models under conditional quantile restrictions. *Econometric Theory*, 14(1):123–138, 1998.

J. X. Zheng. A consistent test of conditional parametric distributions. *Econometric Theory*, 16(5):667–691, 2000.

X.-H. Zhou, H. Lin, and E. Johnson. Non-parametric heteroscedastic transformation regression models for skewed data with an application to health care costs. *Journal of the Royal Statistical Society*, 70:1029–1047, 2009.

Appendix A

Formalities

A.1 Abstract

Predicting a real valued random variable Y depending on the observation of an $\mathbb{R}^{d_X}$-valued explanatory covariate is a classical issue in statistics. In this context, transformation models have attracted more and more attention over the last years and decades. Box and Cox (1964) already introduced transformations of the dependent variable to justify assumptions like normality, homoscedasticity or additivity of the error terms or to reduce skewness of the distribution of Y. The transformation model, which is considered throughout this thesis, can be written in its most general form as

$$h(Y) = g(X) + \sigma(X)\varepsilon, \tag{A.1}$$

where ε is assumed to be independent of X and h, g and σ^2 denote the transformation, regression and variance function, respectively. Apart from questions concerning solvability or uniqueness, statisticians especially deal with estimating and testing problems when looking at models like (A.1). Indeed, each of the four main chapters 2–5 can be classified into one of these three categories. If not specified differently, let $(Y_i, X_i), i = 1, ..., n$, denote independent and identically distributed observations from model (A.1).
Chapter 2 belongs to the last category and deals in the situation of homoscedastic errors with the question if the regression function g belongs to some parametric class $\{g_\beta : \beta \in B\}$. Here, the parameter space B is a subset of $\mathbb{R}^{d_B}$ for some $d_B \in \mathbb{N}$. In contrast to already existing approaches in the literature, the transformation function is modelled nonparametrically. This complicates the estimation of the regression function g as the conditional expectation and the asymptotic analysis of such an estimator insofar, as the existing uniform convergence results for nonparametric estimators $\hat{h}$ of the transformation function h in the literature are restricted to compact sets. Therefore, the hypothesis test of Chapter 2 is based on a comparison of the parametric class and an appropriate estimator of the conditional quantile function of Y given X instead of the estimated conditional expectation. This approach is motivated by two observations: On the one hand, the conditional quantile function and the regression function only differ by an additive constant. On the other hand, the applied estimator of the conditional distribution function of $h(Y)$ conditional on X only requires the nonparametric estimator of the transformation function h to

be evaluated on a compact set. For an appropriate bandwidth h_x, a finite measure μ on $(0,1)$ and an appropriate estimator $\hat{F}^{\hat{h}}_{Y|X}$ of the conditional distribution function of $h(Y)$ conditioned on X, the provided test statistic is given by

$$T_n^{\hat{h}} = nh_x^{\frac{d_X}{2}} \min_{\beta \in B} \int \min_{c \in \mathbb{R}} \int v(x)\big((\hat{F}^{\hat{h}}_{Y|X})^{-1}(\tau|x) - g_\beta(x) - c\big)^2 \, dx \, \mu(d\tau).$$

The asymptotic behaviour is similar to that of the hypothesis test by Härdle and Mammen (1993). First, a version of the test statistic in the nonparametric regression model is provided, before the approach is extended to transformation models. Afterwards, asymptotic normality of the test statistic is shown in both cases. It is found that the estimation of the transformation function does not influence the asymptotic distribution. Some of the techniques, which are applied in the proofs, are also used later in Chapter 5. Finally, two tests for each of the models with and without transforming the dependent variable are provided and the finite sample size behaviour is examined.

Chapter 3 addresses the question of uniqeness of the model components in (A.1), which is called identification of the model. Previous results on identification can not be applied on such a general model as (A.1) and mostly are restricted to homoscedastic errors. Similar to the approach of Chiappori et al. (2015), a quotient of partial derivatives of the conditional distribution function of Y conditional on X is considered. In contrast to the homoscedastic case, this quotient can not be simply integrated to obtain the transformation function, but leads to a differential equation, which can be solved uniquely under some common identification constraints. Moreover, an explicit expression of this solution is given.

Chapter 4 deals with the estimation of the transformation function h in model (A.1). The estimator presented there is based on the expression of h, which was derived in Chapter 3. Depending on the underlying set, uniform convergence rates up to weak convergence of $\sqrt{n}(\hat{h} - h)$ to a centred Gaussian process are proven. Additionally, the behaviour of the estimator for finite sample sizes is examined.

Under the assumption of a constant variance function σ^2, a test for the hypothesis, if the transformation function h belongs to a given class of transformation functions $\{\Lambda_\theta : \theta \in \Theta\}$ for some given parameter space $\Theta \subset \mathbb{R}^{d_\Theta}$ with $d_\Theta \in \mathbb{N}$, is developed in Chapter 5. The presented approach uses the ideas of Colling and Van Keilegom (2018) to construct a test statistic on the base of a comparison of a nonparametric estimator $\hat{h}$ for h and the parametric function class. The asymptotic behaviour of the resulting test statistic

$$\tilde{T}_n = \min_{c_1 \in \mathbb{R}^+, c_2 \in \mathbb{R}, \theta \in \Theta} \sum_{i=1}^n w(Y_i)(\hat{h}(Y_i)c_1 + c_2 - \Lambda_\theta(Y_i))^2$$

is examined and tests for the null hypothesis of a parametric transformation class well as for the null hypothesis of the transformation function not belonging to the parametric class are provided. Especially, the latter null hypothesis is interesting since in this case rejection of the null hypothesis yields evidence for the validity of the parametric model. Quite a sophisticated bootstrap algorithm is presented and the finite sample size behaviour of the corresponding test is examined in a simulation study.

Finally, the results of the thesis are summarized and discussed in Chapter 6 and a brief outlook of possible adjustments and extensions is given.

A.2 Zusammenfassung

Die Vorhersage einer reellwertigen Zufallsgröße Y in Abhängigkeit einer $\mathbb{R}^{d_X}$-wertigen, erklärenden Variable X stellt eine klassische Fragestellung in der Statistik dar. In den letzten Jahren und Jahrzehnten haben dabei sogenannte Transformationsmodelle immer weiter an Bedeutung gewonnen. Bereits Box und Cox (1964) führten Transformationen der abhängigen Variable ein, um Eigenschaften wie Normalität, Homoskedastizität oder Additivität der Fehler zu erreichen oder die Schiefe einer Verteilung zu reduzieren. Das in dieser Arbeit betrachtete Transformationsmodell lässt sich allgemein durch

$$h(Y) = g(X) + \sigma(X)\varepsilon \tag{A.2}$$

beschreiben, wobei Unabhängigkeit von ε und X angenommen wird und h, g und σ^2 die Transformations- und Regressions- beziehungsweise Varianzfunktion des Modells darstellen. Im Rahmen derartiger Modelle beschäftigt sich die Statistik abgesehen von Fragestellungen bezüglich der Lösbarkeit des Modells und der Eindeutigkeit dieser Lösungen insbesondere mit der Schätzung der Modellkomponenten und Hypothesentests, in obigem Fall beispielsweise auf eine parametrische Annahme an eine der drei genannten Funktionen. Tatsächlich lässt sich jedes der vier Hauptkapitel 2–5 dieser Arbeit in eine der drei Kategorien einordnen. Sofern nicht anders spezifiziert, seien $(Y_i, X_i), i = 1, ..., n$, unabhängige und identisch verteilte Realisierungen des Modells (A.2).

Kapitel 2 gehört zur letzten Kategorie und beschäftigt sich unter der Annahme einer konstanten Varianzfunktion σ^2 mit der Frage, ob die Regressionsfunktion g in einer parametrischen Klasse $\{g_\beta : \beta \in B\}$ enthalten ist. Dabei sei der Parameterraum B eine Teilmenge von $\mathbb{R}^{d_B}$ für ein $d_B \in \mathbb{N}$. Im Gegensatz zu bereits existierenden Tests in der Literatur werden keine parametrischen Annahmen an die Transformationsfunktion h gestellt. Dies schränkt die Schätzung der Regressionsfunktion als bedingten Erwartungswert beziehungsweise die asymptotische Behandlung eines solchen Schätzers insofern ein, als gleichmäßige Konvergenzresultate für nichtparametrische Schätzer $\hat{h}$ der Transformationsfunktion h in der Literatur bisher auf Kompakta beschränkt sind. Daher beruht der Hypothesentest in Kapitel 2 auf einem Vergleich der parametrischen Klasse mit der geschätzten bedingten Quantilfunktion von Y gegeben X anstelle der geschätzten bedingten Erwartung. Dieser Idee liegen zwei grundlegende Beobachtungen zugrunde: Zum einen unterscheiden sich die bedingte Quantilfunktion und die Regressionsfunktion nur um eine additive Konstante. Zum anderen bedarf es bei geeigneter Schätzung der bedingten Verteilungsfunktion von $h(Y)$ gegeben X nur der Auswertung des nichtparametrischen Schätzers $\hat{h}$ auf einem Kompaktum. Für eine geeignete Bandbreite h_x, ein endliches Maß μ auf $(0, 1)$ und einen geeigneten Schätzer $\hat{F}^{\hat{h}}_{Y|X}$ für die bedingte Verteilungsfunktion von $h(Y)$ gegeben X lässt sich die vorgestellte Teststatistik als

$$T_n^{\hat{h}} = nh_x^{\frac{d_X}{2}} \min_{\beta \in B} \int \min_{c \in \mathbb{R}} \int v(x)\big((\hat{F}^{\hat{h}}_{Y|X})^{-1}(\tau|x) - g_\beta(x) - c\big)^2 \, dx \, \mu(d\tau)$$

schreiben. Das asymptotische Verhalten erinnert an den von Härdle und Mammen (1993) entwickelten Hypothesentest. Zunächst wird eine Version der Teststatistik im nichtparametrischen Regressionsmodell vorgestellt, bevor der Testansatz auf Transformationsmo-

delle verallgemeinert wird. In beiden Fällen wird asymptotische Normalität der Teststatistik nachgewiesen, wobei sich die Schätzung der Transformationsfunktion h nicht auf die asymptotische Verteilung auswirkt. Einige der verwendeten Beweistechniken finden auch in Kapitel 5 Anwendung. Anschließend werden jeweilige Hypothesentests bereitgestellt und deren Verhalten für endliche Stichprobenumfänge untersucht.
Kapitel 3 behandelt die Frage der Eindeutigkeit der Modellkomponenten in dem Modell (A.2). Bisherige Resultate in der Literatur sind auf ein derart allgemeines Modell nicht anwendbar und meist auf homoskedastische Fehler beschränkt. Ähnlich zu dem Ansatz von Chiappori, Komunjer und Kristensen (2015) wird ein Quotient von partiellen Ableitungen der bedingten Verteilungsfunktion von Y gegeben X betrachtet. Allerdings kann dieser nun nicht mehr einfach integriert werden, um die Transformationsfunktion h zu erhalten, führt aber auf eine Differentialgleichung, deren Lösungen unter üblichen Identifizierbarkeitsbedingungen eindeutig ist. Ferner wird eine explizite Darstellung der Lösung angegeben.
Kapitel 4 beschäftigt sich mit der Schätzung der Transformationsfunktion h in dem Modell (A.2). Der vorgestellte Ansatz fußt wesentlich auf dem in Kapitel 3 herausgearbeiteten, expliziten Ausdruck für h. In Abhängigkeit von der betrachteten Menge werden gleichmäßige Konvergenzraten bis hin zur schwachen Konvergenz von $\sqrt{n}(\hat{h}-h)$ gegen einen zentrierten Gaußprozess bewiesen. Abschließend wird das Verhalten des Schätzers für endliche Stichprobenumfänge untersucht.
Kapitel 5 stellt wieder unter der Annahme einer konstanten Varianzfunktion σ^2 einen Hypothesentest bereit, allerdings diesmal für die Hypothese einer parametrischen Transformationsfunktion h, das heißt, ob h in einer gegebenen Klasse $\{\Lambda_\theta : \theta \in \Theta\}$ für einen Parameterraum $\Theta \subseteq \mathbb{R}^{d_\Theta}$ mit $d_\Theta \in \mathbb{N}$ liegt. Der vorgestellte Ansatz verwendet die Ideen von Colling und Van Keilegom (2018), um eine Teststatistik auf Basis eines Vergleichs eines nichtparametrischen Schätzers $\hat{h}$ für h mit der parametrischen Funktionenklasse zu konstruieren. Das asymptotische Verhalten der resultierenden Teststatistik

$$\tilde{T}_n = \min_{c_1\in\mathbb{R}^+, c_2\in\mathbb{R}, \theta\in\Theta} \sum_{i=1}^{n} w(Y_i)(\hat{h}(Y_i)c_1 + c_2 - \Lambda_\theta(Y_i))^2$$

wird untersucht und Tests sowohl für die Nullhypothese einer parametrischen Transformationsfunktion als auch für die Nullhypothese einer Transformation außerhalb der gegebenen Funktionenklasse werden bereitgestellt. Der Test für die letztere Hypothese ist insbesondere interessant, da eine Ablehnung der Nullhypothese Evidenz für die Gültigkeit des parametrischen Modells liefert. Ein recht ausgeklügelter Bootstrapalgorithmus wird vorgestellt und das Verhalten des zugehörigen Tests für endliche Stichprobenumfänge wird in einer Simulationsstudie untersucht.
Die Ergebnisse der Arbeit werden abschließend in Kapitel 6 zusammengefasst und bewertet. Außerdem wird ein Ausblick in mögliche Ergänzungen und Erweiterungen gegeben.

A.3 Publications Related to this Dissertation

The author is involved in two preprints related to transformation models:

- Kloodt and Neumeyer (2019)

- Kloodt, Neumeyer, and Van Keilegom (2019).

As mentioned in Section 2.2, Kloodt and Neumeyer (2019) contains previous work on goodness of fit tests for a parametric regression function as well as on a significance test for components of the covariate X in the semiparametric transformation model (2.5). The results have not been included in this thesis. A preprint can be found on https://arxiv.org/pdf/1709.06855.pdf.
Chapter 5 addressed the question of how to test for a parametric transformation function in nonparametric transformation models. The preprint of Kloodt et al. (2019) is based on the results of this Chapter and can be found on https://arxiv.org/pdf/1907.01223.pdf.

A.4 Eidesstattliche Versicherung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Nick Kloodt
Hamburg, den 15.10.2019