Universität Hamburg

Dissertation

Non-Commutative Probability Theory and Applications in Finance

Freie wissenschaftliche Arbeit in Form einer

Monographie

zur Erlangung des akademischen Grades eines

"Doctor rerum politicarum"

eingereicht bei der

Fakultät für Betriebswirtschaft
der Universität Hamburg

Hamburg, 19. Juli 2019

eingereicht von:
Marco Breitig
geboren am 16.06.1983 in Dresden

This thesis has been written by Marco Breitig 2010–2019 as partial fulfillment of the requirements for the degree of Doctor rerum politicarum.
It has been submitted on July 19, 2019.

Fakultät für Betriebswirtschaft
Universität Hamburg
Moorweidenstr. 18
20148 Hamburg, Germany

Chaiperson:	Prof. Dr. Knut Haase
First Examiner:	Prof. Dr. Peter Stahlecker
Second Examiner:	Prof. Dr. Wolfgang Drobetz
Date of Disputation:	January 23, 2020

Abstract

We introduce non-commutative probability theory as a tool to analyse sample covariance matrices. We develop the theory necessary for derivation of the spectral distribution of covariance matrix estimates of VARMA(p, q) random matrix models and introduce an extension to VARFIMA(p, d, q) random matrix models. The relationship between sample covariance matrices and there population counterparts are investigated. Specifically, we showcase efficient algorithms for calculating various VARMA(p, q) spectral densities.

Both model classes are implemented so that parameter estimation is possible. For a feasible subset of a high-dimensional data set of stock returns we estimate the model parameters for VARMA$(1, 1)$ random matrix models.

Wir führen in nicht-kommutative Wahrscheinlichkeitstheorie als ein Werkzeug zur Analyse von empirischen Kovarianzmatrizen ein. Wir entwickeln die notwendige Theorie zur Herleitung von Spektralverteilungen von Kovarianz-Matrix-Schätzern für VARMA(p, q) Zufallsmatrix-Modelle. Mit dieser erarbeiten wir eine Erweiterung auf VARFIMA(p, d, q) Zufallsmatrix-Modelle. Die Verbindung zwischen Kovarianz-Matrix-Schätzer und der wahren Kovarianzmatrix werden untersucht. Speziell präsentieren wir effiziente Algorithmen um diverse VARMA(p, q) Spektraldichten zu berechnen.

Beide Modellklassen werden implementiert, so dass Parameterschätzung möglich ist. Für eine praktikable Teilmenge von hochdimensionalen Finanzmarktdaten von Aktienreturns schätzen wir die Modellparameter für VARMA$(1, 1)$ Zufallsmatrix-Modelle.

JEL: C13, C51,

MSC: Primary 46L53, 46L54; Secondary 62H12, 60B20

Keywords: Non-Commutative Probability Theory, Covariance-Matrix Estimation, VARMA, VARFIMA

Dedicated to All my Teachers.

Table of Contents

Abstract i

Table of Contents v

List of Tables ix

List of Figures xi

Symbols and Acronyms xv
Symbols . . . xv
Acronyms and Abbreviations . . . xviii

1 Introduction 1
1.1 High-dimensional Data . . . 2
1.2 Covariance Estimation and Model Parameter Growth . . . 2
1.3 New Framework: Non-Commutative Probability Theory . . . 3
1.4 Simple Comparison with Market Data . . . 4
1.5 Overview . . . 5

2 Introduction to Non-Commutative Probability Theory and Free Independence 7
2.1 Introduction to Non-Commutative Probability Theory . . . 7
2.1.1 Basic Definitions . . . 8
2.1.2 Random Matrices . . . 9
2.2 Wishart Ensemble and Marchenko–Pastur Distribution . . . 12
2.2.1 Wishart Ensemble . . . 12
2.2.2 Marchenko–Pastur Distribution . . . 13
2.3 Free Independence . . . 16
2.3.1 Independence for Classical Random Variables . . . 16
2.3.2 Definition of Free Independence . . . 16
2.3.3 Asymptotic Freeness . . . 18
2.3.4 Free Deterministic Equivalents . . . 19
2.4 Transformations . . . 20
2.4.1 The $\mathcal{M}$-transform . . . 20
2.4.2 The Cauchy-Transform $\mathcal{G}$ . . . 21
2.4.3 $\mathcal{R}$-transform . . . 22
2.4.4 $\mathcal{S}$-transform . . . 23
2.5 Rectangular Random Matrices . . . 24
2.6 Scaling . . . 27

3 Eigenvalue Distributions for Specific Random Matrix Models 29
3.1 Spectral Distributions of Vector-ARMA-Ensembles . . . 29
3.1.1 Basic Definitions and Short Introduction to VARMA(p,q) Processes . . . 29
3.1.2 Assumptions on the General Covariance Matrix and Factorizations . . . 31
3.1.3 Connection of Sample-Covariance Matrix to $\mathcal{M}$-transform of Auto-Correlation Matrix . . . 33
3.1.4 The $\mathcal{M}$-transform of VARMA(p,q) Processes . . . 37
3.1.5 Calculation of distribution generating polynomials for various special VARMA(p,q) Models . . . 40
3.1.6 General Remarks on VARMA(p,q) Eigenvalue Density Distributions . . . 47
3.2 Spectral Distributions of Some Linear Stochastic Processes . . . 48
3.2.1 VARMA(1,1) processes . . . 49
3.2.2 Short Introduction to *Long-Range Dependence* . . . 52
3.2.3 ARFIMA(p,d,q) processes . . . 53
3.2.4 VARFIMA(0,d,0) processes . . . 56
3.2.5 VARFIMA(1,d,1) process . . . 58
3.3 Numerical Methods and Algorithms used . . . 61
3.3.1 Calculating the DGP for VARMA(p,q) processes . . . 61
3.3.2 Determining the right root of the DGP for VARMA(p,q) processes 61
3.3.3 Numerical Integration for VARFIMA(p,d,q) processes . . . 64

4 Overview of Data and Descriptive Statistics 69
4.1 Basic Methodology for Constructing Raw Data Sets . . . 69
4.2 Overview of Raw Data . . . 71
4.2.1 NASDAQ Intraday Data . . . 71
4.2.2 S&P 500 and DJIA Intraday Data . . . 72
4.3 Descriptive Statistics and Data Sets . . . 73
4.3.1 NASDAQ Intraday Data . . . 74
4.3.2 S&P 500 and DJIA Intraday Data . . . 81
4.4 Empirical Spectral Distributions . . . 88
4.4.1 Motivation . . . 88
4.4.2 Market Eigenvalue . . . 89
4.4.3 Plots of Empirical Spectral Distributions . . . 90

5 Application of Specific Random Matrix Models to Financial Returns 97
5.1 Parameter Estimation Methodology . . . 98
5.2 Numerical Minimization Methods . . . 100
5.2.1 VMA(1) and VAR(1) processes . . . 102
5.2.2 VARMA(1,1) processes . . . 105
5.2.3 VARFIMA(p,d,q) processes . . . 106
5.3 Parameter Estimates and Fit to Historical Data . . . 107
5.3.1 Parameter Estimates . . . 108
5.3.2 Fit of Spectral Density and ESD . . . 111
5.4 Reflections on the Results . . . 114

6 Conclusions and Perspectives **117**
6.1 Recapitulation and Conclusions . . . 117
6.2 Motivations and Restrictions . . . 119
6.3 Perspectives . . . 119

Appendices **123**

A Theorems, Proofs and Results **123**
A.1 Definitions . . . 123
A.2 Theorems . . . 123
A.2.1 Some Theorems in Complex Analysis . . . 123
A.3 Results . . . 127
A.3.1 Polynomial for VMA(2) . . . 127
A.3.2 Polynomial for VAR(2) . . . 130
A.3.3 Argument of Roots for a VAR(2) Distribution Generating Polynomial . . . 134
A.3.4 Fit of Spectral Density and ESD . . . 136

B Program Code **139**
B.1 Mathematica Code . . . 139
B.1.1 Code for Generating DGP . . . 140

Bibliography **143**

Index **152**

List of Tables

4.1 NASDAQ-100 key statistics for different aggregation levels. 77
4.2 NASDAQ key statistics for different aggregation levels. 80
4.3 S&P 500 key statistics for different aggregation levels. 84
4.4 DJIA key statistics for different aggregation levels. 87

List of Figures

1.1 Marchenko–Pastur law vs. second week of S&P 500 »long« with $T = 946$ and $N = 224$. Largest eigenvalue at 36.084 not shown. . . . 4

2.1 Density of Marchenko–Pastur law for various shape parameters α. Atom for parameter $0 \leq \alpha < 1$ depicted as vertical line of appropriate length at 0. 14
25

3.1 Comparison of VMA(2) eigenvalue densities for various parameter combinations with Marchenko–Pastur law, all with $\alpha = 4$. 43

. 44
3.3 Simulation of $n = 2^{20}$ realizations of size 32×32 Wishart matrices with rectangularity ratio $\alpha = 4$. 44
3.4 Comparison of VAR(2) eigenvalue densities for various parameter combinations with Marchenko–Pastur law, all with $\alpha = 4$. 45
3.5 Comparison of VARMA$(1, 1)$ eigenvalue densities for various parameter combinations with Marchenko–Pastur law, all with $\alpha = 4$. 47
3.6 Comparison of VARFIMA$(0, d, 0)$ eigenvalue densities for various fractional integration orders d, all with $\alpha = 4$. Estimated with algorithm 4.1. 58
3.7 Comparison of VARFIMA$(1, d, 1)$ eigenvalue densities with fractional integration orders $d \in \{0, 0.4\}$, both with $\alpha = 4$. 60
3.8 »Density« functions implied by all 9 roots of VAR(2) with $\theta_1 = \theta_2 = 1/4$ and $\alpha = 4$. Functions corresponding to the true spectral density in bold, all other dashed. 63

4.1 110 NASDAQ-100 stocks, portion of days non-`NA` and with at least two non-0 returns in trading days between 2007-06-27 and 2015-01-16. 74
4.2 82 NASDAQ-100 stocks, fraction of non-zero returns per trading day between 2007-06-27 and 2015-01-16. 75
4.3 82 NASDAQ-100 stocks, fraction of non-zero returns per stock intraday between 2007-06-27 and 2015-01-16. 76
4.4 82 NASDAQ-100 stocks, change of liquidity per stock for aggregated returns, sorted. In black the average for each series. 76
4.5 1618 NASDAQ stocks, portion of days non-`NA` and with at least two non-0 returns in trading days between 2007-06-27 and 2014-11-28. . . 78
4.6 826 NASDAQ stocks, fraction of non-zero returns per trading day between 2007-06-27 and 2014-11-28. 78

4.7 826 NASDAQ stocks, fraction of non-zero returns per stock intraday between 2007-06-27 and 2014-11-28. 79
4.8 826 NASDAQ stocks, change of liquidity per stock for aggregated returns, sorted. In black the average for each series. 80
4.9 502 S&P 500 stocks, portion of days non-NA and with at least two non-0 returns in trading days between 2002-12-30 and 2016-09-16. . . 81
4.10 393 S&P 500 stocks, fraction of non-zero returns per trading day between 2002-12-30 and 2016-09-16. 82
4.11 393 S&P 500 stocks, fraction of non-zero returns per stock intraday between 2002-12-30 and 2016-09-16. 83
4.12 393 S&P500 stocks, change of liquidity per stock for aggregated returns, sorted. In black the average for each series. 83
4.13 30 DJIA stocks, portion of days non-NA and with at least two non-0 returns in trading days between 2002-12-30 and 2016-09-16. 85
4.14 28 DJIA stocks, fraction of non-zero returns per trading day between 2002-12-30 and 2016-09-16. 85
4.15 28 DJIA stocks, fraction of non-zero returns per stock intraday between 2002-12-30 and 2016-09-16. 86
4.16 28 DJIA stocks, change of liquidity per stock for aggregated returns, sorted. In black the average for each series. 87
4.17 Histograms of averaged eigenvalues of the NASDAQ-100. 92
4.18 Histograms of averaged eigenvalues of the NASDAQ. 93
4.19 Histograms of averaged eigenvalues of the S&P 500. 94
4.20 Histograms of averaged eigenvalues of the S&P 500. 95
4.21 Histograms of averaged eigenvalues of the DJIA. 96

5.1 $\|\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\theta})\|_2$ of VMA(1) (red) and VAR(1) (blue) random matrix models, calculated for parameter values in $(0, 2]$. Depicted is the first day of observation 2007-06-27 of the NASDAQ-100 data set. Algorithm 6.1 was used. 103
5.2 $\|\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\theta})\|_2$ of VARMA$(1, 1)$ random matrix model, calculated for parameter values $(\theta_1, \phi_1) \in [-2, 2] \times (0, 3]$. Depicted is the first day of observation 2007-06-27 of the NASDAQ-100 data set. Algorithm 6.1 was used. 106
5.3 Parameter estimates for VARMA$(1, 1)$ random matrix model, S&P 500 »long« data set. $\hat{\theta}_1(t) \in (0, 3]$ in blue, $\hat{\phi}_1(t) \in [-1, 1]$ in red, for all $t \in \mathbb{T}$. Algorithm 6.1 was used. 109
5.4 empirical spectral distribution (ESD) of NASDAQ-100 data set (daily) for $t =$ 2013-01-16 and VARMA$(1, 1)$ spectral density function. Algorithm 6.1 was used. 111
5.5 Example of model fit for VARMA$(1, 1)$ random matrix model, ESD of S&P 500 »long« data set (weekly, monthly and quarterly) for the first complete observation and VARMA$(1, 1)$ spectral density function with estimated parameters $\hat{\theta}_1$ and $\hat{\phi}_1$. Algorithm 6.1 was used. 112
5.6 Example of model fit for VAR$(1, 1)$ random matrix model, ESD of S&P 500 »big« data set (monthly) for the last complete observation and VAR(1) spectral density function with estimated parameter $\hat{\theta}_1$. . 113

A.1 arg of complex-valued root of polynomial $M^9 - (x + \mathrm{i}y)$. 134
A.2 arg of complex-valued root of VAR(2) polynomial for parameters $\theta_1 = 0.5$, $\theta_2 = 0.25$ and $\alpha = 4$. 135
A.3 Example of model fit for VARMA$(1, 1)$ random matrix model, ESD of S&P 500 »big« data set (quarterly) for the median of all eigenvalues and VARMA$(1, 1)$ spectral density function with estimated parameters $\hat{\theta}_1$ and $\hat{\phi}_1$. 136
A.4 Example of model fit for VARMA$(1, 1)$ random matrix model, ESD of NASDAQ data set (monthly) for the mean of all eigenvalues and VARMA$(1, 1)$ spectral density function with estimated parameters $\hat{\theta}_1$ and $\hat{\phi}_1$. 137
A.5 Example of model fit for VARMA$(1, 1)$ random matrix model, ESD of DJIA data set (daily) for the last complete observation and VARMA$(1, 1)$ spectral density function with estimated parameters $\hat{\theta}_1$ and $\hat{\phi}_1$. 138

Symbols and Acronyms

Symbols

θ	AR polynomial of degree p	30
$\boldsymbol{\Theta}$	matrix-valued AR polynomial of degree p; resp. matrix-valued coefficient of this polynomial	30
ψ	MA polynomial for ARMA process, possibly after Wold decomposition	38
$\mathcal{A}_\sigma$	σ-algebra	7
$\mathbf{A}$	general (random) matrix; or $(T \times T)$ auto-covariance matrix	32
$\gamma_{(X_t)}(\tau)$	autocovariance function for stochastic process (X_t) and lags $\tau \in \mathbb{Z}$	29
B	backshift operator	29
B^d	fractional difference operator, for $d > -1$ non-integer	53
$\mathcal{B}_\sigma(\mathbb{R})$	Borel σ-algebra of $\mathbb{R}$	7
`C` (`C++`)	`C` resp. `C++` programming language	68
$\mathcal{G}$	$\mathcal{G}$-transform	21
$\mathcal{G}^{\langle -1 \rangle}$	inverse $\mathcal{G}$-transform w.r.t. composition of power series	22
X^c	centered random variable X or centered (sub)algebra of random variables	16
$\mathbb{C}$	complex numbers	8
$\hat{\mathbf{c}}$	covariance matrix estimate	33
$\mathbf{C}$	general (random) matrix; or $(N \times N)$ cross-covariance matrix	32
$\otimes$	direct product; or the algebraic tensor product resp. the tensor product of Hilbert spaces; or product σ-algebra; or product probability measure	9
$\hat{x}$	estimator for x	13
$\mathbb{E}$	expected value	7
$\mathbb{E}_{\mathbb{P}}$	expected value w.r.t. probability measure $\mathbb{P}$	7

$\mathbb{F}$	general field . . . 8
$\mathbb{C}[\![X]\!]$	set of formal power series in indeterminates X and coefficients in $\mathbb{C}$ . . . 20
$\mathbb{C}[\![X]\!]_\emptyset$	set of formal power series in indeterminates X and coefficients in $\mathbb{C}$ with non-zero constant term . . . 23
$g(\lambda)$	auxiliary function g . . . 49
$\mathbb{T}$	index set for time . . . 30
$\delta_{k,\ell}$	Kronecker delta, 1 for $k = \ell$, 0 otherwise . . . 37
$\varphi \quad (\psi, \omega)$	linear functional . . . 8
$L^p_{\mathbb{R}}(\Omega, \mathbb{P})$	space of real functions from Ω with finite p-norm, where the norm is given by Lebesgue integration w.r.t. the measure $\mathbb{P}$ . . . 8
$L^{-\infty}_{\mathbb{R}}(\Omega, \mathbb{P})$	$L^p_{\mathbb{R}}$-space of real functions from Ω with finite moments of all orders . . 8
$\mathcal{M}$	$\mathcal{M}$-transform . . . 20
$\mathcal{M}^{\langle -1 \rangle}$	inverse $\mathcal{M}$-transform w.r.t. composition of power series . . . 23
ϕ	MA polynomial of degree q . . . 30
$\mathbf{\Phi}$	matrix-valued MA polynomial of degree q; resp. matrix-valued coefficient of this polynomial . . . 30
$M_{n\times m}(\mathbb{F})$	space of $n \times m$ matrices with entries in $\mathbb{F}$ . . . 9
λ^n	n-dimensional Lebesgue measure . . . 21
μ_{MP}	Marchenko–Pastur measure . . . 13
`NA`	not available . . . 70
Ω	sample space, the set of all possible outcomes ω . . . 7
ω	outcome, element from the sample space Ω; or angular argument for Fourier transformation . . . 8
$\mathbb{P}$	probability measure . . . 7
$\mathcal{R}$	$\mathcal{R}$-transform . . . 22
$\mathcal{R}^{\langle -1 \rangle}$	inverse $\mathcal{R}$-transform w.r.t. composition of power series . . . 23
$\mathcal{R}^{\mathsf{orig.}}$	original $\mathcal{R}$-transform . . . 23
rank	rank . . . 15
$\mathbb{R}$	real numbers . . . 7
Re	real part of complex number . . . 57
$\mathbf{P}$	projection matrix . . . 33
$\mathcal{S}$	$\mathcal{S}$-transform . . . 23
$\Sigma_{\mathbb{F}} \quad (\Sigma^{\emptyset}_{\mathbb{F}})$	space of (non-zero) normed linear functionals over a field $\mathbb{F}$ . . . 20

$\tilde{\gamma}_{(X_t)}(\omega)$	spectral density for stochastic process (X_t) and circular angle $\omega \in (-\pi, \pi]$	38
N	number of stocks	31
T	number of observations in time	31
τ	parameter, usually used for lags	29
$\boldsymbol{\theta}$	parameter vector	98
Tr	trace	9
tr_n	normalized trace	9
$\mathcal{WN}(\sigma^2)$	White noise process with variance σ^2	29
$\mathcal{W}_p(n, \mathbf{\Sigma})$	Wishart distribution with $(n \times n)$ covariance matrix $\mathbf{\Sigma}$	13
$\mathcal{X}$	sample space of data	98
$\mathbf{x}$ $(\mathbf{y})$	generic (random) matrix; or $(T \times N)$ data matrix of columnwise per-stock centered returns	33

Acronyms and Abbreviations

a.s. almost sure . . . 14
AAPL Apple Inc. . . . 73
AR autoregressive . . . 29
ARFIMA fractional integrated autoregressive–moving-average . . . 53
ARIMA integrated autoregressive–moving-average . . . 52
ARMA autoregressive–moving-average . . . 30

CAPM capital asset pricing model . . . 90
CDF cumulative distribution function . . . 12
cf. confer, Latin for »compare« . . . 30
CSCO Cisco Systems . . . 73

d.o.f. degree of freedom . . . 88
DGP distribution generating polynomial . . . 40
DJIA Dow Jones Industrial Average Component Stocks . . . 72

ESCDF empirical spectral cumulative distribution function . . . 12
ESD empirical spectral distribution . . . 11
ETF exchange traded fund . . . 72

GARCH generalized auto-regressive conditional heteroskedasticity . . . 53
GOE Gaussian orthogonal ensemble . . . 1

i.e. id est, Latin for »that is« . . . 4
i.i.d. independent and identically distributed . . . 114

LHS left-hand side . . . 22
LSD limit spectral distribution . . . 27

MA moving-average . . . 29
ML maximum likelihood . . . 13
MLE maximum likelihood estimator . . . 13
MSFT Microsoft Corp. . . . 73

NA not available . . . 72
NCPS non-commutative probability space . . . 8
NCPT non-commutative probability theory . . . 1
NCRV non-commutative random variable . . . 8
NYSE:NBR Nabors Industries Ltd . . . 73
NYSE:SLG SL Green Realty Corp. . . . 73

REIT real estate investment trust . . . 73
resp. respectively . . . 11
RHS right-hand side . . . 22
RMT Random Matrix Theory . . . 114
RV random variable . . . 7

S&P 500 S&P 500 Index Component Stocks 72

U.S. United States 72

V Visa Inc. 73
VAR vector-autoregressive 31
VARMA vector-autoregressive–moving-average 31
VMA vector-moving-average 31

w.r.t. with respect to 7

1 Introduction

We would like to give some perspective on modeling strategies. The world we experience is complicated, the state space of possible influencing quantities is extremly high-dimensional. As processing power of the human minds is limited we can only hope to reduce complexity to an extent so that key dynamics are captured. In general, we advocate a model view where the central part of our model describes a low-dimensional subspace and the complementary dimensions are aggregated to a residual dimension. This residual dimension captures all the discrepancy between model and real world.

The scientific act of modeling can be seen as searching for a suitable parametrization so that the low-dimensional subspace captures as much structure of interest to the researcher as possible.

We are interested in a class of preferably parsimonious models that capture and describe the main dynamics of financial stock markets. We restrict ourselves to stock returns and the data sets created in section 4.3. In the first part of this work, we lay the foundation of non-commutative probability theory and utilize it to introduce basic random matrix models. There are many works that utilize Gaussian orthogonal ensemble (GOE) matrices and the implied Marchenko–Pastur spectral distribution of there covariance matrix estimators and successfully apply those results to financial data.

A structural caveat of those models is the required and thus assumed independence[1] of all matrix entries, modulo some symmetry constraints. Thus one of the main points of interest, the inter-dependence of stock returns in time and between different stocks, is assumed non-existent. It seems therefore natural to relax the very strict assumptions of independent stock returns and allow for some dependence structure. We will do so and allow for vector-valued ARMA(p, q) type dynamics.

With the increase of automated data gathering and the reduced cost of processing power it is now feasible to gather data in quality and quantity unprecedented. Both, the high volume and the broad availability of data permits one to find answers to questions intractable just a decade ago.

[1] In the case of all entries Gaussian uncorrelatedness suffices, because for Gaussians independence and uncorrelatedness are equivalent.

1.1 High-dimensional Data

A classical example of a branch with enormous amounts of data is the financial industry. Financial markets connect different local markets to a truely global economy. The largest companies are traded by computer programs up to thousands of times a minute. Financial transaction data is the sediment of the time gone. By analyzing it we may infer on the structure of financial markets. This structural awareness can be employed to build new models of financial markets to enhance stability or predict certain future key measures of the economy.

In classical statistics, the branch of asymptotic theory gives conditions necessary for convergence. But classical asymptotic theory only works well in situations were the number of observations growth and dominates an eventual growth in new explanatory variables. In addition, many methods are univariate and not easily applicable for multivariate situations or are only reasonable employable for a number of explanatory variables up to about 20.

Many financial markets contain hundreds of stocks, for instance the S&P 500 financial market. Another problem is relevance of historic data, as it might be questionable on what information financial data from the 1950^{th} contains, an era before widespread use of computers and before the invention of the internet. Therefore, there might only exists a medium amount of time with data relevant for certain questions. Both observations lead to the conclusion that the time horizon of relevant data is limited and that more and more very new data becomes available. This calls for high-dimensional methods suitable for situations with possibly hundreds of explanatory variables like stock returns. In addition, high-dimensional methods should be able to cope with situations where the growth of observations does not dominate the growth of explanatory variables.

1.2 Covariance Estimation and Model Parameter Growth

Motivated from elementary statistics one can model the interdependency of two quantities by a relationship between those two quantities that is estimable. One of the most easy measures of dependency is the second moment. Normalized to second centralized moment, this is the ubiquitous covariance. It is an easy to estimate dependency measure between two random variables.

If one wanted to continue and measure global dependencies by the collection of pairwise dependencies, one notices that the number of parameters to estimate growth like $\mathcal{O}(N^2)$, for N the number of random variables under investigation. In concrete numbers, due to the symmetry of the univariate covariance between two random variables the number of covariances growth with $2^{-1}N(N-1)$. For $N = 10$ this amounts only to 45, but for $N = 100$ there are already 4 950 parameters to estimate.

If the S&P 500 had exactly $N = 500$ stocks and one wanted to know the covariance matrix for the whole S&P 500, one needed to estimate 124 750 covariances.

This little example illustrates that it is neither feasible nor desirable to model high-dimensional dependency by pairwise dependency measures. If we model to reduce complexity and understand the world with our limited human mental capacity, we have to reduce complexity. This is what models are for. Even the 45 for $N = 10$ imply a number of parameters way to big to use as model parameters of interest.

The previous paragraphs imply that one should not model high-dimensional dependency by a covariance matrix. But in many applications a covariance estimator is needed or is of interest in its own right. This includes classical modern portfolio theory in finance, introduced 1952 by Markowitz.[2] The estimated covariance matrix, or its inverse, is the key building block to obtain portfolio weights. If one were to estimate a classical covariance matrix for $N = 100$ stocks, this would imply estimation of 4 950 covariances. By pure chance, some of those estimates will over- or underestimate the assumed true covariance of the model. Therefore one could say with classical Markowitz portfolio selection and high-dimensional data one invests ones money into the most pronounced estimation errors.

1.3 New Framework: Non-Commutative Probability Theory

Classical stochastics is concerned with the behavior of random variables and their interactions and combined behavior. At the core are the basic operations of addition and multiplication of random variables. For collections of random variables, one can gather them in a vector. For vectors whose entries are random variables one only has addition. But there is no general inner multiplication of two random vectors, yielding a random vector again.

For, say, real valued special vectors of length N^2 one can identify the vector space $\mathbb{R}^{N^2}$ with $\mathbb{R}^{N\times N}$, the space of square matrices. Often, a matrix structure with random entries is preferable for structural reasons if one wants to model high-dimensional objects. On square matrices there exists a natural multiplication, given by matrix multiplication. It only has the slight caveat that it is not commutative, that is for two matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{N\times N}$ in general we have that $\mathbf{AB} \neq \mathbf{BA}$. But we will show that this non-commutativity is the source for a rich mathematical structure that enables one to work with matrices of classical random variables as new objects in its own right.

In situations where the model building blocks can be written as matrices, each with classical random variables as its entries, non-commutative probability theory is the natural model framework. It is not a generalization of classical probability

[2] See [64].

theory, but more of a parallel framework that builds on and uses classical probability theory, but is distinct from it. This has the caveat that most theory, including its algebraic foundation, has to be learned anew by trained classical probabilists. But the rewarding theory leads to exiting new ways to model the complexity of the world and thus reduce it and help us understand.

In this thesis, we want to give a thorough introduction to non-commutative probability theory and highlight its ability to derive theoretical results previously unavailable. In addition, we showcase its applicability by exemplary parameter estimation of certain parametrizations of covariance matrices for financial stock return data.

1.4 Simple Comparison with Market Data

We show in Figure 1.1 the histogram of the eigenvalues of the sample covariance matrix $\hat{\mathbf{C}}\left(\mathbf{X}_{\text{S\&P 500}}^{(2^{\text{nd}}\text{ week})}\right)$ for the S&P 500 »long« model for the second week between 2003-01-06 and 2003-01-10.[3] To facilitate comparison we re-normalize each asset time series so that it has vanishing mean and unit variance. Because of the latter the resulting empirical covariance matrix is also the empirical correlation matrix with only 1 on its main diagonal. This fixes the matrix trace, id est (i.e.) the sum of all eigenvalues, to equal the number of assets N. This is a reasonable choice as now data from each data set has a mutual scale. One can even compare data from different data sets, for instance the DJIA and the NASDAQ.

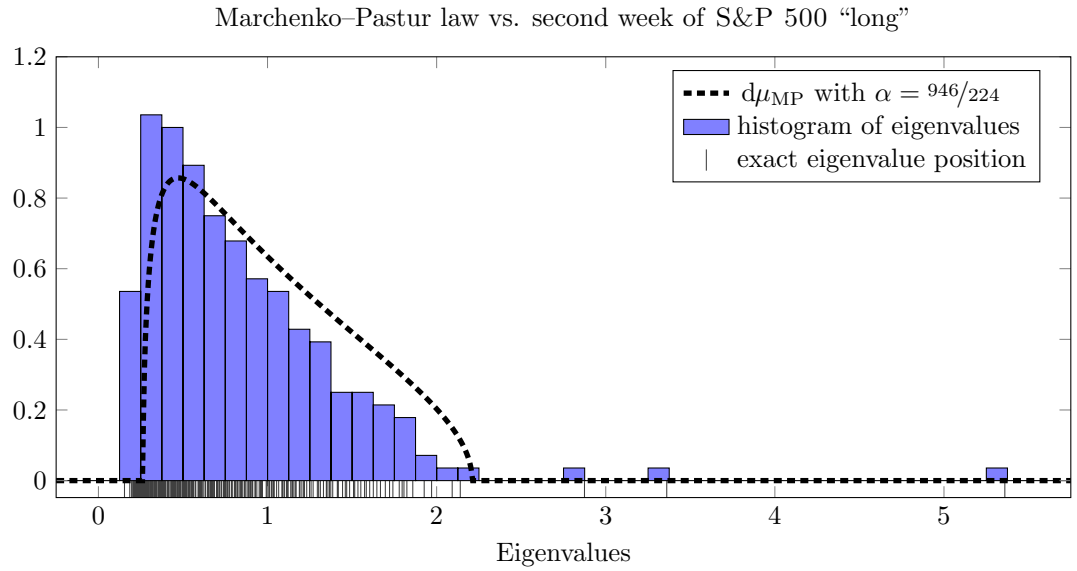

Figure 1.1: Marchenko–Pastur law vs. second week of S&P 500 »long« with $T = 946$ and $N = 224$. Largest eigenvalue at 36.084 not shown.

[3] The first week is not a good representative of the data set as it has fewer trading days due to *New Year's Day*.

One of the first basic objects from RMT is the so-called *Marechenko–Pastur distribution*. It has a shape parameter $\alpha = {}^{946}/_{224}$ and it approximates the data decently. The 4 biggest and the 18 smallest eigenvalues fall outside the Marchenko–Pastur domain. It is well known[4] that empirical spectral distributions of financial return series contain a largest eigenvalue much bigger than all the other eigenvalues. Empirically the associated eigenvector can be seen as a proxy for the so-called *market portfolio*. Therefore the biggest eigenvalue is called the *market eigenvalue*.

We want to refine this analysis and find estimators, so that their spectrum approximates the empirical observable histogram of the sample covariance matrix even better.

1.5 Overview

In the first part of this thesis we introduce the theoretical foundation necessary for application to high-dimensional financial correlation matrices. In chapter 2 we introduce the concept of non-commutative probability theory and motivate that so-called *freeness* is the non-commutative analog to stochastic independence in classical probability theory. We develop the theory to add and multiply free non-commutative random variables and convenient tools to facilitate analysis. The connection between non-commutative random variables and random matrices is shown in chapter 3. After the introduction of certain random matrix ensembles the theoretical spectral density functions for correlation matrix estimates of random matrices with dependence between its entries are derived in sections 3.1 and 3.2.

The second part is devoted to the application of the theoretical framework developed in the first part to high-dimensional financial data. An overview of the raw data available is given in chapter 4. The methodology used to construct the final data sets is explained. The so-called empirical spectral distribution of correlation matrix estimators for the financial stock return data are shown. In chapter 5 we introduce the methodology for estimation of the model parameters. Finally, we exemplary present some results of the estimated model parameters for the data sets.

A recapitulation, conclusion and perspectives conclude this thesis.

[4] See e.g. [17, p. 148].

2 Introduction to Non-Commutative Probability Theory and Free Independence

Depending on the researchers' interest, a mathematical structure can be viewed from different perspectives to accentuate different aspects. In the branch of *measure theory* a measurable space $(\Omega, \mathcal{A}_\sigma)$, consisting of a set Ω and a σ-algebra $\mathcal{A}_\sigma$ of Ω of measurable sets, is endowed with a finite normed measure $\mathbb{P}$ to make it a measure space. Often one adds additional structure to the set Ω and studies its implications under the linear functional $\mathbb{P}$. The original structure of the measure space is vital to this line of research.

From the perspective of classical probability theory the triple $(\Omega, \mathcal{A}_\sigma, \mathbb{P})$ is called a probability space and the main area of interest are measurable functions from an abstract measure space to some other, well-behaved measurable space, e.g. to $(\mathbb{R}, \mathcal{B}_\sigma(\mathbb{R}))$. These measurable functions are called »random variables« and induce a push-forward measure from the original, abstract probability space to $(\mathbb{R}, \mathcal{B}_\sigma(\mathbb{R}))$. Random variables (RVs) and their distributions are the primary objects studied in *probability theory*. Commonly the original probability space $(\Omega, \mathcal{A}_\sigma, \mathbb{P})$ is not fixed but generic. This principle gives great flexibility when working with families of random variables, because one can simply extend the original probability space to a suitable product space. The original probability space is embedded in the product space and can be retrieved by projection on one of the factors.

2.1 Introduction to Non-Commutative Probability Theory

Classical probability theory is the study of commuting random variables and their distributions, i.e. the study of the commutative algebra of random variables with the expectation map $\mathbb{E}$. It suffices to use the linear expectation functional $\mathbb{E}_\mathbb{P}$ with respect to (w.r.t.) a probability measure $\mathbb{P}$, because one can retrieve the probability measure:

$$\mathbb{P}[X \in A] = \mathbb{E}[\mathbb{1}_A(X)] \qquad \text{for a random variable } X \text{ and for every event}^1 A.$$

If one wants to study non-commutative random variables, the logical next step is to change to an non-commutative algebra of then non-commutative random variables. From this perspective, non-commutative probability theory is the study of non-commutative algebras together with a linear functional φ that serves as expectation map.

2.1.1 Basic Definitions

Definition (Non-Commutative Probability Space)
A non-commutative probability space (NCPS) $(\mathcal{A}, \varphi)$ consists of a unital algebra $\mathcal{A}$ over a field $\mathbb{F}$ and a normed linear functional[2]

$$\varphi\colon \mathcal{A} \to \mathcal{F}, \qquad \text{with} \quad \varphi(\mathbf{1}_{\mathcal{A}}) = \mathbf{1}_{\mathbb{F}}.$$

Definition ($\star$-Probability Space)
If, in addition to the definition of non-commutative probability spaces, the unital algebra $\mathcal{A}$ is an $\star$-algebra and if

$$\varphi(\mathbf{a}^\star \mathbf{a}) \geq 0, \qquad (\forall \mathbf{a} \in \mathcal{A}), \qquad \text{(positivity)}$$

we call the tuple $(\mathcal{A}, \varphi)$ an $\star$-probability space.

The elements $\mathbf{a} \in \mathcal{A}$ are called *non-commutative random variables (NCRVs)* in $(\mathcal{A}, \varphi)$. To shorten notation we will usually skip the prefix »non-commutative« and just refer to *random variables*. This is justifiable, as we will see in the next example:

Example 2.1 (Classical Probability Space as Non-Commutative Probability Space)
Let $(\Omega, \mathcal{A}_\sigma(\Omega), \mathbb{P})$ be a classical probability space and define

$$L_{\mathbb{R}}^{-\infty}(\Omega, \mathbb{P}) := \bigcap_{1 \leq p < \infty} L_{\mathbb{R}}^{p}(\Omega, \mathbb{P})$$

and

$$\varphi(\mathbf{a}) := \int_\Omega \mathbf{a}(\omega)\, \mathrm{d}\mathbb{P}(\omega), \qquad \mathbf{a} \in L_{\mathbb{R}}^{-\infty}(\Omega, \mathbb{P}).$$

Then the non-commutative probability space $\left(L_{\mathbb{R}}^{-\infty}(\Omega, \mathbb{P}), \varphi\right)$ is the classical probability space of random variables that have finite moments of all orders. The map φ is the classical expectation map, $\varphi \equiv \mathbb{E}_{\mathbb{P}}$.

[1] An event is an element of a σ-algebra of subsets of the sample space.
[2] The field $\mathbb{F}$ will be most times either $\mathbb{R}$ or $\mathbb{C}$.

Definition
For a $\star$-probability space we call a random variable $\mathbf{a} \in \mathcal{A}$, with

(i) $\mathbf{a} = \mathbf{a}^\star$; (self-adjoint)
(ii) $\mathbf{a}^\star\mathbf{a} = \mathbf{a}\mathbf{a}^\star$; (normal)
(iii) $\mathbf{a}^\star\mathbf{a} = \mathbf{a}\mathbf{a}^\star = \mathbf{1}_{\mathcal{A}}$. (unitary)

2.1.2 Random Matrices

Example 2.2 (Random Matrices)
Let $M_{n\times n}(\mathbb{R})$ be the set of $(n \times n)$-matrices with entries in $\mathbb{R}$ and let $\mathrm{tr}_n := \frac{1}{n}\mathrm{Tr}(\cdot)$, i.e.

$$\begin{aligned} \mathrm{tr}_n \colon M_{n\times n}(\mathbb{R}) &\to \mathbb{R} \\ \mathbf{a} &\mapsto \frac{1}{n}\sum_{k=1}^{n} a_{k,k} \qquad \text{for} \quad \mathbf{a} = (a_{k,l})_{k,l=1}^{n} \in M_{n\times n}(\mathbb{R}), \end{aligned}$$

be the normalized trace[3] of these matrices. Then the non-commutative probability space

$$\left(M_{n\times n}\left(L_{\mathbb{R}}^{-\infty}(\Omega,\mathbb{P})\right), \mathrm{tr}_n \otimes \mathbb{E}_{\mathbb{P}}\right)$$

consists of $(n \times n)$-matrices with each entry being a classical random variable (with all moments finite), i.e. a *random matrix*, and a linear functional which is the normalized trace of the classical expected value of the matrix[4].

Definition (Average Eigenvalue Distribution)
Let the $\star$-probability space be given by

$$\left(M_{n\times n}\left(L^{-\infty}(\Omega,\mathbb{P})\right), \mathrm{tr}_n \otimes \mathbb{E}_{\mathbb{P}}\right).$$

Let $\mathbf{A}(\omega) \in M_{n\times n}(L^{-\infty}(\Omega,\mathbb{P}))$ be a normal matrix for all $\omega \in \Omega$ and let $\lambda_1(\omega), \ldots, \lambda_n(\omega)$ be the eigenvalues of $\mathbf{A}(\omega)$. Then the averaged eigenvalue distribution *is defined to be*

$$\mu_{\mathbf{A}} := \frac{1}{n}\sum_{j=1}^{n} \int_\Omega \delta(\lambda_j(\omega))\, \mathrm{d}\mathbb{P}(\omega).$$

For all $k, n \in \mathbb{N}_0$, it holds that

$$\varphi\left(\mathbf{A}^k(\mathbf{A}^\star)^l\right) = (\mathrm{tr}_n \otimes \mathbb{E})\left(\mathbf{\Lambda}^k(\mathbf{\Lambda}^\star)^l\right) = \frac{1}{n}\sum_{j=1}^{n} \int_\Omega \lambda_j(\omega)^k \overline{\lambda_j(\omega)}^l \, \mathrm{d}\mathbb{P}(\omega) = \int_{\mathbb{C}} z^k \overline{z}^l \, \mathrm{d}\mu_{\mathbf{A}}(z).$$

In general $\mu_{\mathbf{A}}$ depends on the classical joint probability distribution $\mathbb{P}$ of all entries of $\mathbf{A}$, therefore we cannot generally say whether $\mu_{\mathbf{A}}$ is compactly supported or not. But

[3] If it is clear from context, we will suppress the dimension index n and write tr instead of tr_n.
[4] The expected values are taken element-wise.

if we specify a joint probability distribution $\mathbb{P}$ or family of probability distributions $\mathbb{P}_\theta$ indexed by some parameter vector θ, one can calculate the properties of $\mu_{\mathbf{A}}$. For the cases interesting to us in this thesis the $\star$-distribution exists and we identify the average eigenvalue distribution with it.

If one wants to describe the joint behavior of several or all entries of a $(n \times n)$ random matrix $\mathbf{a}$, that is the joint distribution of several or all classical random variables forming the entries, one has to switch to the product probability space. To elucidate, let $\mathcal{I} \coloneqq \{1, \ldots, n^2\}$ and let $(\Omega_i, \mathcal{A}_\sigma(\Omega_i), \mathbb{P}_i)$ be the classical probability space for the i^{th} entry, with $i \in \mathcal{I}$. The product probability space is then given by

$$\bigotimes_{i\in\mathcal{I}} (\Omega_i, \mathcal{A}_\sigma(\Omega_i), \mathbb{P}_i) \coloneqq \left(\bigtimes_{i\in\mathcal{I}} \Omega_i, \bigotimes_{i\in\mathcal{I}} \mathcal{A}_\sigma(\Omega_i), \bigotimes_{i\in\mathcal{I}} \mathbb{P}_i \right).$$

Note that because we chose such a general notation, the probability spaces for every single entry can be chosen differently. If we only want the joint distribution of a subset of all entries, it suffices to consider the projection onto this measurable space from the complete product space. To facilitate matters, we often require the entries to be independent random variables. If we want to describe dependencies among the entries, we work on the product measurable space but replace the product measure $\mathbb{P} \coloneqq \bigotimes_{i\in\mathcal{I}} \mathbb{P}_i$ by a suitable probability measure $\mathbb{P}'$ on that measurable space that describes the dependencies.

On classical random variables the expectation map $\mathbb{E}_\mathbb{P}$ is a linear functional on the space of random variables. The trace Tr as well as the normalized trace tr_n are both linear functionals on the space of $(n \times n)$ matrices. Random matrices combine the two aforementioned spaces, so a linear functional on the space of random matrices can be constructed from the individual linear functionals. As we defined random matrices to be matrices of classical random variables, we can apply the appropriate linear functionals. That is, we can send random variables to the base field via the expectation map $\mathbb{E}_\mathbb{P}$ to get an $(n \times n)$ matrix with all entries deterministic. Next, the normalized trace tr_n maps the deterministic matrix to a scalar in the base field. So we have

$$\mathrm{tr}_n \otimes \mathbb{E}_\mathbb{P} : M_{n\times n}(\mathbb{F}) \otimes L_\mathbb{F}^{-\infty}(\Omega, \mathbb{P}) \to \mathbb{F}, \qquad \mathbf{a} \mapsto \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}_\mathbb{P}[a_{k,k}].$$

If we took the normalized trace first, this would result in a scalar-valued random variable. If it is beneficial, we will adapt this view.

Remark: *We could also use an expectation map that conditions on a non-trivial*[5] *sub-σ-algebra $\mathcal{A}_1$ to get a conditional expectation. The resulting map can then be thought of as having $\mathcal{A}_1$-measurable classical random variables as values. Though the term »linear functional« does not necessarily apply anymore.*

[5] The »full« expectation $\mathbb{E}_\mathbb{P}$ conditions on the trivial σ-algebra $\mathcal{A}_\varnothing \coloneqq \{\varnothing, \Omega\}$, so only constant random variables are $\mathcal{A}_\varnothing$-measurable and thus the expectation $\mathbb{E}_\mathbb{P}$ takes values in the constant random variables, that is the constant scalars.

The vector space of $(n \times n)$ matrices together with any submultiplicative norm forms a Banach space. This extends naturally to a unital Banach algebra for algebra multiplication given by matrix multiplication. It can be made to a unital $C^\star$-algebra by the involution operation $\star$ of *transposition* for real respectively *conjugate transposition* for complex valued matrices, under which it is obviously closed and we have $\|\mathbf{a}^\star\mathbf{a}\| = \|\mathbf{a}\|^2$. De-randomization by taking the expected value first makes the vector space of random matrices to a vector space of deterministic $(n \times n)$ matrices. This extends naturally to a $C^\star$-algebra, as just described.

Conclusion 2.3 (Random Matrices as Non-Commutative Random Variables)
The $C^\star$-algebra $\mathcal{A}_n \coloneqq M_{n\times n}\big(L_{\mathbb{F}}^{-\infty}(\Omega, \mathbb{P})\big)$ together with the linear functional

$$\varphi_n \coloneqq \mathrm{tr}_n \otimes \mathbb{E}_{\mathbb{P}} : M_{n\times n}(\mathbb{F}) \times L_{\mathbb{F}}^{-\infty}(\Omega, \mathbb{P}) \to \mathbb{F}, \quad \mathbf{a}(\omega) \mapsto \int \mathrm{tr}_n(\mathbf{a}(\omega))\, \mathrm{d}\mathbb{P}(\omega),$$

form the *$C^\star$-probability space* $(\mathcal{A}_n, \varphi_n)$ with elements $\mathbf{a} \in \mathcal{A}_n$ as *random matrices*.

Remark: *Since the normalized trace* tr_n *is the finite sum of* n *summands, one can*[6] *interchange the expectation map* $\mathbb{E}_{\mathbb{P}}$ *with* tr_n.

The connection of random matrices and free probability theory was established early[7] on by Voiculescu et al., because the semicircle distribution from free probability was known much earlier since 1958 as the limit distribution of eigenvalues of large Gaussian random matrices [114], which was called »Wigner's Semicircle Distribution«.

Definition (Empirical Spectral Distribution (ESD))
Let $\mathbf{M} \in \mathbb{R}^{n\times n}$ *be a deterministic (or random) matrix and let* $\lambda_1, \lambda_2, \ldots, \lambda_n$ *be its eigenvalues*[8]. *The* empirical spectral distribution *or* eigenvalue distribution *is the deterministic (respectively (resp.) random) probability measure*[9] *on* $\mathbb{C}$ *given by*

$$\mu_{\mathbf{M}} \coloneqq \frac{1}{n}\sum_{j=1}^{n} \delta(\lambda_j).$$

Remark: *If matrix* $\mathbf{M} \in \mathbb{R}^{n\times n}$ *in the definition above is* self-adjoint*, all its eigenvalues exist and are real and thus the probability measure* $\tilde{\mu}_{\mathbf{M}}$ *is constricted to the reals. Note that* symmetric *matrices are in particular* self-adjoint.

For a random matrix $\mathbf{M}$ obviously its eigenvalues are random variables, too. So the eigenvalue distribution, which gives n^{-1} probability mass to each eigenvalue, is also random. But one can take the expected value of the eigenvalues, which results in the expected eigenvalues distribution.

[6] The associated series with only finite many non-zero summands is absolutely convergent.
[7] See [111, p. 43].
[8] Respectively, if the matrix is not diagonalizable, the diagonal elements of its Jordan normal form.
[9] The dependence on the dimension n is suppressed in the notation, if no confusion can arise.

For a self-adjoint matrix $\mathbf{M}$ with only real eigenvalues we can define its empirical cumulative distribution function.

Definition (empirical spectral cumulative distribution function (ESCDF))
Let $\mathbf{M} \in \mathbb{R}^{n\times n}$ *be a self-adjoint*[10] *deterministic (or random) matrix and let* $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ *be its real eigenvalues. The* empirical spectral cumulative distribution function *is defined to be*

$$F_{\mathbf{M}}\colon \mathbb{R} \to [0,1], \qquad \lambda \mapsto F_{\mathbf{M}}(\lambda) \coloneqq \mu_{\mathbf{M}}((-\infty, \lambda]) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\lambda_i \leq \lambda}.$$

We will work with either the empirical spectral measure or the empirical spectral CDF, whichever is more convenient.

2.2 Wishart Ensemble and Marchenko–Pastur Distribution

2.2.1 Wishart Ensemble

Definition (Wishart Ensemble and Wishart Distribution, see [72, p. 82, Definition 3.1.3])
Let $(\mathbf{x}_i)_{i=1}^n$ *be a family of* n *independent classical random* p*-dimensional real vectors, each distributed according to* $\mathbf{x}_i \sim \mathcal{N}_p(\mathbf{0}, \mathbf{\Sigma})$*, with* $\mathbf{\Sigma} \in \mathbb{R}^{p\times p}$ *symmetric, positive semidefinite. Let* $\mathbf{X} \coloneqq \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_n \end{pmatrix}^\intercal$ *be the* $(n \times p)$ *matrix with each row given by an independent, centered Gaussian vector* $\mathbf{x}_i$*. The* Wishart Ensemble *is given by the ensemble of scatter matrices* $\mathbf{W} \coloneqq \mathbf{X}^\intercal\mathbf{X}$.
The probability distribution of $\mathbf{W}$ *on the symmetric, positive semidefinite real random matrices is denoted as* $\mathcal{W}_p(n, \mathbf{\Sigma})$*, i.e. we have* $\mathbf{W} \sim \mathcal{W}_p(n, \mathbf{\Sigma})$*. It is called the* Wishart distribution.

Notation: If the covariance matrix $\mathbf{\Sigma}$ of the Gaussian vectors is the identity matrix $\mathbf{I}_p$, we simple use the shorthand notation $\mathcal{W}_p(n) \coloneqq \mathcal{W}_p(n, \mathbf{I}_p)$. We call n the *degrees of freedom*, noting that this quantity only bears the usual meaning if $n \geq p$.

Remark: *The Wishart distribution, introduced 1928 by Wishart in [117], can be thought of as a multivariate generalization of the chi-square distribution. But we follow Seber, so*

> *"[. . .] we do not call the Wishart distribution the multivariate chi-square distribution, as the marginal distribution [. . .] is not chi square. We*

[10] For $\mathbf{M}$ a non-self-adjoint matrix with complex eigenvalues[8], one can define a 2-dimensional cumulative distribution function (CDF) by the isomorphic relation $\mathbb{C} \cong \mathbb{R}^2$.

normally reserve the term multivariate for the case when all univariate marginals belong to the same family." ([88, p. 22])

The probability density[11] function of a Wishart distributed matrix $\mathbf{W} \sim \mathcal{W}_p(n, \mathbf{\Sigma})$ is given by

$$p(\mathbf{W}) \coloneqq \frac{2^{-\frac{np}{2}}}{\Gamma_p\left(\frac{n}{2}\right)} |\mathbf{\Sigma}|^{-\frac{n}{2}} |\mathbf{W}|^{\frac{n-p-1}{2}} \exp\left(-\frac{1}{2}\operatorname{Tr}\left(\mathbf{\Sigma}^{-1}\mathbf{W}\right)\right), \qquad n \geq p,$$

for $\mathbf{\Sigma} \in \mathbb{R}^{p\times p}$ a positive definite covariance matrix. $\Gamma_p(x)$ denotes the multivariate gamma function given in terms of the univariate gamma function in [53, p. 483, eq. (57)] as

$$\Gamma_p(x) \coloneqq \pi^{\frac{p(p-1)}{4}} \prod_{k=1}^{n} \Gamma\left(x + \frac{1-k}{2}\right).$$

By construction, the Wishart distribution is the (rescaled) distribution of the maximum likelihood (ML) covariance matrix estimator for a sample from a multivariate (centered[12]) normal distribution. The maximum likelihood estimator (MLE) for such a covariance matrix is

$$\hat{\mathbf{\Sigma}} \coloneqq \frac{1}{n}\mathbf{X}^{\intercal}\mathbf{X}. \tag{2.1}$$

If we rescaled $\mathbf{X} \mapsto n^{-\frac{1}{2}}\mathbf{X}$ or, equivalently, rescaled the variance-covariance matrix $\mathbf{\Sigma} \mapsto n^{-\frac{1}{2}}\mathbf{\Sigma}$, the resulting Wishart ensemble has elements $\mathbf{W} \sim \mathcal{W}_p\left(n, n^{-\frac{1}{2}}\mathbf{\Sigma}\right)$, which are distributed like (2.1) for unscaled $\mathbf{X}$ resp. $\mathbf{\Sigma}$.

2.2.2 Marchenko–Pastur Distribution

Theorem 2.4 (Marchenko–Pastur Distribution for Random Matrices)
Let $\mathbf{W} \sim \mathcal{W}_p\left(n, n^{-\frac{1}{2}}\mathbf{I}_p\right)$ be a Wishart matrix and let the dimensions $n, p \to \infty$ in such a way that $\frac{n}{p} \to \alpha \in (0, \infty)$. Then the empirical eigenvalue distribution $\tilde{\mu}_{\mathbf{W}}$ converges in the weak* topology, in probability, to the deterministic Marchenko–Pastur distribution given by

$$\mathrm{d}\mu_{\mathrm{MP}} = \begin{cases} (1-\alpha)\delta_0 + \frac{\alpha}{2\pi\lambda}\sqrt{(\alpha_+ - \lambda)(\lambda - \alpha_-)}\mathbb{1}_{[\alpha_-,\alpha_+]}\,\mathrm{d}\lambda, & \text{if } 0 \leq \alpha < 1 \\ \frac{\alpha}{2\pi\lambda}\sqrt{(\alpha_+ - \lambda)(\lambda - \alpha_-)}\mathbb{1}_{[\alpha_-,\alpha_+]}\,\mathrm{d}\lambda, & \text{if } \alpha \geq 1 \end{cases} \tag{2.2}$$

with $\alpha_\pm \coloneqq \left(1 \pm \sqrt{\alpha^{-1}}\right)^2$.

Proof See e.g. [9, Chapter 3] ■

[11] See the original paper [117] by Wishart and [72, p. 85, eq. (1)].
[12] If the rows of $\mathbf{X}$ are distributed like $\mathbf{x}_i \sim \mathcal{N}_p(\boldsymbol{\mu}, \mathbf{\Sigma})$ with $\boldsymbol{\mu} \neq \mathbf{0}$, we speak of a *non-central Wishart distribution*.

Remark: *The name* Marchenko–Pastur *law was given because in 1967 Marchenko and Pastur derived the density function for the first time in [63]. The name* free Poisson *stems from the fact that the limit distribution of*

$$\lim_{n\to\infty}\left(\left(1-\frac{\alpha}{n}\delta_0+\frac{\alpha}{n}\delta_1\right)\right)^{\boxplus n}, \tag{2.3}$$

that is the free additive convolution of free projections $\left(\mathbf{p}_{\frac{\alpha}{n}}\right)_{n\in\mathbb{N}}$*, is the Marchenko–Pastur distribution. Because* (2.3) *resamples the construction of a Poisson-distributed classical random variable, it is called the* free Poisson *distribution, see [78, pp. 203–206].*

As in the Wigner matrix case, the results of Marchenko and Pastur were subsequently strengthened. The strongest result is almost sure convergence proved in [93] under the finite second moment assumption for matrix $\mathbf{X}$. Due to a Wishart matrix being $\mathbf{W} = \mathbf{X}^T\mathbf{X}$, this amounts to finite fourth moments of the resulting Wishart matrix itself.

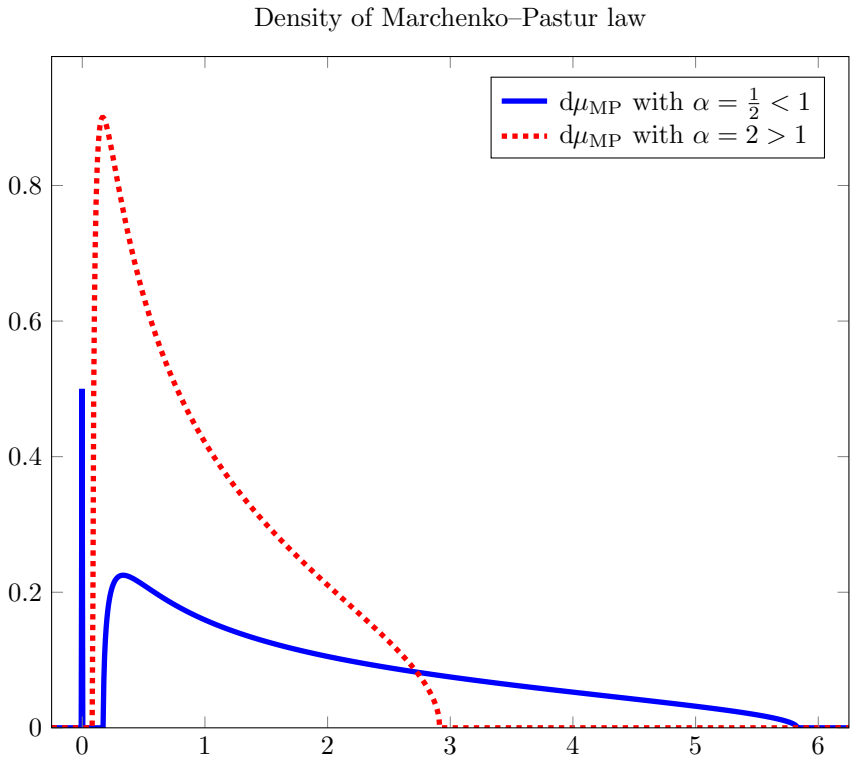

Figure 2.1: Density of Marchenko–Pastur law for various shape parameters α. Atom for parameter $0 \leq \alpha < 1$ depicted as vertical line of appropriate length at 0.

A graph of the Marchenko–Pastur distribution in shown in Figure 2.1. For parameter values of $0 \leq \alpha < 1$ the Marchenko–Pastur distribution has an atom of probability mass $1 - \alpha$ at 0. For $\alpha \to \infty$ its supports converges to its expected value 1. The variance of a Marchenko–Pastur distribution is given by α^{-1}, provided $\alpha \geq 1$.

We already gave a connection of the Wishart ensemble with covariance matrix estimators, namely the ML covariance matrix estimator in (2.1). The covariance matrix of a classical random vector $\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\mathbf{x} \in \mathbb{R}^p$ is given by

$$\boldsymbol{\Sigma} = \mathbb{E}\big[(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{x} - \mathbb{E}[\mathbf{x}])^\intercal\big] = \mathbb{E}[\mathbf{x}\mathbf{x}^\intercal] - \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{x}]^\intercal. \tag{2.4}$$

Let a sample of size n be given by n independent realizations $(\mathbf{x}_i)_{i=1}^n$ of a random vector $\mathbf{x}$ distributed according to $\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. One could estimate the expected value of $\mathbf{x}$ by the arithmetic average over the whole sample, that is one uses

$$\hat{\boldsymbol{\mu}} \coloneqq \bar{\mathbf{x}} \coloneqq \hat{\mathbb{E}}[\mathbf{x}] \coloneqq \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i.$$

Applying this procedure at covariance matrix (2.4) results in the plug-in covariance matrix estimator

$$\hat{\boldsymbol{\Sigma}}_n^{(\mathrm{u})} \coloneqq \hat{\mathbb{E}}[\mathbf{x}\mathbf{x}^\intercal] - \hat{\mathbb{E}}[\mathbf{x}]\hat{\mathbb{E}}[\mathbf{x}]^\intercal \coloneqq \left(\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\intercal\right) - \left(\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\right)\left(\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\right)^\intercal. \tag{2.5}$$

This estimator differs only by the rank-1 matrix $\bar{\mathbf{x}}\bar{\mathbf{x}}^\intercal$ from the estimator

$$\hat{\boldsymbol{\Sigma}}_n \coloneqq \hat{\mathbb{E}}[\mathbf{x}\mathbf{x}^\intercal] = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\intercal. \tag{2.6}$$

We want to estimate bounds on how the empirical spectral CDF changes by subtraction of a rank-1 matrix. The next theorem gives a result slightly more general than we require.

Theorem 2.5 (Inequality for ESCDFs, see [94, p. 179, Lemma 2.4])
Let $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{n\times n}$ be two self-adjoint matrices and let their respective empirical spectral CDFs be given by $F_{\mathbf{X}}$ resp. $F_{\mathbf{Y}}$. Then we have

$$\|F_{\mathbf{X}} - F_{\mathbf{Y}}\|_\infty \leq \frac{1}{n}\mathrm{rank}(\mathbf{X} - \mathbf{Y}).$$

Proof See [94, Proof of Lemma 2.4]. ■

To conclude, the empirical spectral CDFs of the two estimators $\hat{\boldsymbol{\Sigma}}_n^{(\mathrm{u})}$ in (2.5) and $\hat{\boldsymbol{\Sigma}}_n$ in (2.6) differ in supremum norm only by n^{-1}, i.e. by Theorem 2.5 we have

$$\left\|F_{\hat{\boldsymbol{\Sigma}}_n^{(\mathrm{u})}} - F_{\hat{\boldsymbol{\Sigma}}_n}\right\|_\infty \leq \frac{1}{n}.$$

Thus, for $\lim n \longrightarrow \infty$ the empirical spectral CDFs of both estimators converge uniformly. When we are interested in the large n behavior of those estimators, we therefore can restrict our investigation on the structurally simpler second moment estimator $\hat{\boldsymbol{\Sigma}}_n$ in (2.6).

2.3 Free Independence

2.3.1 Independence for Classical Random Variables

As we are able to embed classical commutative probability spaces in the framework of non-commutative probability theory, we like to extend classical independence to non-commutative probability spaces. Therefore we recapitulate the definition of independence for classical random variables, which is also called *tensor independence*.

Definition (Tensor Independence, [78, Definition 5.1])
Let $(\mathcal{A}, \varphi)$ be a non-commutative probability space and let $\mathcal{I}$ be a fixed index set. Unital subalgebras $(\mathcal{A}_i)_{i\in\mathcal{I}}$ are called tensor independent *or* classical independent*, if the subalgebras commute and if φ factorizes in the following way:*

$$\varphi\left(\prod_{j\in\mathcal{J}} \mathbf{a}_j\right) = \prod_{j\in\mathcal{J}} \varphi(\mathbf{a}_j),$$

for all finite subsets $\mathcal{J} \subset \mathcal{I}$ and all $\mathbf{a}_j \in \mathcal{A}_j$.

In the sense of categories, the product of two, possibly non-commutative, probability spaces $(\mathcal{A}_i, \varphi_i)$ with $i \in \{1, 2\}$, is the tensor product $(\mathcal{A}_1 \otimes \mathcal{A}_2, \varphi_1 \otimes \varphi_2)$. Therefore classical independence is sometimes also called *tensor independence*. The tensor product probability space contains two canoncial tensor independent subalgebras, namely $\mathcal{A}_1 \otimes \mathbf{1}_{\mathcal{A}_2}$ and $\mathbf{1}_{\mathcal{A}_1} \otimes \mathcal{A}_2$.

As commutativity of all subalgebras is a critical requirement for tensor independence, it does not make much sense to use it for explicitly non-commutative situations. But one can mimic the structure of tensor independence and try to amend it to non-commutative probability theory.

2.3.2 Definition of Free Independence

Before we define the non-commutative analog to classical independence, we need the following definition:

Definition (Centering)
Let $(\mathcal{A}, \varphi)$ be an non-commutative probability space with $\mathcal{A}$ a unital algebra and $\varphi : \mathcal{A} \to \mathbb{F}$ a linear functional. The centered *subalgebra $\mathcal{A}^c$ is defined to be*

$$\mathcal{A}^c := \ker(\varphi).$$

Elements $\mathbf{a} \in \mathcal{A}$ are called centered *if $\mathbf{a} \in \mathcal{A}^c$. Elements $\mathbf{a} \in \mathcal{A}$ can be centered by*

$$(\mathbf{a})^c := \mathbf{a} - \varphi(\mathbf{a})\mathbf{1}_{\mathcal{A}}.$$

Remark: *Centering in a non-commutative probability space depends on the linear functional φ and is best thought of as analogous to centered classical random variables, which also depend on the chosen measure.*

The analog of classical independence for non-commuting subalgebras is called *free independence* and is defined as follows:

Definition (Free Independence, [78, Definition 5.3 (1)])
Let $(\mathcal{A}, \varphi)$ be a non-commutative probability space and let $\mathcal{I}$ be a fixed index set. Let $\mathcal{A}_i \subset \mathcal{A}$ be a unital subalgebra for each $i \in \mathcal{I}$. The subalgebras $(\mathcal{A}_i, \varphi_i)_{i\in\mathcal{I}}$ are called freely independent w.r.t. *φ, if for every $n \in \mathbb{N}$ we have*

$$\varphi(\mathbf{a}_1\mathbf{a}_2\cdots\mathbf{a}_n) = 0,$$

whenever, for all $k \in \{1, \ldots, n\}$, we have

$$\begin{aligned} \mathbf{a}_k \in \mathcal{A}_{i_k}, \qquad & \text{for indices } i_1 \neq i_2 \neq \cdots \neq i_n && \text{(non-neighboring)} \\ \varphi(\mathbf{a}_k) = 0. & && \text{(centrality)} \end{aligned}$$

All factors $\mathbf{a}_k$ in the product $\mathbf{a}_1\mathbf{a}_2\cdots\mathbf{a}_n$ have to be from different subalgebras than their neighbors, otherwise one could reduce two neighboring elements from the same subalgebra by their algebra multiplication. Furthermore, all factors $\mathbf{a}_k$ have to be centered. So we can combine both conditions into

$$\mathbf{a}_k \in \mathcal{A}_{i_k}^c, \qquad \text{for indices } i_1 \neq i_2 \neq \cdots \neq i_n.$$

The concept of freeness is defined with respect to a linear functional φ and thus depends on it. In classical probability theory *tensor independence* is also defined with respect to a product measure[13]. Subalgebras $(\mathcal{A}_i)_{i\in\mathcal{I}}$ that are free with respect to a linear functional φ therefore need not be free with respect to another linear functional ψ.

Remark: *The index set $\mathcal{I}$ might be chosen to have countable or even uncountable cardinality. Free independence of the family $(\mathcal{A}_i, \varphi_i)_{i\in\mathcal{I}}$ for $|\mathcal{I}| = \infty$ is defined by free independence of $(\mathcal{A}_j, \varphi_j)_{j\in\mathcal{J}}$ for all finite subsets $\mathcal{J} \subset \mathcal{I}$.*

Notation: If it is clear from context to which linear functional we refer we will just speak about freeness and suppress the dependence on the linear functional. This is the case when we fixed a non-commutative probability space $(\mathcal{A}, \varphi)$.

The definition of free independence parallels that for tensor independence in classical probability theory. There, tensor independence (or short *independence*) is also an property of subalgebras. Free independence is related to the kernel of a linear functional φ as it requires that all mixed products stay in the kernel. With slight

[13] The product measure can be induced by an expectation functional via $\mathbb{P}[X \in A] = \mathbb{E}[\mathbb{1}_A(X)]$ for some random variable X and event A.

abuse of notation, one speaks of »independence of subsets $\mathcal{X}_i$« or »independence of random variables X_i« and means independence of the subalgebras generated by this subsets or random variables. For convenience, we will do the same for free independence:

Definition (Free Independence of Subsets $\mathcal{X}_i \subset \mathcal{A}$, [78, Definition 5.3 (2)])
Let $(\mathcal{A}, \varphi)$ be a non-commutative probability space and let $\mathcal{X}_i \subset \mathcal{A}$ be subsets of $\mathcal{A}$. Then $(\mathcal{X}_i)_{i\in\mathcal{I}}$ are called freely independent *if the unital algebras $\mathcal{A}_i := \mathrm{alg}(\mathbf{1}_\mathcal{A}, \mathcal{X}_i)$ generated by $\mathcal{X}_i$ are freely independent.*

Remark: *As a special case one can take subsets $\{\mathbf{a}_i\}$ with just one element $\mathbf{a}_i \in \mathcal{A}$. The $(\mathbf{a}_i)_{i\in\mathcal{I}}$ are called* freely independent non-commutative random variables*. In slight abuse of notation we identify the non-commutative random variable $\mathbf{a}_i$ with the set that contains it.*

If the two subalgebras $\mathcal{A}_1, \mathcal{A}_2$ are freely independent, so are $\mathcal{A}_2, \mathcal{A}_1$. Thus, free independence is commutative.

2.3.3 Asymptotic Freeness

We have now all prerequisites to define when some sets of non-commutative random variables are *asymptotically free*:

Definition (Asymptotic Freeness)
Let the situation be as in the definition of convergence in distribution and let further $\mathcal{I} = \bigcup_{j\in\mathcal{J}} \mathcal{I}_j$ be a decomposition of the index set $\mathcal{I}$ in $|\mathcal{J}|$ disjoint subsets. We say that a sequence of families $\left(\left\{\mathbf{a}_i^{(n)} \;\middle|\; i \in \mathcal{I}_j\right\}\right)_{j\in\mathcal{J}}$ of sets of non-commutative random variables is asymptotically free *as $n \to \infty$, if it converges to a family $(\{\mathbf{a}_i \mid i \in \mathcal{I}_j\})_{j\in\mathcal{J}}$ of sets of non-commutative random variables in some non-commutative probability space $(\mathcal{A}, \varphi)$ and if the limits $(\{\mathbf{a}_i \mid i \in \mathcal{I}_j\})_{j\in\mathcal{J}}$ are free in $(\mathcal{A}, \varphi)$.*

Remark: *The definition of asymptotic freeness is twofold. A sequence of a family of non-commutative random variables has to* converge in distribution *to some family of non-commutative random variables and this family has to be* free*. Thus,* asymptotic freeness *can be seen as approximate freeness. In the definition of asymptotic freeness, the index set $\mathcal{J}$ numerates the $|\mathcal{J}|$ sets of random variables which shall be asymptotically free. If we do not want sets, each possibly with several random variables, to be asymptotically free but the random variables itself, we set $\mathcal{J} := \mathcal{I}$. This gives us the decomposition $\mathcal{I} = \bigcup_{i\in\mathcal{I}} i$ in the elements of the index set $\mathcal{I}$.*

An important results concerns asymptotic freeness of Gaussian and constant matrices:

Theorem 2.6 (Asymptotic Freeness of Gaussian and Constant matrices, [96, pp. 14–15, Theorem 1.])
Let $\mathbf{A}_N$ be a Gaussian $(N \times N)$-random matrix and $\mathbf{D}_N$ a constant (non-probabilistic) matrix, such that all limiting moments

$$\lim_{N\to\infty} \mathrm{tr}_N\big(\mathbf{D}^k\big), \qquad \text{for } k \in \mathbb{N}$$

exist. Then $\mathbf{A}$ and $\mathbf{D}_N$ are asymptotically free.

Proof See [110] ■

A general recipe on how to construct asymptotic free matrices is given by the following theorem:

Theorem 2.7 (Freeness by Random Rotation, [96, pp. 17–18, Theorem 4.])
Let $\mathbf{A}_N$ and $\mathbf{B}_N$ be two sequences of constant $(N \times N)$-matrices with limiting distributions

$$\mathbf{A}_N \to \mathbf{A} \qquad \text{and} \qquad \mathbf{B}_N \to \mathbf{B}.$$

Let $\mathbf{U}_N$ be a Haar unitary $(N \times N)$-random matrix. Then we have

$$\mathbf{A}_N, \mathbf{U}_N\mathbf{B}_N\mathbf{U}_N^\star \to \mathbf{A}, \mathbf{B} \qquad \text{and} \qquad \mathbf{B}_N \to \mathbf{B}.$$

where $\mathbf{A}$ and $\mathbf{B}$ are free.

Proof See [110] ■
As Gaussian matrices are rotational invariant, this also proves Theorem 2.6.

2.3.4 Free Deterministic Equivalents

Applying the concept of asymptotic freeness to random matrices, the former states that the mixed moments of a tuple of random matrices behave, for matrix dimension going to infinity, equal to a tuple of associated operators. So two independent symmetric normalized Gaussian matrices behave in the large dimension limit as two free semicircle elements. By results of Voiculescu in [109, Chapter 4], given a deterministic matrix ensemble that converges to a fixed asymptotic joint distribution and certain random matrix ensembles, the associated limiting operators are free. We view deterministic matrices as special (degenerated) random matrices. The linear functional is just the normalized trace tr_n, because the expectation is always the deterministic quantity itself.

In [97], Speicher et al. give a short introduction to free deterministic equivalents. In [97, Proposition 2.5 and Remark 2.6] they state that if in a non-commutative probability space $(\mathcal{A}_n, \tau_n)$ and $\left(\mathbf{S}_{i\in\mathcal{I}}^{(n)}\right)$, $\left(\mathbf{D}_{j\in\mathcal{J}}^{(n)}\right)$ collections of certain random resp. deterministic matrices for index sets $\mathcal{I}$, $\mathcal{J}$ that converge in distribution, then all

matrices are asymptotically free. The certain random matrices include Haar[14] Unitary, Gaussian and symmetric-Gaussian matrices. In the last paragraph in this section they also remark that this result can be generalized in a straightforward way to the more general class of Wigner instead of Gaussian matrices.

Free deterministic equivalents are also compatible with rectangular non-commutative random variables resp. rectangular matrices.[15]

2.4 Transformations

2.4.1 The $\mathcal{M}$-transform

Definition ($\mathcal{M}$-Transform)
Let $(\mathcal{A}, \varphi)$ be a non-commutative probability space. For $\mathbf{a} \in \mathcal{A}$ and $z \in \mathbb{C}$ the $\mathcal{M}$-Transform *is the formal power series*

$$\begin{aligned} \mathcal{M}_{\mathbf{a}} &: \mathcal{A} \to \mathbb{C}[\![X]\!] \\ \mathbf{a} &\mapsto \mathcal{M}_{\mathbf{a}}(z) = \sum_{k=1}^{\infty} \varphi\left(\mathbf{a}^k\right) z^k = \sum_{k=1}^{\infty} m_{\mathbf{a}}^{(k)} z^k, \end{aligned} \tag{2.7}$$

in the variable $z \in \mathbb{C}$.

Notation: We define the analogous object for a distribution $\mu \in \Sigma^{\emptyset}$, instead of $\mathbf{a} \in \mathcal{A}$. This is consistent, because all moments are given by a distribution μ as well.[16]

The $\mathcal{M}$-transform is a power series with its k^{th} coefficient $m_{\mathbf{a}}^{(k)} = \varphi\left(\mathbf{a}^k\right)$ the k^{th} non-commutative moment of the NCRV $\mathbf{a}$.

Remark: *A formal power series is invertible with respect to composition, if and only if its constant term vanishes and the linear term does not.*[17] *The $\mathcal{M}$-transform does not have a constant, so if $\varphi(\mathbf{a}) \neq 0$ it has a compositional inverse. If we attempt to write the $\mathcal{M}$-transform as formal power series with summand $n = 0$, this first summand would be $\varphi(\mathbf{a}^0)z^0 = \varphi(\mathbf{1}_{\mathcal{A}})z^0 = 1$. But if we were to add this constant to (2.7), the resulting formal power series would not be invertible anymore. We need to multiply by z*

$$z(1 + \mathcal{M}_{\mathbf{a}}(z)) = z\left(1 + \sum_{k=1}^{\infty} m_{\mathbf{a}}^{(k)} z^k\right) = \sum_{k=0}^{\infty} m_{\mathbf{a}}^{(k)} z^{k+1}, \tag{2.8}$$

in order to obtain a transformed $\mathcal{M}$-transform with summation starting at 0 that is invertible with respect to composition.

[14] Introduced 1933 by Haar in [37]. For a definition, see [18, VII 2, Definition 2].
[15] See [97, Section 3.2].
[16] $\Sigma_{\mathbb{F}}$ is the space of normed linear functionals over a field $\mathbb{F}$ and $\Sigma^{\emptyset}$ has non-zero elements.
[17] See [26, pp. 26–27], Proposition 9.1 and the proof thereof.

2.4.2 The Cauchy-Transform $\mathcal{G}$

Definition (Cauchy-Transform $\mathcal{G}$)
Let $(\mathcal{A}, \varphi)$ be a non-commutative probability space. For self-adjoint $\mathbf{a} \in \mathcal{A}$ and $z \in \mathbb{C}$ the Cauchy-transform $\mathcal{G}$ *is the formal power series*

$$\begin{aligned} \mathcal{G}_{\mathbf{a}} &: \mathcal{A} \to \mathbb{C}[\![X]\!] \\ \mathbf{a} &\mapsto \mathcal{G}_{\mathbf{a}}(z) = \sum_{k=0}^{\infty} \varphi\big(\mathbf{a}^k\big) z^{-(k+1)} = \sum_{k=0}^{\infty} m_{\mathbf{a}}^{(k)} z^{-(k+1)}, \end{aligned}$$

in the variable $z \in \mathbb{C}$. If $\mu_{\mathbf{a}}$ is the real-valued distribution of $\mathbf{a}$, we also have for all $z \in \mathbb{C} \backslash \operatorname{supp}(\mu_{\mathbf{a}})$ the analytic expression

$$\mathcal{G}_{\mathbf{a}}(z) = \varphi\big((z\mathbf{1}_{\mathcal{A}} - \lambda)^{-1}\big) = \int_{\mathbb{R}} (z - \lambda)^{-1} \, \mathrm{d}\mu_{\mathbf{a}}(\lambda),$$

where it converges.

Notation: Depending on the context, we will speak about Cauchy-transforms of elements $\mathbf{a}$ of some algebra $\mathcal{A}$ or of Cauchy-transforms of a probability distribution. The latter will often be the distribution of an element $\mathbf{a}$ from a non-commutative probability space $(\mathcal{A}, \varphi)$. We will also use to notation $\mathcal{G}_{\mu}$ for the Cauchy-transform of a probability measure μ.

Remark: *In some works the Cauchy-transform is called Stieltjes-transform*[18]*. In other works, predominantly in free probability theory, the Cauchy-transform is the negative of the Stieltjes-transform*[19]*.*

As we want to work with measures, we would like to reconstruct the measure $\mu_{\mathbf{a}}$ from its Cauchy-transform $\mathcal{G}_{\mu_{\mathbf{a}}}$. This can be done by Theorem A.1. A weak version of the inverse Cauchy-transform theorem states that if $\mathrm{d}\mu_{\mathbf{a}}$ is a continuous density w.r.t. Lebesgue measure λ, we have

$$\lim_{\varepsilon \searrow 0} -\frac{1}{\pi} \operatorname{Im} \mathcal{G}_{\mu_{\mathbf{a}}}(\lambda + \mathrm{i}\varepsilon) \, \mathrm{d}\lambda = \mathrm{d}\mu_{\mathbf{a}}(\lambda). \tag{2.9}$$

The $\mathcal{M}$-transform and $\mathcal{G}$-transform are related to each other via

$$\mathcal{G}_{\mathbf{a}}\big(z^{-1}\big) = \sum_{k=0}^{\infty} m_{\mathbf{a}}^{(k)} z^{(k+1)} \overset{(2.8)}{=} z(1 + \mathcal{M}_{\mathbf{a}}(z)). \tag{2.10}$$

So knowing the distribution of $\mu_{\mathbf{a}}$ amounts to knowing all moments $m_{\mathbf{a}}^{(k)}$ and thus the moment generating function $\mathcal{M}_{\mathbf{a}}(z)$. By (2.10) this information can be transformed

[18] The Cauchy/Stieltjes-transform first appeared in [98] and got its name by the so-called *Stieltjes moment problem* posed by Titchmarsh in [102, pp. 320–322].
[19] The name Stieltjes-transform was probably coined by Widder in [112], see also [113, chapter VIII].

in the Cauchy-transform. From its knowledge one can recover the spectral measure $\mu_{\mathbf{a}}$. This shows us that, by functional calculus, the collection of moments of a non-commutative self-adjoint random variable is related to a real probability measure.

Remark: *The problem to find an inversion formula for the Cauchy resp. Stieltjes-transform was first solved by Wintner in [116].*

2.4.3 $\mathcal{R}$-transform

We already derived the right-hand side (RHS) of (2.10) as left-hand side (LHS) of (2.8). The motivation for this transformed $\mathcal{M}$-transform was that we wanted to ensure invertibility with respect to composition of power series. Therefore, the $\mathcal{G}$-transform $\mathcal{G}_{\mathbf{a}}(z^{-1})$ is invertible with respect to composition of power series and this inverse will be denoted as $\mathcal{G}_{\mathbf{a}}^{\langle -1\rangle}(z)$.

Notation: In the context of power series, we mean by *inverse* the inverse with respect to composition of power series.

The inverse $\mathcal{G}_{\mathbf{a}}^{\langle -1\rangle}(z)$ can be obtained by the Lagrange inversion formula and is given by the Laurent series

$$\mathcal{G}_{\mathbf{a}}^{\langle -1\rangle}(z) = z^{-1} + \sum_{k=0}^{\infty} \kappa_k z^k.$$

The inverse series has only one term of negative power. Subtracting it, in [107] Voiculescu defined the resulting formal power series as the *$\mathcal{R}$-transform* (of $\mathbf{a}$). We will define a modified version of the $\mathcal{R}$-transform and will refer to Voiculescu's original $\mathcal{R}$-transform as *original* $\mathcal{R}$-transform, if needed.

Definition ($\mathcal{R}$-transform (modified))
Let $(\mathcal{A}, \varphi)$ be a non-commutative probability space. For self-adjoint $\mathbf{a} \in \mathcal{A}$ and $z \in \mathbb{C}$ the $\mathcal{R}$-transform *is the formal power series*

$$\begin{aligned} \mathcal{R}_{\mathbf{a}} &: \mathcal{A} \to \mathbb{C}[\![X]\!] \\ \mathbf{a} &\mapsto \mathcal{R}_{\mathbf{a}}(z) = \sum_{k=1}^{\infty} \kappa_k z^k, \end{aligned}$$

where κ_k denotes the k^{th} free cumulant. As with classical cumulants from classical probability theory, free cumulants linearize addition of moments of free non-commutative random variables. For NCRVs $\mathbf{a}, \mathbf{b}$ free, we have the desired behavior, i.e.

$$\mathcal{R}_{\mathbf{a}+\mathbf{b}} = \mathcal{R}_{\mathbf{a}} + \mathcal{R}_{\mathbf{b}}.$$

Remark: *Voiculescu's original $\mathcal{R}$-transform is defined to be*

$$\mathcal{R}_{\mathbf{a}}^{\text{orig.}}(z) \coloneqq \mathcal{G}_{\mathbf{a}}^{\langle -1\rangle}(z) - z^{-1} = \sum_{k=0}^{\infty} \kappa_k z^k.$$

It is related to the modified $\mathcal{R}$-transform by

$$\mathcal{R}_{\mathbf{a}}(z) = z\mathcal{R}_{\mathbf{a}}^{\text{orig.}}(z).$$

If $\varphi(\mathbf{a}) = m_1^{(\mathbf{a})} = \kappa_1^{(\mathbf{a})} \neq 0$, the modified $\mathcal{R}$-transform has a vanishing constant term and a non-vanishing linear term and is therefore invertible. There exists a simple relationship[20] for the inverses of $\mathcal{R}$-transform and $\mathcal{M}$-transform, given by

$$\mathcal{R}_{\mathbf{a}}^{\langle -1\rangle}(z) = (1+z)\mathcal{M}_{\mathbf{a}}^{\langle -1\rangle}(z). \tag{2.11}$$

So getting from an $\mathcal{M}$-transform to a $\mathcal{R}$-transform is simple: One just has to invert, multiply by $(1+z)$ and invert again.

2.4.4 $\mathcal{S}$-transform

As the $\mathcal{R}$-transform is the transform that linearizes addition of free non-commutative random variables, the $\mathcal{S}$-transform does the same for multiplication of free non-commutative random variables.

Definition ($\mathcal{S}$-transform)
Let $(\mathcal{A}, \varphi)$ be a non-commutative probability space. For self-adjoint $\mathbf{a} \in \mathcal{A}$ with $\varphi(\mathbf{a}) \neq 0$ and $z \in \mathbb{C}$ the $\mathcal{S}$-transform *is the formal power series*

$$\begin{aligned} \mathcal{S}_{\mathbf{a}} &: \mathcal{A} \to \mathbb{C}[\![X]\!]_{\emptyset} \\ \mathbf{a} &\mapsto \mathcal{S}_{\mathbf{a}}(z) \coloneqq z^{-1}\mathcal{R}_{\mathbf{a}}^{\langle -1\rangle}(z) = \sum_{k=0}^{\infty} \alpha_k z^k. \end{aligned}$$

For two non-commutative random variables free $\mathbf{a}, \mathbf{b}$, we state the following property:

$$\mathcal{S}_{\mathbf{ab}} = \mathcal{S}_{\mathbf{a}}\mathcal{S}_{\mathbf{b}}$$

Remark: *By the close relationship between the inverse $\mathcal{R}$-transform and inverse $\mathcal{M}$-transform stated in (2.11), one also obtains*

$$\mathcal{S}_{\mathbf{a}}(z) = \frac{1+z}{z}\mathcal{M}_{\mathbf{a}}^{\langle -1\rangle}(z), \tag{2.12}$$

connecting the moments of $\mathbf{a}$ with the $\mathcal{S}$-transform.

[20] See [78, p. 270], and also the remark about invertibility of the $\mathcal{M}$-transform.

2.5 Rectangular Random Matrices

We want to model a class of non-commutative random variables in such a way that they have a property which can be translated as being rectangular. So first, we have to reflect on what are the constitutive properties of some element being *rectangular*. From geometric intuition, a proper rectangle is not rotational symmetric to a perpendicular rotation. On the other hand, the composition of a perpendicular rotation with itself should be the identity with respect to rectangles. On reflexive vector spaces the *adjoint* resembles our deliberations. But it is too weak of an analogy to link rectangularity with non-self-adjointness, because for genuine rectangularity we do not only want the values of such an operator to be different, we also want a whole range of possible values. This stipulation closely resembles the rectangular matrix case, where proper rectangular matrices have different dimensions of their domain and image. As we ultimately motivate this rectangularity to model rectangular random matrices, this perception is valuable.

One could think of just defining a rectangular random matrix as rectangular matrix with its entries classical random variables. Unfortunately, with this definition the rectangular matrices do not have eigenvalues anymore and we also can not use the transformations from section 2.4. Therefore, one would lose the spectral information. One could argue that one could switch from eigenvalues to singular values, but the latter are much harder to analyze. So one seeks quadratic matrices which *behave* like rectangular matrices.

If we assume, for convenience, that the dimension of the image is smaller than the dimension of the domain, this can equivalently be accomplished by constraining the subspace on which the operator maps. The easiest method would be to compose the operator with a projection, where the latter projects onto the constrained subspace. We want to formalize this point of view. So let $(\mathcal{A}, \varphi)$ be a non-commutative probability space and let $(\mathbf{P}_1, \ldots, \mathbf{P}_m)$ be a family of self-adjoint projection matrices that are pairwise orthogonal and which partition unity, i.e. we have $\mathbf{P}_i\mathbf{P}_j = \mathbf{0}$ for $i \neq j$ and $\mathbf{P}_1 + \cdots + \mathbf{P}_m = \mathbf{I}_A$. Then any random matrix $\mathbf{A} \in \mathcal{A}$ can be decomposed as

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{1,1} & \ldots & \mathbf{A}_{1,m} \\ \vdots & & \vdots \\ \mathbf{A}_{m,1} & \ldots & \mathbf{A}_{m,m} \end{pmatrix}, \qquad \text{with } \mathbf{A}_{i,j} \coloneqq \mathbf{P}_i\mathbf{A}\mathbf{P}_j \text{ for all } i, j \in \{1, \ldots, m\}. \tag{2.13}$$

Remark: *This concept was advocated by Benaych-Georges in [13, 12], but the basic idea can be found as early as 1991 in [108, p. 213, 3.6.].*

With the structure depicted in (2.13) we have a unifying framework to describe non-commutative *rectangular* random variables of different sizes, all over one (possibly very big) encompassing non-commutative probability space. We will not need

to compare rectangular matrices of different dimensions, so we only require two orthogonal projections $\mathbf{P}_\alpha + \mathbf{P}_{1-\alpha} = \mathbf{I}_\mathcal{A}$ that divide the whole space into blocks of size $\alpha \times \alpha$, $\alpha \times (1-\alpha)$, $(1-\alpha) \times \alpha$ and $(1-\alpha) \times (1-\alpha)$. The union of such two adjacent blocks can be realized by just considering a composition of a random matrix with just one projection $\mathbf{P}_\alpha$, i.e. either $\mathbf{P}_\alpha \mathbf{A} \mathbf{I}_\mathcal{A}$ or $\mathbf{I}_\mathcal{A} \mathbf{A} \mathbf{P}_\alpha$. We will mostly use this most basic scenario to model non-commutative rectangular random matrices.

The parameter α of a projection matrix $\mathbf{P}_\alpha$ describes the »ratio« of the respective »sizes« of the subspace on which $\mathbf{P}_\alpha$ projects and the subspace which $\mathbf{P}_\alpha$ suppresses. This is also apparent from the measure $\mu_\alpha := \alpha\delta_1 + (1-\alpha)\delta_0$ of a projection $\mathbf{P}_\alpha$: The eigenspace associated with the eigenvalue 1 is the subspace the projection matrix projects onto, the eigenspace associated with eigenvalue 0 gets suppressed. A measure of the ratio of these two eigenspaces is precisely α by construction.

In Figure 2.2 we have depicted schematically what happens to spaces of various sizes when they are being compressed by projections.

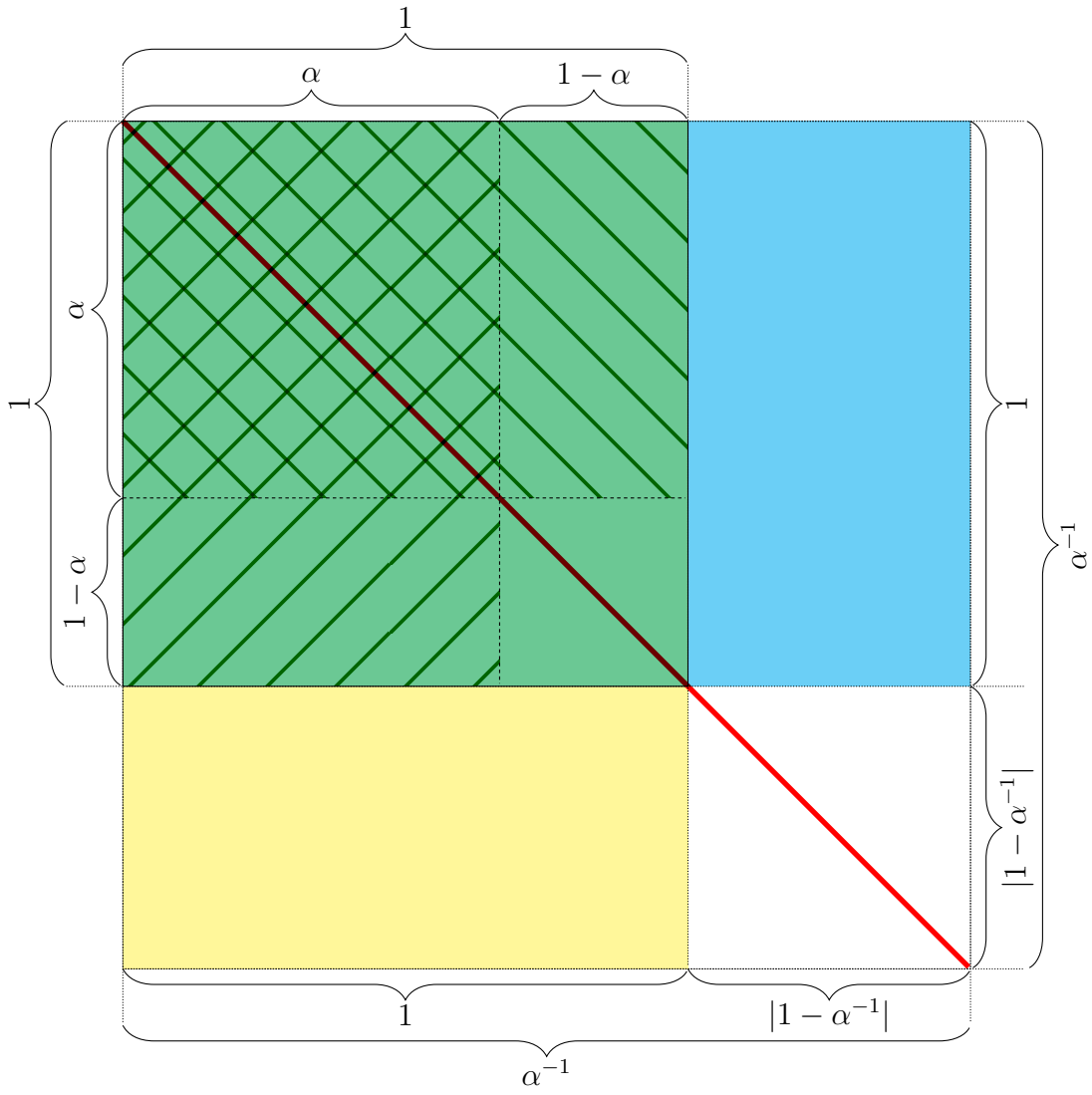

Figure 2.2: Schematic presentation of compression by projection $\mathbf{P}_\alpha$ with $\alpha := \frac{2}{3}$.

First, we want to take the perspective that the main space $\mathcal{A}$ is represented by the solid framed green square of unit length. Therein, we have the $(1 \times \alpha)$ rectangle in green, north-east hatched. This represents the image of $\mathbf{A}\mathbf{P}_\alpha$ for $\mathbf{A} \in \mathcal{A}$. Also in green, but north-west hatched is the rectangular space $(\alpha \times 1)$ of random matrices of the form $\mathbf{P}_\alpha \mathbf{A}$ for $\mathbf{A} \in \mathcal{A}$. The intersection of both spaces has size $(\alpha \times \alpha)$ and is represented by the green square with hatched crosses. A random matrix $\mathbf{P}_\alpha \mathbf{A} \mathbf{P}_\alpha$ is an element of the space of relative size $(\alpha \times \alpha)$. If we are interested in the spectrum of an element $\mathbf{P}_\alpha \mathbf{A} \mathbf{P}_\alpha$ relative to the main space $\mathcal{A}$, it is clear that $1 - \alpha$ of the main space gets suppressed by the projections. Therefore the spectrum, represented by the brown diagonal, must be 0 at a fraction of $1 - \alpha$ and can only be non-zero at the remaining fraction of α.

Second, we want to take the perspective that the main space $\mathcal{A}^{\left(\alpha^{-1}\right)}$ is represented by the big dotted square of length $\alpha^{-1} \coloneqq \frac{3}{2} > 1$. In this »big« space we have a projection matrix $\mathbf{P}_\alpha^{\left(\alpha^{-1}\right)} = \mathbf{P}_1$ that projects again on two third of the main space. The rectangular space generated by right composition with the projection matrix $\mathbf{A}^{\left(\alpha^{-1}\right)} \mathbf{P}_\alpha^{\left(\alpha^{-1}\right)} = \mathbf{A}^{\left(\alpha^{-1}\right)} \mathbf{P}_1$ is represented by the cyan rectangle. On the other hand, the rectangular space generated by left compositions of the form $\mathbf{P}_\alpha^{\left(\alpha^{-1}\right)} \mathbf{A}^{\left(\alpha^{-1}\right)} = \mathbf{P}_1 \mathbf{A}^{\left(\alpha^{-1}\right)}$ is represented by the yellow rectangle. The intersection of both spaces is represented by the green square, which has size $(\alpha \times \alpha)$ relative to the big main space $\mathcal{A}^{\left(\alpha^{-1}\right)}$. An analog argument to the first one holds, so the whole spectrum of dimension α^{-1} has a fraction of $\frac{\left|1-\alpha^{-1}\right|}{\alpha^{-1}}$ that vanishes and a remaining part of $\frac{1}{\alpha^{-1}} = \alpha$.

We can change perspective one more time and consider the main space to be $\mathcal{A}$, represented by the green square. If some random matrices are being projected from an encompassing ambient »big« space to the main space $\mathcal{A}$, from the perspective of $\mathcal{A}$ the spectrum does not need to vanish. Thus, depending on the perspective we choose and the ambient space from which we project, we have to renormalize as just described. From the perspective of the »unit« space $\mathcal{A}$, projection matrices with values $\alpha < 1$ give rise to an atom with probability mass (at least) $1 - \alpha$. On the other hand, for $\alpha > 1$ we imagine a »bigger« ambient space and the resulting distribution has, in general, no atom at 0.

2.6 Scaling

With rectangular random matrices introduced, it makes sense to revisit the different transformations from section 2.4 and examine their behavior under the *dual* rectangular transformation. So let $\mathbf{X} \in \mathbb{R}^{T\times N}$ be a rectangular *long* random matrix, i.e. $T > N$. One can calculate second moments by either

$$\mathbf{A} \coloneqq \mathbf{X}^\intercal\mathbf{X} \quad \text{or} \quad \mathbf{B} \coloneqq \mathbf{X}\mathbf{X}^\intercal.$$

It is well known that the $(T \times T)$ matrix $\mathbf{B}$ has $T - N$ eigenvalues euqal to 0 and the remaining N eigenvalues coincide with the eigenvalues of the $(N \times N)$ matrix $\mathbf{A}$.

Let $\mu_{\mathbf{A}}, \mu_{\mathbf{B}}$ be the resp. limit spectral distribution (LSD) of matrices $\mathbf{A}, \mathbf{B}$. The above translates directly to the measure via

$$\mu_{\mathbf{A}}(\lambda) = \frac{\alpha^{-1}}{\alpha^{-1}-1}\delta_0(\lambda) + \alpha\mu_{\mathbf{B}}(\lambda) \qquad \text{and} \qquad \mu_{\mathbf{B}}(\lambda) = \big(1-\alpha^{-1}\big)\delta_0(\lambda) + \alpha^{-1}\mu_{\mathbf{A}}(\lambda).$$

We can now calculate the scaling of the $\mathcal{G}$-transform to be

$$\begin{aligned}
\mathcal{G}_{\mathbf{B}}(z) &= \int \frac{1}{\lambda - z}\,\mathrm{d}\mu_{\mathbf{B}}(\lambda) \\
&= \int \frac{1}{\lambda - z}\,\mathrm{d}\big(\big(1-\alpha^{-1}\big)\delta_0(\lambda) + \alpha^{-1}\mu_{\mathbf{A}}(\lambda)\big) \\
&= \int \frac{1}{\lambda - z}\,\mathrm{d}\big(1-\alpha^{-1}\big)\delta_0(\lambda) + \int \frac{1}{\lambda - z}\,\mathrm{d}\alpha^{-1}\mu_{\mathbf{A}}(\lambda) \\
&= \big(1-\alpha^{-1}\big)\int \frac{1}{0 - z}\,\mathrm{d}\delta_0(\lambda) + \alpha^{-1}\int \frac{1}{\lambda - z}\,\mathrm{d}\mu_{\mathbf{A}}(\lambda) \\
&= \frac{1-\alpha^{-1}}{-z}\int \mathrm{d}\delta_0(\lambda) + \alpha^{-1}\int \frac{1}{\lambda - z}\,\mathrm{d}\mu_{\mathbf{A}}(\lambda) \\
&= \frac{\alpha^{-1}-1}{z} + \alpha^{-1}\mathcal{G}_{\mathbf{A}}(z).
\end{aligned}$$

An analogous calculation yields

$$\mathcal{G}_{\mathbf{B}}(z) = \frac{1-\alpha^{-1}}{\alpha^{-1}z} + \alpha\mathcal{G}_{\mathbf{A}}(z).$$

3 Eigenvalue Distributions for Specific Random Matrix Models

3.1 Spectral Distributions of Vector-ARMA-Ensembles

We introduced all prerequisites necessary to examine the spectral distribution of sample covariance matrices for samples from certain stochastic processes with specific structure. Up to now, we only introduced the Wishart distribution as eigenvalue distribution for rectangular centered Gaussian matrices. Next, we want to extend this most prominent, but also most basic eigenvalue distribution to situations more involved than the simple Gaussian case. Specifically we want to examine the spectral distribution of the sample covariance matrix for a class of vector–valued autoregressive (AR) moving-average (MA) processes.

Remark: *In this section we mainly follow Burda, Jarosz, et al. and replicate much of their work. But we see this necessary to motivate and show our extensions of their work in [23]. We specifically mention new results not previously known.*

3.1.1 Basic Definitions and Short Introduction to $\mathrm{VARMA}(p, q)$ Processes

Define the *backshift operator* B by its action on some random variable in an indexed collection of classical random variables (X_t) to be

$$\mathrm{B}X_t \coloneqq X_{t-1} \quad \Longrightarrow \quad \mathrm{B}^k X_t = X_{t-k}, \tag{3.1}$$

where the second part is implied by induction. Also, define a *white noise* process as a classical[1] stochastic process with the property that it has zero mean and its auto-covariance $\gamma_{\mathcal{WN}}(\tau)$ vanishes for all lags τ but $\tau = 0$. We denote a white noise process as

$$(Z_t) \sim \mathcal{WN}\big(\sigma^2\big) \quad \text{with auto-covariance function} \quad \gamma_{\mathcal{WN}}(h) = \begin{cases} \sigma^2 & \text{if } h = 0 \\ 0 & \text{if } h \neq 0 \end{cases}.$$

[1] One can also define a stochastic process of non-commutative random variables, but in this context we mean an indexed collection of classical random variables.

Remark: *Every independent distributed stochastic process has auto-covariance function vanishing for lags $\tau \neq 0$. But if the process is not identically distributed the variances of each classical random variable, if they exist, might differ. This implies that every independent and identically distributed stochastic process with finite variance is white noise. The opposite is not true as there exist white noise processes which do not consist of identical distributed random variables.*

Definition (ARMA(p, q))
A stochastic process (X_t) is called an ARMA(p, q) *process if there exist polynomials θ and ϕ of degree p resp. q and a white noise process Z_t such that*

$$\theta(\mathrm{B})X_t(\omega) = \phi(\mathrm{B})Z_t(\omega) \tag{3.2}$$

holds almost surely (a.s.).

Remark: *Note that we did not define* ARMA(p, q) *processes to be stationary*[2]*. Doing so makes the process indirectly defined as the solution to (3.2) unique. But if not otherwise stated we just speak of an* ARMA(p, q) *process without explicitly mentioning the associated white noise process. As there are (at least) 2^{p+q} possible*[3] *white noise processes as solution to the defining* ARMA *equation (3.2), we will use a white noise process that makes the* ARMA(p, q) *processes stationary. Also, questions about* causality *or* invertibility *of an ARMA process are not well posed because both,* causality *and* invertibility *are not properties of an* ARMA *process itself but of an* ARMA *process with respect to a suitable white noise process. This subtlety will prove useful in later applications.*

Definition (Multivariate ARMA$_N(p, q)$ process)
*An N-dimensional vector-valued stochastic process $(\mathbf{X}_t)$ is called an N-*dimensional multivariate ARMA$_N(p, q)$ process *if there exist $(N \times N)$ matrix-valued polynomials $\boldsymbol{\Theta}$ and $\boldsymbol{\Phi}$ of degree p resp. q and an N-dimensional white noise process $(\mathbf{Z}_t) \sim \mathcal{WN}_N(\sigma^2)$ such that*

$$\boldsymbol{\Theta}(\mathrm{B})\mathbf{X}_t(\omega) = \boldsymbol{\Phi}(\mathrm{B})\mathbf{Z}_t(\omega)$$

holds almost surely. The matrix-valued polynomials are given by[4]

$$\begin{aligned} \boldsymbol{\Theta}(x) &\coloneqq \boldsymbol{\Theta}_0 - \boldsymbol{\Theta}_1 x - \cdots - \boldsymbol{\Theta}_q x^q, \\ \boldsymbol{\Phi}(x) &\coloneqq \boldsymbol{\Phi}_0 + \boldsymbol{\Phi}_1 x + \cdots + \boldsymbol{\Phi}_p x^p, \end{aligned} \tag{3.3a}$$

with $\boldsymbol{\Theta}_i, \boldsymbol{\Phi}_j \in \mathbb{R}^{N \times N}$ for all $(i, j) \in \{1, \ldots, q\} \times \{1, \ldots, p\}$.

[2] A stochastic process $(X_t)_{t \in \mathbb{T}}$ is said to be *stationary* if any joint CDF of finite many RVs X_t is invariant w.r.t. translation in time. It is *weakly stationary*, if for all $t \in \mathbb{T}$ the first two moments exist, the first moments are all equal and the auto-covariance does only depend on differences in time. Compare (cf.) [20, p. 12].
[3] This is a consequence of [20, Propositions 3.5.1 and 4.4.2], as one can »invert« every root.
[4] See [20, eq. (11.3.1)].

Note that these matrix-valued polynomials have matrix-valued coefficients but its argument is a (commutative) scalar. So every entry of the matrix-valued $\mathbf{\Theta}(z)$ polynomial is a scalar polynomial of real coefficients and of degree at most q.

Remark: *The auto-covariance at lag $\tau = 0$ is simply the variance and it is captured by σ^2. For convenience and to adhere to convention we normalize at lag $\tau = 0$ and set*

$$\mathbf{\Theta}_0 \coloneqq \mathbf{I}_N \quad \textit{and} \quad \mathbf{\Phi}_0 \coloneqq \mathbf{I}_N. \tag{3.4}$$

Multivariate $\mathrm{ARMA}_N(p, q)$ processes give great flexibility but for situations where we do not have $T \gg N$, there is no realistic chance to ever estimate the whole coefficient matrices with satisfactory precision. Therefore one needs to impose restrictions on the coefficient matrices. We will impose the very strict restriction that all coefficient matrices have to be multiples of the identity matrix $\mathbf{I}_N$. Diagonality decouples the N-dimensional vector-valued $\mathrm{ARMA}_N(p, q)$ process into N scalar-valued $\mathrm{ARMA}_1(p, q)$ processes, so $\mathrm{ARMA}(p, q)$ process k cannot influence $\mathrm{ARMA}(p, q)$ process ℓ anymore for $k \neq \ell$. By the even stricter restriction that the diagonal matrix coefficients have to be multiples of the identity matrix $\mathbf{I}_N$, we insist that every one of the N scalar-valued $\mathrm{ARMA}(p, q)$ processes is an independent copy of all the others. So we just have a vectorization of a scalar-valued $\mathrm{ARMA}(p, q)$ process. We therefore call this type of processes suggestively *vector-autoregressive–moving-average* resp. $\mathrm{VARMA}(p, q)$ processes.

Definition ($\mathrm{VARMA}(p, q)$ process, $\mathrm{VAR}(p)$ process and $\mathrm{VMA}(q)$ process)
A multivariate $\mathrm{ARMA}_N(p, q)$ process with the restriction that the coefficient matrices of its matrix-valued polynomials $\mathbf{\Theta}$ and $\mathbf{\Phi}$ are only multiples of the identity matrix $\mathbf{I}_N$ is called a $\mathrm{VARMA}(p, q)$ process. Specifically, it suffices to know the factors (θ_i) resp. (ϕ_j) for matrix coefficients $\mathbf{\Theta}_i = \theta_i \mathbf{I}_N$ and $\mathbf{\Phi}_j = \phi_j \mathbf{I}_N$ for all $i \in \{1, \ldots, p\}$ and $j \in \{1, \ldots, q\}$.

We further simplify notation by defining the vector-autoregressive (VAR) process $\mathrm{VAR}(p) \coloneqq \mathrm{VARMA}(p, 0)$ *and the vector-moving-average (VMA) process* $\mathrm{VMA}(q) \coloneqq \mathrm{VARMA}(0, q)$.

3.1.2 Assumptions on the General Covariance Matrix and Factorizations

Assume we are given a matrix $\mathbf{x} \in \mathbb{R}^{T \times N}$ of returns for N stocks at T points in time. We want to model each return $(x_{t,n})$ as a classical random variable $X_{t,n}$, so that $(X_{t,n})_{(t,n) \in \{1,\ldots,T\} \times \{1,\ldots,N\}} =: \mathbf{X} \in \mathbb{R}^{T \times N}$ denotes the random matrix comprised of all classical random variables. Clearly, by rearranging the classical random variables in matrix form, one does not lose any information. Depending on our assumptions on the classical random variables, the random matrix $\mathbf{X}$ will have different characteristics. We hope to infer from the characteristics of $\mathbf{X}$ the characteristics of return matrix $\mathbf{x}$.

We are mainly interested in correlations between the random variables. Modulo normalization[5], we are interested in

$$\mathbb{E}_{\mathbf{X}}[X_{t,k}X_{s,\ell}] = C_{t,s,k,\ell},$$

where $C_{t,s,k,\ell}$ denotes the correlation of the return RV of stock k at time t with the return RV of stock ℓ at time s. In this most general setting inference on $C_{t,s,k,\ell}$ by sampling would be futile, because we only have one observation for each combination. Therefore we need to impose some structure in order to generate invariance. This invariance can then be exploited by statistical methods.

As a first restriction on the possible shape of $(C_{t,s,k,\ell})$ we assume that the correlation factors in a cross-correlation $C_{k,\ell}$ and an auto-correlation $A_{t,s}$, that is

$$\mathbb{E}_{\mathbf{X}}[X_{t,k}X_{s,\ell}] = C_{k,\ell}A_{t,s}.$$

The subscript of $\mathbb{E}_{\mathbf{X}}$ denotes that we take the expectation w.r.t. all returns, i.e. the joint distribution of all classical random variables $X_{t,n}$.

The cross-correlation between stock k and stock ℓ is captured by the $N \times N$ cross-correlation matrix $(C_{k,\ell})_{k,\ell\in\{1,\dots,N\}^2} =: \mathbf{C} \in \mathbb{R}^{N\times N}$ for two arbitrary points in time. So we assume that time is an invariant for the cross-correlation. This invariance gives a straightforward recipe on how to determine the cross-correlation from given observations: As we assume that the cross-correlation is stable in time, estimate the cross-correlations at different points in time and average out variations.

The auto-correlation between points in time t and s is captured by the $T \times T$ auto-correlation matrix $(A_{t,s})_{t,s\in\{1,\dots,T\}^2} =: \mathbf{A} \in \mathbb{R}^{T\times T}$ for two arbitrary stocks. Here the specific stocks are invariant with respect to the autocorrelation structure of the whole system. As before, this invariance opens up the possibility for non-trivial and conceptually easy inference.

As the random returns in $\mathbf{X}$ contain both, the cross-correlation and the auto-correlation, we want to normalize:

$$\mathbb{E}_{\mathbf{X}}\left[\mathbf{X}\mathbf{C}^{-1}\mathbf{X}^{\intercal}\mathbf{A}^{-1}\right] = \mathbf{I}_T.$$

Therefore, for a suitable matrix root[6] we have

$$\mathbf{X} = \mathbf{A}^{\frac{1}{2}}\mathbf{Z}\mathbf{C}^{\frac{1}{2}} \quad \Longleftrightarrow \quad \mathbf{Z} = \mathbf{A}^{-\frac{1}{2}}\mathbf{X}\mathbf{C}^{-\frac{1}{2}}, \tag{3.5}$$

where $\mathbf{Z} \in \mathbb{R}^{T\times N}$ is a random matrix of the same size as $\mathbf{X}$ but with vanishing non-trivial[7] correlation.

[5] For centered classical random variables with unit variance the second moment is the correlation. Therefore we will use the term correlation in a broad sense.

[6] A matrix root for a positive-(semi)definite matrix exists and can be obtained by functional calculus. We will always choose the so-called principal square root, which is itself positive-(semi)definite.

[7] By definition, self-correlation is of course 1.

If $\mathbf{X}$ were multivariate Gaussian, the joint density for centered random variables is solely and completely determined by the correlations. The multivariate density function would be

$$p(\mathbf{X}) \coloneqq (2\pi)^{-\frac{TN}{2}} |\mathbf{A}|^{-\frac{N}{2}} |\mathbf{C}|^{-\frac{T}{2}} \exp\left(-\frac{1}{2} \operatorname{Tr}\left(\mathbf{X}\mathbf{C}^{-1}\mathbf{X}^{\intercal}\mathbf{A}^{-1}\right)\right).$$

In terms of the de-correlated matrix $\mathbf{Z}$ the joint density simplifies to

$$p(\mathbf{Z}) \coloneqq (2\pi)^{-\frac{TN}{2}} \exp\left(-\frac{1}{2} \operatorname{Tr}(\mathbf{Z}\mathbf{Z}^{\intercal})\right).$$

3.1.3 Connection of Sample-Covariance Matrix to $\mathcal{M}$-transform of Auto-Correlation Matrix

Capitalizing on the invariance in time a simple estimator for the cross-correlation matrix $\mathbf{C}$ is given by time-averaging over the single second moment estimates $x_{t,k}\tilde{X}_{t,\ell}$ for all t:

$$\hat{\mathbf{c}} \coloneqq \frac{1}{T}\mathbf{x}^{\intercal}\mathbf{x}.$$

By the arguments following (2.5) and (2.6) in combination with Theorem 2.5, by dividing by $T-1$ instead of T the resulting empirical spectral cumulative distribution function differ at most by $\frac{1}{T}$ in the supremum norm. We therefore conclude that the difference is practically negligible for large T. Bearing in mind that dividing by $T-1$ makes the correlation estimator $\hat{\mathbf{c}}$ unbiased, we opt for ease of notation and divide by T. The results still hold for division by $(T-1)$.

In the following, we want to analyze general properties of the correlation estimator $\hat{\mathbf{c}}$. In order to facilitate this goal, we switch perspective and examine the theoretical model $\mathbf{X}$ we imposed for the matrix of return observations $\mathbf{x}$. In addition, if we substitute (3.5) we arrive at

$$\hat{\mathbf{C}} \coloneqq T^{-1}\mathbf{X}^{\intercal}\mathbf{X} = T^{-1}\left(\mathbf{A}^{\frac{1}{2}}\mathbf{Z}\mathbf{C}^{\frac{1}{2}}\right)^{\intercal}\left(\mathbf{A}^{\frac{1}{2}}\mathbf{Z}\mathbf{C}^{\frac{1}{2}}\right) = T^{-1}\mathbf{C}^{\frac{1}{2}}\mathbf{Z}^{\intercal}\mathbf{A}\mathbf{Z}\mathbf{C}^{\frac{1}{2}}, \tag{3.6}$$

where we use the fact that $\mathbf{A}, \mathbf{C}$ are symmetric, and by functional calculus, so is its principal square root.

Without loss of generality assume $T > N$ and set[8] $N/T =: \alpha \in (0,1)$ and define the two projections[9] $\mathbf{P}_\alpha + \mathbf{P}_{1-\alpha} = \mathbf{I}_T \in \mathbb{R}^{T\times T}$ partitioning the unit. Now we can realize all random matrices $\mathbf{A}, \mathbf{C}^{\frac{1}{2}}, \mathbf{Z}$ as rectangular elements over the same rectangular non-commutative probability space $(\mathcal{A}_T, \tau_T)$ for $T \in \mathbb{N}$.[10] Specifically we have

$$\mathbf{C} = \mathbf{P}_\alpha \tilde{\mathbf{C}} \mathbf{P}_\alpha \quad \text{and} \quad \mathbf{Z} = \tilde{\mathbf{Z}}\mathbf{P}_\alpha,$$

[8] By the remarks to Figure 2.2 α has to be defined relativ to the big ambient space $\mathbb{R}^{T\times T}$ and not relativ to $\mathbb{R}^{N\times N}$ so that we can partition into α and $1-\alpha$.

[9] For some basis $\mathbf{P}_\alpha$ is the projection matrix of a specific projection map. The ratio of the dimensions of the eigenspace associated with eigenvalue 1 and the dimension of the whole image is denoted by α.

[10] To ease notation we suppress the dependence of the random matrices on T.

where the matrices with tilde denote random $(T \times T)$ matrices »cut« to the right size by the projection matrix $\mathbf{P}_\alpha$, which itself is a deterministic random matrix.[11] We will keep in mind this construction of rectangular matrices, but for ease of notation we will abstract away as much notation as possible regarding the difference between rectangularized matrices and their squared-sized »big« counterparts.

We want to view all matrix factors of (3.6) as random matrices. This is certainly in line with our construction of $\mathbf{Z}$ and the cross-correlation resp. auto-correlation matrix $\mathbf{C}$ resp. $\mathbf{A}$. They can be viewed as a degenerated special subclass of deterministic matrices, as can be their square roots.

By the theorem about asymptotic freeness of Gaussian and constant matrices the random matrix $\mathbf{Z}$ and the deterministic matrices $\mathbf{P}_\alpha$, $\mathbf{A}$ and $\mathbf{C}$ are asymptotically free, provided $\tilde{\mathbf{Z}}$ is a Gaussian matrix and the asymptotic distribution of the deterministic matrices converges.

Henceforth we will assume both conditions, specifically that $\tilde{\mathbf{Z}}$ is a Gaussian matrix. By universality it is expected that asymptotic freeness stays true for broader classes of random matrices. Projection matrices have an easy spectrum and their distribution obviously converges. For trivial matrices like $\tilde{\mathbf{A}} = \tilde{\mathbf{C}} = \mathbf{I}_T$ this would also be the case, but we will argue later that this is also the case for the resulting matrices we get by applying the free probability toolkit.[11] Mathematically, we can simply assume all matrices to be mutually free, as we provided reasoning that this additional assumption does not reduce the model to the empty set.

Calculating distributions or checking freeness is done with non-commutative monomials in the moments, viewed under the linear functional τ_T. For random matrices, we have as linear functional $\tau_T = \mathrm{tr}_T \otimes \mathbb{E}_{\mathbf{Z}}$. So this linear functional has the trace property and therefore we are permitted to cyclically permute factors.

As by (3.6) $\hat{\mathbf{C}}$ is a product of free factors, we investigate the $\mathcal{S}$-transform of this product as it linearizes multiplication:

$$\begin{aligned}\mathcal{S}_{\hat{\mathbf{C}}}(z) &\overset{(3.6)}{=} \mathcal{S}_{T^{-1}\mathbf{C}^{\frac{1}{2}}\mathbf{Z}^\intercal\mathbf{A}\mathbf{Z}\mathbf{C}^{\frac{1}{2}}}(z) \overset{\text{cycl.}}{=} \mathcal{S}_{T^{-1}\mathbf{Z}^\intercal\mathbf{A}\mathbf{Z}\mathbf{C}}(z) \\ &\overset{\text{mult.}}{=} \mathcal{S}_{T^{-1}\mathbf{Z}^\intercal\mathbf{A}\mathbf{Z}}(z)\mathcal{S}_{\mathbf{C}}(z) \overset{\text{cycl.}}{=} \mathcal{S}_{T^{-1}\mathbf{A}\mathbf{Z}\mathbf{Z}^\intercal}(z)\mathcal{S}_{\mathbf{C}}(z)\end{aligned}$$

Here, *cycl.* denotes cyclic permutation and *mult.* utilizes that factorization of the $\mathcal{S}$-transform of a product of free factors. The $\mathcal{S}$-transform $\mathcal{S}_{\hat{\mathbf{C}}}(z)$ and all subsequent $\mathcal{S}$-transforms are for $(N \times N)$ matrices, as is $\hat{\mathbf{C}}$. We want to isolate the $\mathcal{S}$-transform of the auto-correlation matrix $\mathbf{A}$, but in terms of its $\mathcal{S}$-transforms for $(T \times T)$ matrices. So we proceed by another cycling permutation and transposition:

$$\mathcal{S}_{\hat{\mathbf{C}}}(z) \overset{\text{cycl.}}{=} \mathcal{S}_{T^{-1}\mathbf{A}\mathbf{Z}\mathbf{Z}^\intercal}(z)\mathcal{S}_{\mathbf{C}}(z) \overset{\text{trans.}}{=} \mathcal{S}_{T^{-1}\mathbf{Z}\mathbf{Z}^\intercal\mathbf{A}}\left(\alpha^{-1}z\right)\mathcal{S}_{\mathbf{C}}(z).$$

[11] This also holds true for the matrices $\mathbf{C}^{\frac{1}{2}}$ and $\left(\mathbf{C}^{\frac{1}{2}}\right)^\intercal$.

When transposing, the $\mathcal{S}$-transform has to be rescaled, which is done by scaling its argument by α^{-1}. We utilize once more the factorization property of the $\mathcal{S}$-transform to arrive at

$$\mathcal{S}_{\hat{\mathbf{C}}}(z) \overset{\text{mult.}}{=} \mathcal{S}_{T^{-1}\mathbf{Z}\mathbf{Z}^\intercal}\big(\alpha^{-1}z\big)\mathcal{S}_{\mathbf{A}}\big(\alpha^{-1}z\big)\mathcal{S}_{\mathbf{C}}(z). \tag{3.7}$$

For $\mathbf{Z}$ a rectangular Gaussian matrix the term of the first $\mathcal{S}$-transform in (3.7) is a Wishart matrix with parameter α^{-1}. By Theorem 2.4 the Wishart ensemble has Marchenko–Pastur distribution given by (2.2). By [106, p. 52, eq. (2.87)], the $\mathcal{S}$-transform is given by

$$\mathcal{S}_{T^{-1}\mathbf{Z}\mathbf{Z}^\intercal}\big(\alpha^{-1}z\big) = \frac{1}{1+\alpha^{-1}z}. \tag{3.8}$$

Multiplying both sides of the equation by $z(1+z)^{-1}$ we arrive at

$$\begin{aligned}\frac{z}{1+z}\mathcal{S}_{\hat{\mathbf{C}}}(z) &\overset{(3.8)}{=} \frac{1}{1+\alpha^{-1}z}\mathcal{S}_{\mathbf{A}}\big(\alpha^{-1}z\big)\frac{z}{1+z}\mathcal{S}_{\mathbf{C}}(z) \\ &= \big(\alpha^{-1}z\big)^{-1}\frac{\alpha^{-1}z}{1+\alpha^{-1}z}\mathcal{S}_{\mathbf{A}}\big(\alpha^{-1}z\big)\frac{z}{1+z}\mathcal{S}_{\mathbf{C}}(z).\end{aligned} \tag{3.9}$$

Ultimately we want $\hat{\mathbf{C}}$ to be expressed in simple expressions to be able to connect some of them to real world applications. We aim for the moment-transform $\mathcal{M}$, as it is the power series decoding all information as basic moments. The $\mathcal{M}$-transform and the $\mathcal{S}$-transform are related via the $\mathcal{M}^{\langle -1\rangle}$-transform, which is directly connected to the $\mathcal{S}$-transform by (2.12). We notice that (3.9) has already the right form and we arrive at

$$\mathcal{M}^{\langle -1\rangle}_{\hat{\mathbf{C}}}(z) \overset{(2.12)}{=} \big(\alpha^{-1}z\big)^{-1}\mathcal{M}^{\langle -1\rangle}_{\mathbf{A}}\big(\alpha^{-1}z\big)\mathcal{M}^{\langle -1\rangle}_{\mathbf{C}}(z).$$

As the $\mathcal{M}^{\langle -1\rangle}$-transform is the inverse of the $\mathcal{M}$-transform we can eliminate one term by choosing the argument z appropriately. Since we are ultimately interested in the cross- resp. auto-correlation matrices $\mathbf{C}$ resp. $\mathbf{A}$ we set $z \coloneqq \mathcal{M}_{\hat{\mathbf{C}}}(z)$.

$$\mathcal{M}^{\langle -1\rangle}_{\hat{\mathbf{C}}}(\mathcal{M}_{\hat{\mathbf{C}}}(z)) = \big(\alpha^{-1}\mathcal{M}_{\hat{\mathbf{C}}}(z)\big)^{-1}\mathcal{M}^{\langle -1\rangle}_{\mathbf{A}}\big(\alpha^{-1}\mathcal{M}_{\hat{\mathbf{C}}}(z)\big)\mathcal{M}^{\langle -1\rangle}_{\mathbf{C}}(\mathcal{M}_{\hat{\mathbf{C}}}(z)).$$

At the left-hand side we get z by construction, we ease notation on the right-hand side by abbreviating $M \coloneqq \mathcal{M}_{\hat{\mathbf{C}}}(z)$ and arrive at

$$z = \big(\alpha^{-1}M\big)^{-1}\mathcal{M}^{\langle -1\rangle}_{\mathbf{A}}\big(\alpha^{-1}M\big)\mathcal{M}^{\langle -1\rangle}_{\mathbf{C}}(M). \tag{3.10}$$

We managed to get an equation with only transforms of the cross- and auto-correlation appearing. To simplify further we have to pose assumptions on either of the correlation matrices. We are mainly interested in the dynamics for dependence structures in time, i.e. the auto-correlation $\mathbf{A}$. To make matters simple let us assume the true cross-correlation matrix $\mathbf{C}$ to be the identity matrix $\mathbf{I}_N$. This is clearly an oversimplification of the true cross-correlation matrix — but it is virtually the same oversimplification as assuming the auto-correlation $\mathbf{A}$ to be the identity matrix $\mathbf{I}_T$ as it is normally done in the literature. Equation (3.10) is a central starting point for further research on other correlation structures.

The moments of the identity matrix $\mathbf{C} = \mathbf{I}_N$ are given by $\tau_{\mathcal{A}}\big(\mathbf{I}_N^k\big)$ for $k \in \mathbb{N}$. As the identity matrix is idempotent, all higher moments for $k \geq 2$ coincide with the first

moment $\tau_{\mathcal{A}}(\mathbf{I}_N) = 1$. Because of (2.7), the moment series respectively $\mathcal{M}$-transform of $\mathbf{C}$ is given by

$$\mathcal{M}_{\mathbf{I}}(z) = \sum_{k=1}^{\infty} z^k = \frac{z}{1-z}. \tag{3.11}$$

The inverse of (3.11) is the $\mathcal{M}^{\langle -1 \rangle}$-transform and it is given by

$$\mathcal{M}_{\mathbf{I}}^{\langle -1 \rangle}(z) = \frac{z}{1+z}. \tag{3.12}$$

We substitute (3.12) in (3.10) and obtain

$$\begin{aligned} z &= \alpha M^{-1} \mathcal{M}_{\mathbf{A}}^{\langle -1 \rangle}\big(\alpha^{-1} M\big) \frac{M}{1+M} \\ &= \alpha \mathcal{M}_{\mathbf{A}}^{\langle -1 \rangle}\big(\alpha^{-1} M\big) M^{-1} \frac{M}{1+M} \frac{1+M}{M} \frac{M}{1+M} \\ &= \frac{\alpha}{1+M} \mathcal{M}_{\mathbf{A}}^{\langle -1 \rangle}\big(\alpha^{-1} M\big) \\ \iff \quad \frac{z(1+M)}{\alpha} &= \mathcal{M}_{\mathbf{A}}^{\langle -1 \rangle}\big(\alpha^{-1} M\big). \end{aligned}$$

We are finally in a position to invert the $\mathcal{M}^{\langle -1 \rangle}$-transform and get

$$\mathcal{M}_{\mathbf{A}}\left(\frac{z(1+M)}{\alpha}\right) = \frac{M}{\alpha}. \tag{3.13}$$

This is the final equation for the auto-correlation matrix. Starting from the sample covariance matrix $\hat{\mathbf{c}}$ we switched to an ex-ante investigation and wrote $\hat{\mathbf{C}}$ as product of free random matrices. By some algebra and simplifying assumptions we finally were able to derive equation (3.13), which connects MLE $\hat{\mathbf{C}}$ to the moments of the auto-correlation matrix $\mathbf{A}$.

We are now in a position that if we can work out closed form solutions to a model of auto-correlation dynamics, we can derive the limiting spectral density. Of course, this argument holds only in the large T, N limit for $\frac{N}{T} \to \alpha$, because we argued with asymptotic freeness which is only available in the large T, N limit.

Remark: *The idea of connecting the various $\mathcal{M}$-transform stems back to at least [22, eq. (22)], but the resulting equation is implicit as it has an $\mathcal{M}$-transform on the left and on the right-hand side as a function argument. In [24] Burda, Jurkiewicz, et al. developed a formula comparable to (3.10), but with different techniques and much more effort. The physicist Burda, Jarosz, et al. took in [23, eq. (14)] a similar route as we did, but their proof omitted some technical details and lacked some foundation. Nevertheless, the ideas from this paper motivated and influenced the derivation presented here. Their [23, eq. (14)] is essentially our (3.13). But Burda, Jurkiewicz, et al. defined their $\mathcal{M}$-transform in a slightly unusual way which amplifies subsequent differences. They essentially work with inverse (and rescaled) arguments than the ones we use and they opted to not use the $\mathcal{S}$-transform as a natural transform for working with products of free matrices.*

Our subsequent arguments in this line of reasoning are qualitatively similar to theirs modulo rescaling. To our knowledge we are the first to apply this line of reasoning and the theoretical results to financial return time series and their specific properties.

3.1.4 The $\mathcal{M}$-transform of VARMA(p, q) Processes

Recapitulating that in order to simplify (3.10) we chose to model the auto-correlation at the price of trivial cross-correlation, we have to sastisfy this restriction and can only model dynamics with non-trivial auto-correlation, but trivial cross-correlation. As defined on p. 31, VARMA(p, q) processes are a suitable model class.

Assume we are given the N-dimensional stationary vector-valued $\mathbf{X}_t \sim \text{VARMA}_N[p, q]$. By construction, each of the N scalar-valued processes is an identical copy of all the others and thus there cannot be any cross-correlation present. So we only have the auto-correlation left:

$$\mathbb{E}_{\mathbf{X}}[X_{t,k} X_{s,\ell}] = \delta_{k,\ell} A_{t,s}.$$

Let $\gamma_{(\mathbf{X}_t)}(\tau)$ be the auto-correlation function of the VARMA(p, q) process $(\mathbf{X}_t)$. Because $(\mathbf{X}_t)$ is stationary, the entries $A_{t,s}$ of its auto-correlation matrix $\mathbf{A}$ do not depend on the specific points in time t, s, but only on absolute differences $\tau := |t - s|$. The auto-correlation matrix $\mathbf{A}$ has identical entries $\gamma_{(\mathbf{X}_t)}(0)$ on its main diagonal, entries $\gamma_{(\mathbf{X}_t)}(|\pm 1|)$ on its sub- and super diagonal and so forth:

$$\mathbf{A} = \begin{bmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \cdots & & & \\ \gamma(1) & \gamma(0) & \gamma(1) & \gamma(2) & \cdots & & \\ \gamma(2) & \gamma(1) & \gamma(0) & \gamma(1) & \gamma(2) & \cdots & \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & & \ddots & \ddots & \ddots & \vdots \\ & & & & \ddots & \ddots & \gamma(1) \\ & & & & \cdots & \gamma(1) & \gamma(0) \end{bmatrix}$$

So one gets the value of the $(t, s)^{\text{th}}$ entry of $\mathbf{A}$ by $\gamma_{(\mathbf{X}_t)}(|t - s|) = \gamma_{(\mathbf{X}_t)}(\tau)$, i.e. all the information of the auto-correlation matrix is contained in the auto-covariance function $\gamma_{(\mathbf{X}_t)} : \mathbb{Z} \to \mathbb{R}$.

Because of the *Wold decomposition* every covariance-stationary stochastic process has a decomposition in a, possible infinite, MA process with square-summable coefficients and a purely deterministic process.[12] Applying the Wold decomposition to ARMA(p, q) processes one gets its representation as a, possible infinite, MA process and the purely deterministic process vanishes[13]. But it is known[14] that the coefficients of an ARMA(p, q) auto-covariance are absolutely summable, which

[12] See [20, Theorem 5.7.1].
[13] See [20, p. 189].
[14] See [20, p. 111].

is a slightly stronger property as the square summability implied by the Wold decomposition. By iterating on the interpretation of the MA coefficients (ψ_k), one directly has that

$$\gamma_{\mathrm{MA}(\infty)}(\tau) \coloneqq \sum_{k=0}^{\infty} \psi_k \psi_{k+|\tau|}, \tag{3.14}$$

where $\psi_k \psi_{k+|\tau|}$ on the RHS are the coefficients of the ARMA(p, q) process and

$$\psi(z) \coloneqq \sum_{k=0}^{\infty} \psi_k z^k = \frac{\phi(z)}{\theta(z)} \quad (\forall z \in \mathbb{C},\ |z| \leq 1).$$

We set the coefficients ψ_k of the MA representation of the stationary ARMA(p, q) process to zero for some $k > \tilde{q}$ if the MA representation is MA$(\tilde{q})$ with $\tilde{q} < \infty$.

The auto-correlation function with values in $[0, \infty)$ is by its very meaning defined in the *time* domain that we assumed to be discrete and which spans all integer numbers. This domain $\mathbb{Z}$ is not nicely analytically tractable. We therefore want to transform the auto-covariance function in such a way that both its domain and its range have nice topologies and the auto-covariance function is thus easy to work with. In addition, its new domain should have a nice statistical interpretation to foster imagination. As the other parts of our model are in the *frequency* domain, we therefore want the auto-correlation function $\gamma_{(\mathbf{X}_t)}(\tau)$ to act on the frequency domain as well. The requested transform is exactly the discrete Fourier transform applied to the auto-correlation function, the resulting quantity is the so-called *spectral density*:

Definition (Spectral Density)
The spectral density *of a stochastic process (X_t) is the discrete[15] Fourier transform[16] of its auto-covariance function,*

$$\tilde{\gamma}_{(X_t)}(\omega) \coloneqq \frac{1}{2\pi} \sum_{\tau \in \mathbb{Z}} \mathrm{e}^{-\mathrm{i}\tau\omega}\, \gamma_{(X_t)}(\tau), \tag{3.15}$$

whenever the auto-covariance series is absolutely convergent, uniformly in $\omega \in (-\pi, \pi]$.

For a fixed value τ the function $\mathrm{e}^{-\mathrm{i}\tau\omega}$ with argument $\omega \in (-\pi, \pi]$ runs though the unit circle, therefore by Euler's formula[17] ω parametrizes all sinusoidal basis functions for the Fourier transformation.

Remark: *For stationary stochastic processes one can expand the RHS of (3.15) and apply trigonometric identities and de Moivre's formula to arrive at*[18]

$$\tilde{\gamma}_{(X_t)}(\omega) = \frac{1}{2\pi} \sum_{\tau \in \mathbb{Z}} \mathrm{e}^{-\mathrm{i}\tau\omega}\, \gamma_{(X_t)}(\tau) \overset{18}{=} \frac{1}{2\pi} \left(\gamma_0 + 2 \sum_{k=0}^{\infty} \gamma_k \cos(k\omega) \right). \tag{3.16}$$

[15] Henceforth, we omit to mention the property *discrete* for the Fourier transform if there is no room for confusion.

[16] We let the factor of $(2\pi)^{-1}$ be present at the Fourier transform and do not split it between the Fourier transform and its inverse to adhere to convention.

[17] See [31, §138, p. 104].

[18] See [60, eq. (2.1)].

For ARMA(p,q) *processes the coefficients* (γ_k) *of the autocorrelation function* γ *are related to the coefficients of the* ARMA *polynomials because of* (3.14) *by*

$$\gamma_{\mathrm{MA}(q)}(\tau) \coloneqq \sum_{j=0}^{q-|\tau|} \psi_j \psi_{j+|\tau|}. \tag{3.17}$$

The spectral density of the ARMA(p,q) process (X_t) implicitly specified by the equation $\theta(\mathrm{B})X_t = \phi(\mathrm{B})Z_t$, with $Z_t \sim \mathcal{WN}(\sigma^2)$, is given by[19]

$$\tilde{\gamma}_{(X_t)}(\omega) = \frac{\sigma^2}{2\pi}\left|\frac{\phi(\mathrm{e}^{-\mathrm{i}\omega})}{\theta(\mathrm{e}^{-\mathrm{i}\omega})}\right|^2, \qquad \omega \in (-\pi, \pi]. \tag{3.18}$$

For the auto-correlation instead the auto-covariance function, the variance σ^2 is normed to 1.

With the spectral density we have a quantity that encodes the same information as the auto-correlation function or the auto-correlation matrix of the ARMA(p,q) process (X_t), but which also exist in the frequency domain and is analytically tractable. This last property is crucial and being used in [23, Appendix A.3, eq. (A.1)] to connect $\mathcal{G}$-transform and spectral density by

$$\mathcal{G}_{\mathbf{A}}(z) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{z - 2\pi\tilde{\gamma}_{(X_t)}(\omega)}\,\mathrm{d}\omega, \tag{3.19}$$

with $\tilde{\gamma}_{(X_t)}(\omega)$ the spectral density of (X_t) and $\mathbf{A}$ its autocorrelation function.

Remark (Exploiting symmetries): *Because the auto-correlation matrix* $\mathbf{A}$ *is real and symmetric, the auto-correlation function* $\gamma(\tau)$ *is a real symmetric function and the spectral density is also a real symmetric and non-negative function. Therefore it suffices, by symmetry, to only know the values of* $\tilde{\gamma}_{(X_t)}(\omega)$ *in the range* $\omega \in [0,\pi]$. *Thus one could calculate integral* (3.19) *just between* 0 *and* π *and double its value.*

In a final step we transform to the $\mathcal{M}$-transform and arrive at

$$\begin{aligned}
\mathcal{M}_{\mathbf{A}}(z) &\overset{(2.10)}{=} z^{-1}\mathcal{G}_{\mathbf{A}}\big(z^{-1}\big) - 1 \overset{(3.19)}{=} \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{z^{-1}}{z^{-1} - \tilde{\gamma}_{(X_t)}(\omega)} - 1\,\mathrm{d}\omega \\
&\overset{\text{sym.}}{=} \frac{1}{\pi}\int_{0}^{\pi} \frac{z^{-1}}{z^{-1} - \tilde{\gamma}_{(X_t)}(\omega)} - 1\,\mathrm{d}\omega \\
&\overset{(3.18)}{=} -1 + \frac{1}{\pi}\int_{0}^{\pi} \frac{1}{1 - \left|\frac{\phi(\mathrm{e}^{-\mathrm{i}\omega})}{\theta(\mathrm{e}^{-\mathrm{i}\omega})}\right|^2 z}\,\mathrm{d}\omega.
\end{aligned} \tag{3.20}$$

[19] See [20, Theorem 4.4.2].

3.1.5 Calculation of distribution generating polynomials for various special VARMA(p, q) models

We are now in a position where we can apply the results from subsection 3.1.4 to (3.13).

If we combine (3.13) with (3.20), we get

$$0 \overset{(3.13)}{=} \mathcal{M}_{\mathbf{A}}\left(\frac{z(1+M)}{\alpha}\right) - \frac{M}{\alpha}. \tag{3.21a}$$

$$\overset{(3.20)}{=} -1 + \frac{1}{\pi}\int_0^\pi \frac{1}{1 - \left|\frac{\phi(\mathrm{e}^{-\mathrm{i}\omega})}{\theta(\mathrm{e}^{-\mathrm{i}\omega})}\right|^2 \frac{z(1+M)}{\alpha}}\,\mathrm{d}\omega - \frac{M}{\alpha}$$

$$= \frac{1}{\pi}\int_0^\pi \frac{\alpha}{\alpha - (1+M)z\left|\frac{\phi(\mathrm{e}^{-\mathrm{i}\omega})}{\theta(\mathrm{e}^{-\mathrm{i}\omega})}\right|^2}\,\mathrm{d}\omega - \frac{M+\alpha}{\alpha}. \tag{3.21b}$$

In principle, this equation can be solved for M provided one is able to calculate the integral. The solution will be a function for M in the variables z, α and the ARMA(p, q) coefficients. Since z is complex, M will in general be complex, too. Remembering the notational abbreviation $M \coloneqq \mathcal{M}_{\hat{\mathbf{C}}}(z)$ before (3.10), this complex function $\mathcal{M}_{\hat{\mathbf{C}}}(z, \alpha, (\psi_k))$ encodes all assumptions and information of the ML estimator $\hat{\mathbf{C}}$ of a VARMA(p, q) process. As it still is the $\mathcal{M}$-transform for the ML estimator $\hat{\mathbf{C}}$, by transforming back to the $\mathcal{G}$-transform and applying the inverse Cauchy-transform theorem (2.9) we finally arrive at the spectral distribution function for $\hat{\mathbf{C}}$.

Unfortunately, (3.21b) will not have a unique solution for M, so one has to test every possible solution whether it possesses all properties of a valid $\mathcal{M}$-transform resp. if the implied $\mathcal{G}$ possesses all properties of a true $\mathcal{G}$-transform. These properties are specifically that $\lim_{|z|\to\infty} z\mathcal{G}(z) = 1$, for $z \in \mathbb{C}^+ \coloneqq \{a + \mathrm{i}b \mid a, b \in \mathbb{R},\, b > 0\}$,[20] and that $\mathcal{G}$ is analytic and that $\mathcal{G} : \mathbb{C}^+ \to \mathbb{C}^-$.[21] But it is often more practical to simply check whether the solution under review gives a probability density function when the inverse Cauchy-transform theorem is applied to the »$\mathcal{G}$-transform« resulting from the specific M. Usually we apply both tests, that is we discard potential solutions if the respective $\mathcal{G}$-transform fails to have the necessary properties. After that we check by integration whether the remaining solutions generate a viable probability density function whose infinite integral is approximately unity. In almost all cases this procedure suffices to choose the appropriate solution of (3.21b) that generates the probability density function for the ML estimator $\hat{\mathbf{C}}$.

Let us give some results for specific VARMA(p, q) models. We will see that for VARMA(p, q) models (3.21a) gives a polynomial in M. Since one root of this polynomial is the $\mathcal{M}$-transform $\mathcal{M}_{\hat{\mathbf{C}}}(z)$ that lets one subsequently calculate the probability density function of the eigenvalues of $\mathbf{C}$, we call this polynomial a »distribution generating polynomial« (distribution generating polynomial (DGP)).

[20] Set $\mathbb{C}^- \coloneqq -\mathbb{C}^+$.

[21] See [78, Remark 2.19 (1), eq. (2.19)] and subsection 2.4.2.

VMA(q) processes

We start with VMA(q) = VARMA($0, q$) and note that this consists of independent copies of MA(q) processes. The spectral density of a MA(q) process (X_t) is by (3.18) given by

$$\tilde{\gamma}_{(X_t)}(\omega) = \frac{\sigma^2}{2\pi}\left|\phi\left(\mathrm{e}^{-\mathrm{i}\omega}\right)\right|^2, \qquad \omega \in (-\pi, \pi]. \tag{3.22}$$

By the definition of VMA(q) processes and the ϕ polynomial[22] (3.3a) and subsequent application of trigonometric identities and de Moivre's formula one arrives at

$$\begin{aligned}\tilde{\gamma}_{\mathrm{MA}(q)}(\omega) &\overset{(3.16)}{=} \frac{1}{2\pi}\left(\gamma_0 + 2\sum_{\tau=1}^{q}\gamma_\tau\cos(\tau\omega)\right)\\ &\overset{(3.17)}{=} \frac{1}{2\pi}\left(\sum_{j=0}^{q}\phi_j\phi_{j+|\tau|} + 2\sum_{\tau=1}^{q}\left(\cos(\omega)\sum_{j=0}^{q-\tau}\phi_j\phi_{j+|\tau|}\right)\right).\end{aligned} \tag{3.23}$$

Remembering (3.4), i.e. $\phi_0 = 1$, we observe that (3.23) is a function in $2\cos(\omega)$. Applying iteratively de Moivre's formula resp. using the trigonometric identity of Chebyshev polynomials of the first kind and cos one can rewrite every occurrence of $2\cos(\tau\omega)$ for some $\tau \in \mathbb{Z}$ as polynomial in $2\cos(\omega)$. If $(T_n)_{n\in\mathbb{N}}$ are the Chebyshev polynomials of the first kind one has $\cos(\tau\omega) = T_\tau(\cos(\omega))$. Since T_n is defined by a recurrence relation this eliminates the integer factors in the cos argument and gives a polynomial in $\cos(\omega)$ of order τ. Doing so for every $\tau \in \{1, \dots, q\}$ one gets a polynomial in $2\cos(\omega)$ of order q. The factor 2 is present for each occurrence of cos and thus stays.

Next, we change the integration variable to $x := 2\cos(\omega)$, which implies $\omega = \arccos\left(\frac{x}{2}\right)$ for $\omega \in [0, \pi]$ and thus $\mathrm{d}x = -\frac{d\omega}{\sqrt{4-x^2}}$. The principal values of arccos are in $[0, \pi]$ for arguments in $[-1, 1]$, so the integration limits have to be $\{-2, 2\}$.

Comparing with (3.19) we finally have to consider $z - 2\pi\tilde{\gamma}_{(X_t)}(\omega)$, where the factor -2π of $\tilde{\gamma}_{(X_t)}$ is easily incorporated and z is added to the constant term. The resulting polynomial in x can be written as $c_q\prod_{k=1}^{q}(x - x_k)$, with (x_k) the q roots and c_k the coefficient of x^q.[23] So by (3.19) we have to integrate

$$\mathcal{G}_{\mathrm{VMA}(q)}(z) = \frac{1}{\pi}\int_{-2}^{2}\frac{1}{z - c_q\prod_{k=1}^{q}(x - x_k)}\frac{-1}{\sqrt{4-x^2}}\,\mathrm{d}x.$$

This integral has been solved in the literature, one gets[24]

$$\mathcal{G}_{\mathrm{VMA}(q)}(z) = \frac{1}{c_q}\sum_{\tau=1}^{q}\frac{1}{\prod_{\substack{k=1\\k\neq\tau}}^{q}(x_\tau - x_k)}\frac{1}{\sqrt{x_\tau - 2}}\frac{1}{\sqrt{x_\tau + 2}}. \tag{3.24}$$

[22] See (3.3a) with $N = 1$ or [20, Example 4.4.1].

[23] For ease of exposition, we assume all roots (x_k) to be pairwise different.

[24] See [23, eq. (A.3)], but note the slight error in the integration measure induced by the change of variables, one has to substitute $q^2 \mapsto y^2$ in their notation.

Remark: *All distribution generating polynomials were calculated with aid of the computer algebra system Mathematica.*[25] *We checked the simpler calculations for* VMA(1) *and* VAR(1) *»by hand« and also calculated diagnostic statistics to test whether certain mathematical objects possess the theorized properties at least in a statistical sense. For instance, we checked that a proper probability density function should integrate to* 1.

Let us give some explicit solutions. For VMA(1) one gets for (3.21a) a polynomial in M of order 4. The polynomial reads

$$\begin{aligned}
&M^4\big(\phi_1^2-1\big)^2z^2\\
&-M^3\Big(2z\Big(\big(\phi_1^2+1\big)\alpha-\big(\phi_1^2-1\big)^2(\alpha+1)z\Big)\Big)\\
&-M^2\Big(-\big(\phi_1^2-1\big)^2((\alpha+4)\alpha+1)z^2+2\big(\phi_1^2+1\big)(2r+1)\alpha z-\alpha^2\Big)\\
&-M^1\Big(2\alpha\Big(-\big(\phi_1^2-1\big)^2(\alpha+1)z^2+\big(\phi_1^2+1\big)(\alpha+2)\alpha z-\alpha^2\Big)\Big)\\
&-\alpha^2z\Big(2\big(\phi_1^2+1\big)\alpha-\big(\phi_1^2-1\big)^2z\Big).
\end{aligned}\tag{3.25}$$

Since polynomials over $\mathbb{C}$ up to order 4 can in general be solved analytically, polynomial (3.25) has a closed solution for each of its 4 roots. But these 4 roots are very complicated objects. We use Mathematica's `LeafCount` function[26] as a proxy of term length and complexity. The polynomial (3.25) has a `LeafCount` of 147. In contrast, each of the 4 roots has a `LeafCount` of 13 020.

For VMA(2) we apply (3.24) and subsequently transform back to the $\mathcal{M}$-transform and solve (3.21a). The resulting polynomial in M is of order 9 and has a `LeafCount` of 4 227. Because the polynomial degree exceeds 4 for general parameters ϕ_1, ϕ_2, r, there do not exist analytic solutions for its roots. Therefore solutions have to be approximated numerically. One sees from the sheer length of the expression that this is quite cumbersome. The explicit expression can be found in subsection A.3.1.

Remark: *As mentioned in* [24]*, Burda, Jarosz, et al. calculated* (3.24)*. But to the best of our knowledge we are the first to follow the described procedure for something different than* VMA(1)*,* VAR(1) *or* VARMA(1, 1)*. Specifically we are the first to utilize* (3.24) *in order to calcuate the distribution generating polynomial for the* VMA(2) *case. As we describe below, the hard part is to nummerically find the suitable root and construct an algorithm that returns the spectral density of an* VMA(2) *process. The same holds true for* VAR(2) *processes.*

[25] See [118].

[26] Mathematica's `LeafCount` gives the total number of indivisible subexpressions of an expression.

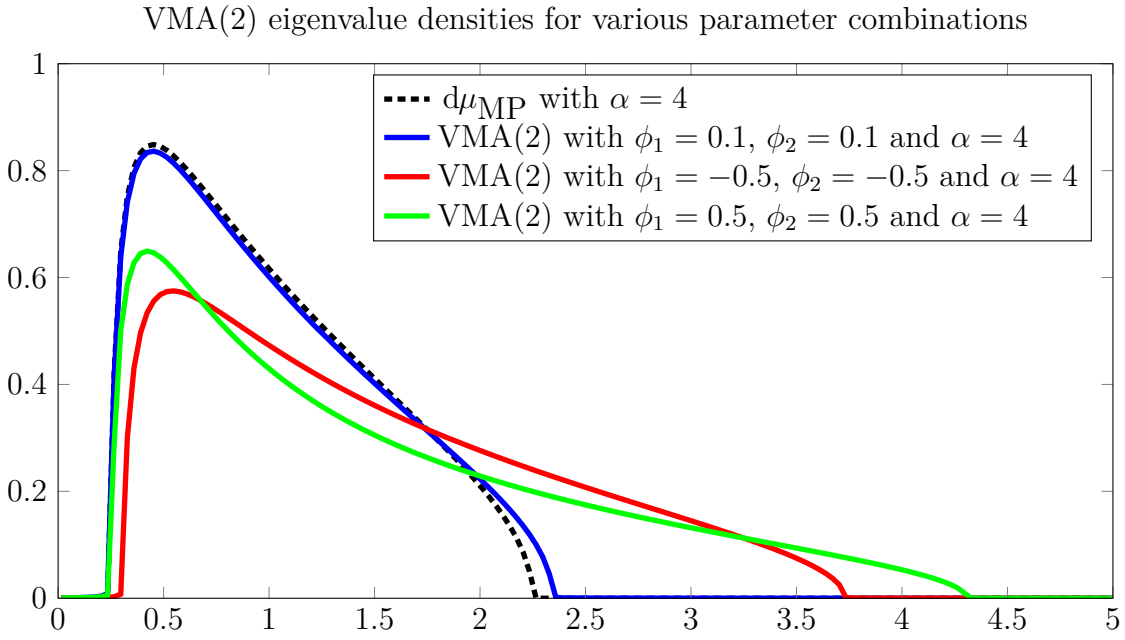

Figure 3.1: Comparison of VMA(2) eigenvalue densities for various parameter combinations with Marchenko–Pastur law, all with $\alpha = 4$.

VAR(q) processes

We note again that a VAR(p) process $(\mathbf{X}_t)$ consists of independent copies of AR(p) processes (X_t). By (3.18), the spectral density of an AR(p) process is

$$\tilde{\gamma}_{(X_t)}(\omega) = \frac{1}{2\pi}\left|\frac{1}{\theta(\mathrm{e}^{-\mathrm{i}\omega})}\right|^2, \qquad \omega \in (-\pi, \pi].$$

Incidentally, the spectral density is connected to the spectral density of a MA(p) process shown in (3.22). Regarding the $\theta(p)$ polynomial, all coefficients $(\theta_k)_{k\in\{1,\dots,p\}}$ have negative sign and all the coefficients $(\phi_k)_{k\in\{1,\dots,p\}}$ of the $\phi(p)$ polynomial have positive sign. So if one sets $\phi_k \coloneqq -\theta_k$ for all $k \in \{1, \dots, p\}$ and inverts, one arrives at a MA(p) representation. We therefore transform the original $\theta(p)$ coefficients (θ_k) as explained and invert the autocorrelation function, which in turn gives us the spectral density in the MA(p) parametrization we already solved.

For VAR(1) the explicit distribution generating polynomial as solution to (3.21a) reads

$$\begin{aligned}
&M^4 z^2 \\
&+ M^3\big(2z^2 - 2\big(\theta_1^2 + 1\big)\alpha z\big) \\
&+ M^2\big(\big(\theta_1^2 - 1\big)^2\alpha^2 - 2\big(\theta_1^2 + 1\big)\alpha z - \big(\alpha^2 - 1\big)z^2\big) \\
&- 2M\alpha^2 z^2 \\
&- \alpha^2 z^2,
\end{aligned}$$

and is a polynomial of order 4. The `LeafCount` is 86, but any of the roots in closed form has a `LeafCount` of 3 836. In Figure 3.2 we compare the probability density for the eigenvalues calculated with the above described method with simulation results of appropriate VAR(1) matrices with an autocorrelation of $\theta_1 = \rho = \frac{1}{2}$ and a rectangularity ratio of $\alpha = 4$.

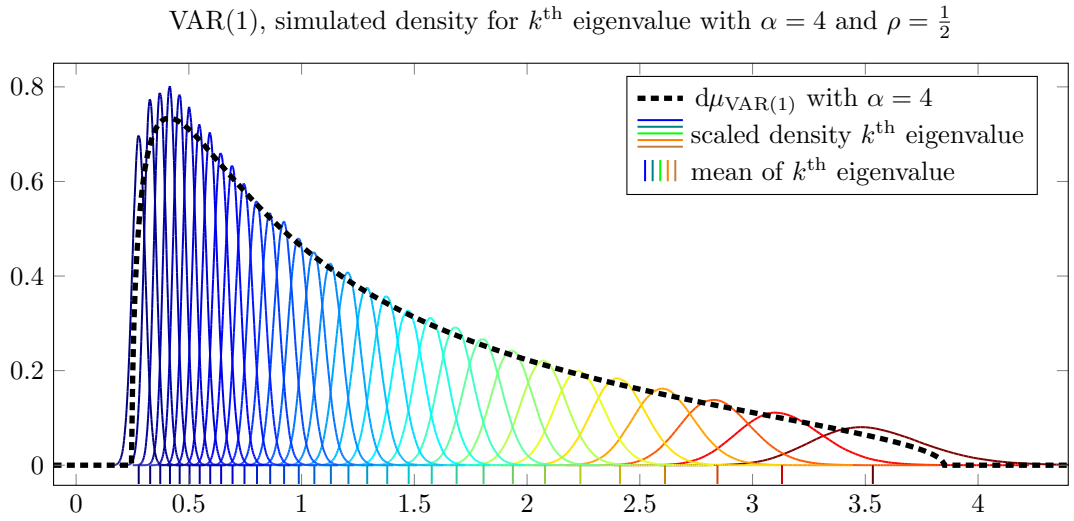

Figure 3.2: Simulation of $n = 2^{20}$ realizations of size 32×32 VAR(1) covariance matrices with rectangularity ratio $\alpha = 4$ and $\rho = \frac{1}{2}$.

The fit of the theoretical results with the simulation outcomes is striking. Numerical integration of the plotted density gives 1 with a precision of at least 6 decimal places.

In comparison with Figure 3.2 we ran the same simulation for uncorrelated Gaussian processes, which gives the Marchenko–Pastur density.

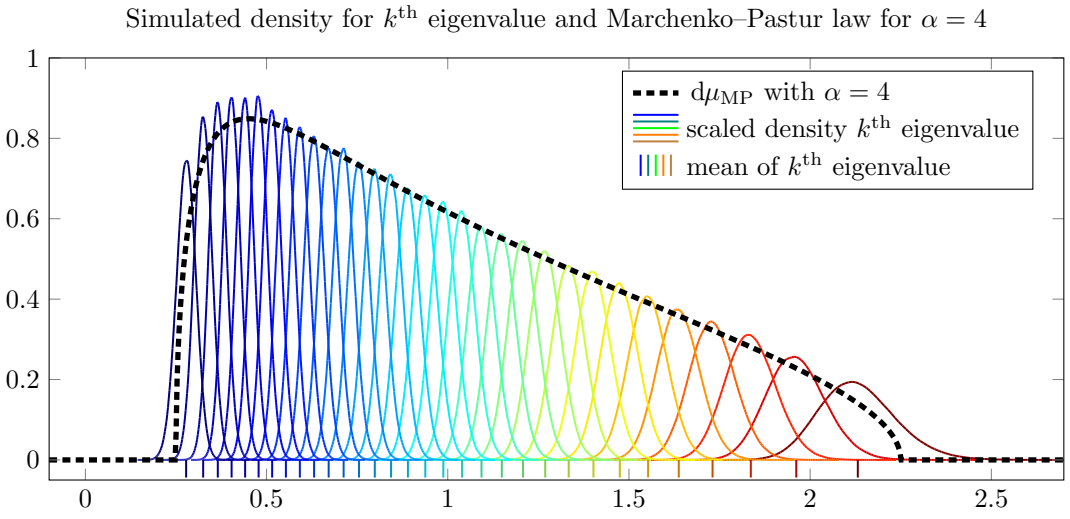

Figure 3.3: Simulation of $n = 2^{20}$ realizations of size 32×32 Wishart matrices with rectangularity ratio $\alpha = 4$.

Comparing Figure 3.3 with Figure 3.2 one notices that by inducing auto-correlation of $\rho = \frac{1}{2}$ the right tail and thus the domain of the probability density function increases. To conserve probability mass at unity the shape of the distribution has to change. It should come at no surprise that introduction of a parametric family, which contains the original Marchenko–Pastur distribution as a special case but possesses an extra parameter, gives rise to more flexibility. But seen the other way, one is able to model more distinct dynamics respectively one is able to differentiate between certain dynamics by comparing different spectral densities.

For VAR(2) we get again a 9^{th} order distribution generating polynomial in M, which has a `LeafCount` of 1 862. Its explicit expression can be found in subsection A.3.2. In Figure 3.4 we compare the spectral density of a VAR(2) process with small parameters $\theta_1 = \theta_2 = 0.1$ with the Marchenko–Pastur law, both with rectangularity ratio $\alpha = 4$. We also plot two additional densities for two different parameter combinations chosen in such a way that the resulting process is stationary.

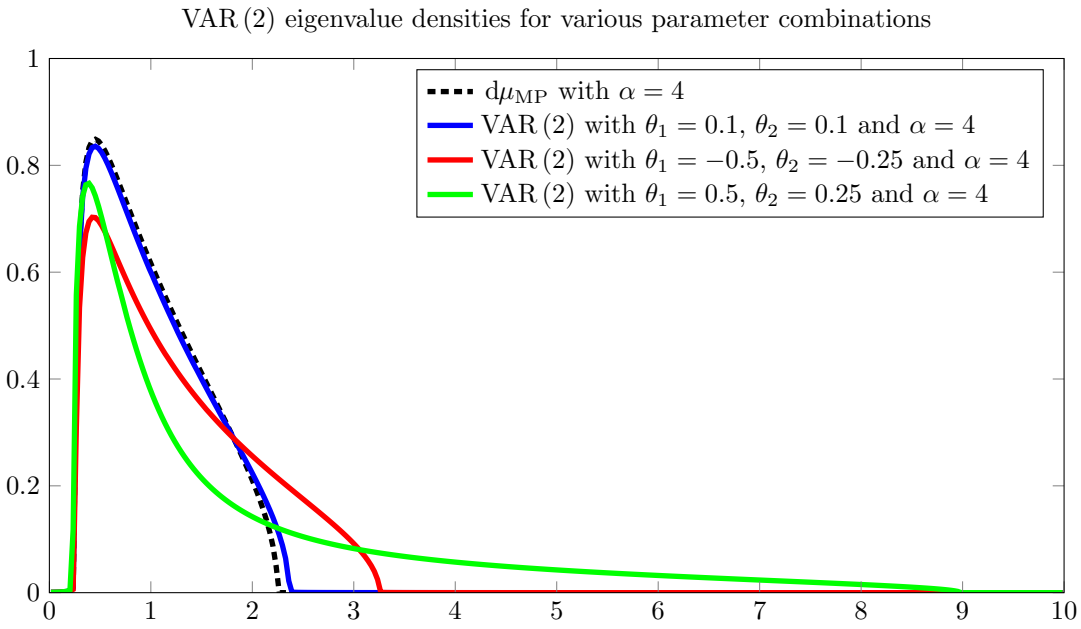

Figure 3.4: Comparison of VAR(2) eigenvalue densities for various parameter combinations with Marchenko–Pastur law, all with $\alpha = 4$.

The shape is different from VMA(2) processes depicted in Figure 3.1. The most prominent difference is that for positive auto-correlations of order 1 and 2 the spectral density exhibits a long right tail. This is to be expected, as two positive auto-correlation coefficients mean that consecutive realizations of the AR(2) process are positively correlated and thus cluster. The VAR(2) process consists of N independent copies of AR(2) processes, where clusters of high resp. low values differ between different trajectories. This naturally spreads the singular values of the observed matrix $\mathbf{X} \sim \text{VAR}(2)$ with $\mathbf{X} \in \mathbb{R}^{T \times N}$, which in turn are implied by the eigenvalues $\mathbf{X}^\intercal\mathbf{X}$.[27] But this, modulo renormalization, is the ML covariance matrix estimator. So, after normalizing the variances to 1, this also holds for the ML correlation matrix estimator.

[27] We disregarded the $T - N$ zero singular values, which would be present in $\mathbf{X}\mathbf{X}^\intercal$ for $T > N$.

We were not able to calculate the distribution generating polynomials for VMA(3) or VAR(3) as Mathematica eventually ran out of memory,[28] but both of these distribution generating polynomials are expected to be quite complicated and long.

VARMA(p, q) processes

We saw above that knowing the explicit form of the spectral density was key to explicitly calculate the $\mathcal{M}$-transform (3.21a). In principal, one always follows the same route: the definite integral (3.21b) has to be solved for a specific spectral density given by (3.18) for ARMA(p, q) processes. This is a rational function, because

$$\begin{aligned}\frac{\sigma^2}{2\pi}\left|\frac{\phi(\mathrm{e}^{-\mathrm{i}\omega})}{\theta(\mathrm{e}^{-\mathrm{i}\omega})}\right|^2 &= \frac{\sigma^2}{2\pi}\frac{\phi(\mathrm{e}^{-\mathrm{i}\omega})\overline{\phi(\mathrm{e}^{-\mathrm{i}\omega})}}{\theta(\mathrm{e}^{-\mathrm{i}\omega})\overline{\theta(\mathrm{e}^{-\mathrm{i}\omega})}}, \qquad \omega \in (-\pi, \pi] \\ &= \frac{\sigma^2}{2\pi}\frac{\phi(\mathrm{e}^{-\mathrm{i}\omega})\phi(\mathrm{e}^{\mathrm{i}\omega})}{\theta(\mathrm{e}^{-\mathrm{i}\omega})\theta(\mathrm{e}^{\mathrm{i}\omega})},\end{aligned} \tag{3.26}$$

where the second equality holds because all coefficients of the $\phi(z)$ and $\theta(z)$ polynomials have real coefficients for which we have $\overline{\phi_k} = \phi_k$ for $k \in \{1, \ldots, q\}$ respectively $\overline{\theta_\ell} = \theta_\ell$ for $\ell \in \{1, \ldots, p\}$. The earlier mentioned »Chebyshev polynomial trick« allows us to express both, the numerator and denominator of the rational function (3.26), as polynomials in $x \coloneqq 2\cos\omega$. So at last one only has to solve a definite integral (3.21b) of a rational function. The integrand is a rational function in the numerator and denominator polynomials in x, but with complex coefficients induced by the complex M.

In principle this kind of integrand is solvable by an application of the residue theorem. For it to apply, we must not have any poles of the integrand on $(0, \pi)$, as the integration path must not go through any of the residues. Then we have to integrate big circular or rectangular paths, whichever is more convenient, which encloses all residues.

For VARMA$(1, 1)$, we were able to calculate integral (3.21b) with the aid of Mathematica and obtained the following 5^{th} order distribution generating polynomial:

$$\begin{aligned}&(M+1)^2\alpha z^2(M+\alpha) \\ &\cdot\big(M\big(\phi_1\big(-2\big(\phi_1^2+1\big)\theta_1^2+\phi_1\big(\phi_1^2-10\big)\theta_1-2\big(\phi_1^2+1\big)\big)+\theta_1\big) \\ &\quad -\alpha\big(\phi_1^2+2\phi_1\theta_1+1\big)\big(\phi_1(\phi_1\theta_1+2)+\theta_1\big)\big) \\ &-M(M+1)\alpha^2 z\Big(M\big(2\phi_1^2\big(\theta_1^3+\theta_1\big)-\phi_1\big(\theta_1^4-10\theta_1^2+1\big)+2\big(\theta_1^3+\theta_1\big)\big) \\ &\quad -2\phi_1\big(\theta_1^2-1\big)^2\alpha\Big)+\phi_1\big(\phi_1^2-1\big)^2(M+1)^3z^3(M+\alpha)^2+\theta_1\big(\theta_1^2-1\big)^2M^2\alpha^3.\end{aligned}$$

In this compact form, it has a `LeafCount` of 160, but multiplied out so every monomial of M has its own coefficient the `LeafCount` is 467.

[28] On a 16 GB RAM Workstation, ca. 14 GB RAM usable.

Next, we compare the VARMA$(1,1)$ law we calculated with the Marchenko–Pastur law. For Gaussian VARMA$(1,1)$ the Marchenko–Pastur law $\mathrm{d}\mu_{\mathrm{MP}}$ is equal to a situation with $\phi_1 = \theta_1 = 0$. In Figure 3.5 we plotted the Marchenko–Pastur law $\mathrm{d}\mu_{\mathrm{MP}}$ versus the eigenvalue densities of VARMA$(1,1)$ processes with three different parameter pairs and a rectangularity ratio for $\alpha = 4$.

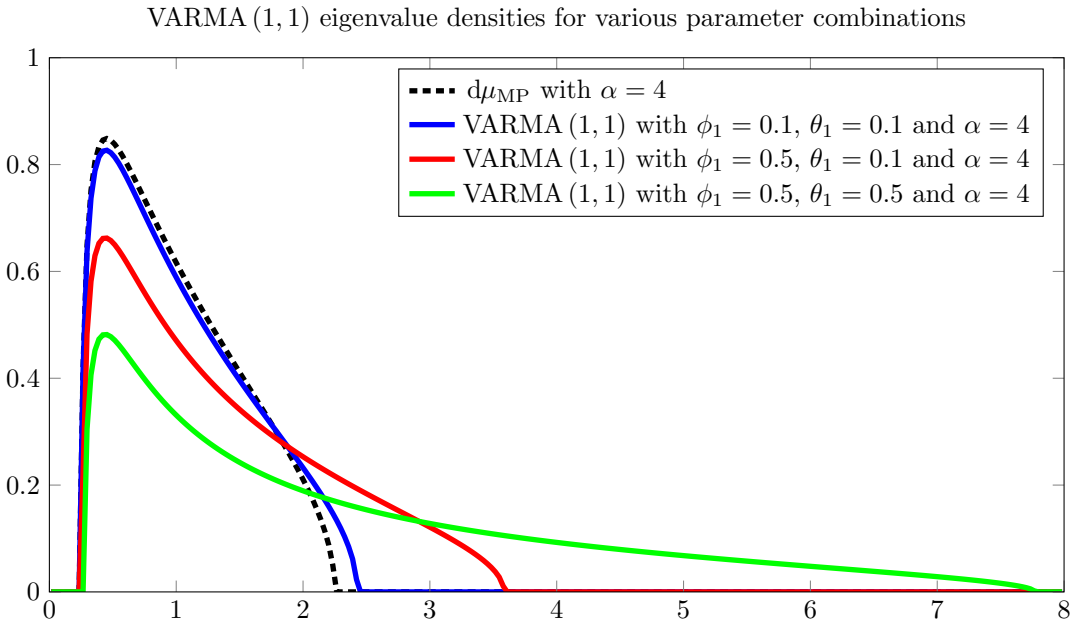

Figure 3.5: Comparison of VARMA$(1,1)$ eigenvalue densities for various parameter combinations with Marchenko–Pastur law, all with $\alpha = 4$.

For $\phi_1 = \theta_1 = 0.1$ the eigenvalue density deviates only slightly from the Marchenko–Pastur law. The shape changes for more distinct parameter values.

3.1.6 General Remarks on VARMA(p,q) Eigenvalue Density Distributions

One also speaks of *rational* spectra when describing ARMA(p,q) processes, because (3.26) is a rational function. The other way round, for appropriate polynomials in numerator and denominator we can find an ARMA(p,q) process. By making sure that the numerator and denominator are relative prime to each other and, if necessary, by exchanging roots with modulus < 1 with their inverses, one can always construct a stationary ARMA$(\tilde{p},\tilde{q})$ process w.r.t. some white noise process.

Having the previous paragraph in mind, it is not surprising that in time series theory of spectral densities every real-valued stationary process (X_t) with continuous spectral density can be approximated with either an invertible MA(q) or a causal AR(p) process. The approximation is uniform for the respective spectral densities.[29] If one can establish either of both, for instance the approximation for arbitrary MA(p)

[29] See [20, Corollary 4.4.1 and 4.4.2].

processes, by Wold's decomposition one gets an appropriate approximation for AR(q) processes, too. From this perspective a VARMA$(1,1)$ process combines both, the direct MA(1) component and the indirect MA(p) components, with possibly $p = \infty$, from the Wold decomposition of the AR(1) part. Therefore, in general VARMA$(1,1)$ processes are quite parsimonious as their spectral density can assume shapes that MA(p) or AR(p) processes could not for low number of parameters p. Because of this rationale, having been able to calculate the explicit distribution generating polynomial for VARMA$(1,1)$ processes is likely of much more practical relevance than being able to calculate the explicit distribution generating polynomials for VMA(p) resp. VAR(p) processes for moderate $p \geq 3$.

In the special case of a linear stochastic process with dependence structure given by the Gaussian MA processes in [40] Hasegawa et al. give a nice interpretation of the sample covariance matrix $\hat{\mathbf{C}}$ as a random matrix model for the compound free Poisson law. In (2.3) we already defined the free Poisson law and remarked that this is a synonym in free probability for the Marchenko–Pastur law in random matrix theory. The compound generalization of the free Poisson law is given by

$$\lim_{T\to\infty}\left(\left(1-\frac{\alpha^{-1}}{T}\delta_0+\frac{\alpha^{-1}}{T}\rho\right)\right)^{\boxplus T}, \tag{3.27}$$

where ρ is a probability measure over $\mathbb{R}$. For $\rho = \delta_1$ one retrieves the special non-compound version, i.e. the Marchenko–Pastur law. So VMA processes give rise to a sample covariance matrix, which is ex ante a random matrix with spectral distribution (3.27), where ρ is the spectral measure of the auto-covariance matrix of the VMA process. If the auto-covariance matrix is trivial, that is the identity matrix $\mathbf{I}$, all its eigenvalues are 1 and therefore its spectral measure is δ_1.

3.2 Spectral Distributions of Some Linear Stochastic Processes

In the previous section 3.1 we deduced first a connection of the spectral density $\tilde{\gamma}_{(X_t)}(\tau)$ of a univariate stochastic process (X_t) and the spectral measure of the sample covariance matrix estimator of a matrix of N independent copies of (X_t). We present a direct method to calculate the $\mathcal{G}$-transform of the sample covariance matrix estimator from a given spectral density and a noise matrix $\mathbf{Z}$, subject to some technical restrictions.

Remark: *This method was introduced 2011 by Pfaffel and Schlemm in [81]. In this section we primarily follow [81], restate their main theorem and replicate some of their know results as starting point for our own extensions. We specifically mention new results not previously known.*

Theorem 3.1 (Cauchy Transform of Linear Stochastic Process[30])
For each $i = 1, \dots, N$, let $X_{t,i} = \sum_{j=0}^{\infty} c_j Z_{t-j,i}$, $t \in \mathbb{Z}$, be a linear stochastic process that satisfies

$$\mathbb{E}[Z_{t,i}] = 0, \quad \mathbb{E}\big[Z_{t,i}^2\big] = 1, \quad \text{and} \quad \sup_{t,i} \mathbb{E}\big[Z_{t,i}^4\big] < \infty$$

$$\lim_{n\to\infty} \frac{1}{TN} \sum_{t=1}^{T} \sum_{i=1}^{N} \mathbb{E}\Big[Z_{t,i}^2 \mathbb{1}_{Z_{t,i}^2 \geq \varepsilon T}\Big] = 0,^{31} \tag{3.28}$$

and has a continuously differentiable spectral density $\tilde{\gamma}$. Assume that

i) there exist positive constants C and δ such that $|c_j| \leq C\,(j+1)^{-1-\delta}$ $(\forall j \geq 0)$,

ii) for almost all $\lambda \in \mathbb{R}$, $f(\omega) = \lambda$ for at most finitely many $\omega \in [0, 2\pi]$, and

iii) $\frac{\mathrm{d}}{\mathrm{d}\omega}\tilde{\gamma}(\omega) \neq 0$ for almost every ω.

Then the empirical spectral distribution $F^{T^{-1}\mathbf{X}^\intercal\mathbf{X}}$ of $T^{-1}\mathbf{X}^\intercal\mathbf{X}$ converges, as T tends to infinity, almost surely to a non-random probability distribution $\hat{F}$ with bounded support. Moreover, there exist positive numbers λ_-, λ_+ such that the $\mathcal{G}$-transform $z \mapsto \mathcal{G}_{\hat{F}}(z)$ of $\hat{F}$ is the unique mapping $\mathbb{C}_+ \to \mathbb{C}_+$ satisfying

$$\frac{\alpha}{\mathcal{G}_{\hat{F}}(z)} = -\alpha z + \frac{\alpha}{2\pi} \int_{\lambda_-}^{\lambda_+} \frac{\lambda}{1 + \lambda\alpha^{-1}\mathcal{G}_{\hat{F}}(z)} \sum_{\omega\in[0,2\pi]:\tilde{\gamma}(\omega)=\lambda} \frac{1}{\left|\frac{\mathrm{d}}{\mathrm{d}\omega}\tilde{\gamma}(\omega)\right|}\,\mathrm{d}\lambda. \tag{3.29}$$

Remark: *We use the convention that sample matrices have its observations in time in rows and not in columns as the authors in [81]. We therefore have to rescale their variables according to $m \mapsto \alpha^{-1}\mathcal{G}_{\hat{F}}(z)$, $z \mapsto \alpha z$ and $y \mapsto \alpha$ to match ours. We already rescaled their original version [81, Theorem 1.1.] to our notation in Theorem 3.1.*

Condition (3.28) is a Lindeberg-type condition that ensures that the contribution of every single random variable to the variance of their sum is arbitrary small for sufficiently large T.

3.2.1 VARMA$(1,1)$ processes

To utilize Theorem 3.1 for a linear stochastic process one needs to differentiate the function

$$g\colon [\lambda_-, \lambda_+] \to \mathbb{R}_+, \qquad \lambda \mapsto \frac{1}{2\pi} \sum_{\omega\in[0,2\pi]:\tilde{\gamma}(\omega)=\lambda} \frac{1}{\left|\frac{\mathrm{d}}{\mathrm{d}\omega}\tilde{\gamma}(\omega)\right|}\,\mathrm{d}\lambda. \tag{3.30}$$

[30] See [81, Theorem 1.1.].
[31] $T \in \mathbb{N}$

Remark: *Note that we took the normalizing constant of* $(2\pi)^{-1}$ *inside the integral as part of function* g *to ease subsequent notation of integrals.*

From (3.26) we know that the spectral density of an ARMA$(1,1)$ process is given by

$$\tilde{\gamma}_{\mathrm{ARMA}(1,1)}(\omega) = \frac{1}{2\pi}\frac{1+\phi_1^2+2\phi_1\cos(\omega)}{1+\theta_1^2-2\theta_1\cos(\omega)}, \quad \omega \in [0,\pi]. \tag{3.31}$$

Since this is the spectral density of a real process, because of symmetry it suffices to analyze (3.31) only on the domain $[0,\pi]$. Because the sum in (3.30) is over all $\tilde{\gamma}(\omega) = \lambda$ we need to invert $\tilde{\gamma}_{\mathrm{ARMA}(1,1)}(\omega)$ and arrive at

$$\tilde{\gamma}^{(-1)}_{\mathrm{ARMA}(1,1)}(\lambda) \coloneqq \arccos\left(\frac{(\theta_1^2+1)\lambda-\phi_1^2-1}{2(\phi_1+\theta_1\lambda)}\right).$$

The principal branch of arccos is monotonically decreasing from $\arccos(-1)=\pi$ to $\arccos(1)=0$, so the domain of $\tilde{\gamma}^{(-1)}_{\mathrm{ARMA}(1,1)}(\lambda)$ can be found by calculating the argmin resp. argmax of $\tilde{\gamma}^{(-1)}_{\mathrm{ARMA}(1,1)}(\lambda)$, that is solving for the argument of arccos for $\{1,-1\}$. Doing so, we get

$$\begin{aligned} 1 &= \frac{(\theta_1^2+1)\lambda-\phi_1^2-1}{2(\phi_1+\theta_1\lambda)} \quad &\Longleftrightarrow \quad \lambda_1 &= \frac{(\phi_1-1)^2}{(\theta_1+1)^2} \\ -1 &= \frac{(\theta_1^2+1)\lambda-\phi_1^2-1}{2(\phi_1+\theta_1\lambda)} \quad &\Longleftrightarrow \quad \lambda_2 &= \frac{(\phi_1+1)^2}{(\theta_1-1)^2}. \end{aligned}$$

We can now set $\lambda_- \coloneqq \min\{\lambda_1,\lambda_2\}$ and $\lambda_+ \coloneqq \max\{\lambda_1,\lambda_2\}$, which gives the range of $\tilde{\gamma}_{\mathrm{ARMA}(1,1)}(\omega)$ resp. the domain of $\tilde{\gamma}^{(-1)}_{\mathrm{ARMA}(1,1)}(\lambda)$ with $[\lambda_-,\lambda_+]$.

It only remains to calculate

$$\frac{1}{\left|\frac{\mathrm{d}}{\mathrm{d}\omega}\tilde{\gamma}(\omega)\right|} = \left|-\frac{(1+\theta_1^2-2\theta_1\cos(\omega))^2}{2(\phi_1+\theta_1)(\phi_1\theta_1+1)\sin(\omega)}\right|,$$

to substitute $\omega \coloneqq \tilde{\gamma}^{(-1)}_{\mathrm{ARMA}(1,1)}(\lambda)$ and ensure that the resulting expression is non-negative due to the absolute value requirement. Doing so and utilizing the trigonometric identity $\frac{\cos(\arccos(x))}{\sin(\arccos(x))} = \frac{x}{\sqrt{1-x^2}}$, one arrives at

$$\begin{aligned} g(\lambda) &\coloneqq \frac{\left(1+\theta_1^2-2\theta_1\cos\left(\arccos\left(\frac{(\theta_1^2+1)\lambda-\phi_1^2-1}{2(\phi_1+\theta_1\lambda)}\right)\right)\right)^2}{2(\phi_1+\theta_1)(\phi_1\theta_1+1)\sin\left(\arccos\left(\frac{(\theta_1^2+1)\lambda-\phi_1^2-1}{2(\phi_1+\theta_1\lambda)}\right)\right)}\mathbb{1}_{[\lambda_-,\lambda_+]} \\ &= \frac{(\phi_1+\theta_1)(\phi_1\theta_1+1)}{\pi(\phi_1+\theta_1\lambda)\sqrt{\left((\phi_1+1)^2-(1-\theta_1)^2\lambda\right)\left((1+\theta_1)^2\lambda-(1-\phi_1)^2\right)}}. \end{aligned} \tag{3.32}$$

Due to the definition of g on the domain $[\lambda_-,\lambda_+]$ in (3.30), the indicator function is not needed and was therefore omitted in (3.32).

Remark: *Pfaffel and Schlemm give function $g(\lambda)$ in [81, p. 10, bottom], but note that they forgot the factor $(\phi_1+\theta_1)(\phi_1\theta_1+1)$ in the numerator. But for subsequent calculations one only arrives at their results with (3.32) and not with their version. So this missing factor seems to be only some minor inadvertence on their side.*

We now have all the prerequisites to calculate integral (3.29). After tedious calculations one can bring the integrand

$$\frac{\lambda}{1+\lambda\alpha^{-1}\mathcal{G}_{\hat{F}}(z)}g(\lambda)$$

in the following easy to integrate form

$$\int \frac{1}{a+b\lambda^2}\,\mathrm{d}\lambda = \frac{\arctan\left(\sqrt{\frac{b}{a}}\lambda\right)}{\sqrt{ab}},$$

with lengthy expressions for a, b and quite complicated integration limits. The arctan expression can be written in a difference of log's along the imaginary axis. Rearranging, collecting terms and simplifying, one finally arrives at the concrete form of (3.29) for a VARMA$(1,1)$ process, given by

$$\begin{aligned} 0 = {} & \frac{\alpha z}{1-\alpha(1+\mathcal{G}_{\hat{F}}(z)z)} - \alpha z + \frac{\phi_1}{\phi_1\mathcal{G}_{\hat{F}}(z)-\theta_1} \\ & + \frac{(\phi_1+\theta_1)(\phi_1\theta_1+1)}{(\theta_1-\phi_1\mathcal{G}_{\hat{F}}(z))\sqrt{\left((\phi_1+1)^2\mathcal{G}_{\hat{F}}(z)+(\theta_1-1)^2\right)\left((\phi_1-1)^2\mathcal{G}_{\hat{F}}(z)+(\theta_1+1)^2\right)}}. \end{aligned} \tag{3.33}$$

This is, modulo a different scaling, exactly[32] the solution of [81, eq. (3.2)]. Solving (3.33) gives a quintic polynomial in $\mathcal{G}_{\hat{F}}(z)$ and one of its roots is the $\mathcal{G}_{\hat{F}}(z)$ we searched for.

We calculated the explicit distribution generating polynomial for the $\mathcal{G}$-transform, which is the following quintic polynomial in $\mathcal{G} = \mathcal{G}_{\mathrm{VARMA}(1,1)}(z)$:

$$\begin{aligned} & \phi_1^5\mathcal{G}^3(\mathcal{G}z-r+1)^2 \\ & + \phi_1^4\theta_1\mathcal{G}^2 r(\mathcal{G}z+r+1)(-\mathcal{G}z+r-1) - 2\phi_1^3\mathcal{G}^2\left(\mathcal{G}-\left(\theta_1^2+1\right)r\right)(\mathcal{G}z-r+1)^2 \\ & + 2\phi_1^2\theta_1\mathcal{G}r\Big(\mathcal{G}\left(r\left(-z\left(\theta_1^2(\mathcal{G}z+2)+\mathcal{G}(z+8)+2\right)-8\right)+5(\mathcal{G}z+1)^2+3r^2\right) \\ & \quad -\left(\theta_1^2+1\right)r\Big) \\ & + \phi_1\mathcal{G}\Big(2\mathcal{G}^3 z\left(r\left(\theta_1^2 z+z-1\right)+1\right)+\mathcal{G}^2 \\ & \quad \cdot\left(r\left(-4\left(\theta_1^2+1\right)(r-1)z+\left(\theta_1^4-10\theta_1^2+1\right)rz^2+r-2\right)+1\right) \\ & \quad + 2\mathcal{G}r\left(\left(\theta_1^2+1\right)(r-1)^2+rz\left(\theta_1^4-\left(\theta_1^2-1\right)^2 r-10\theta_1^2+1\right)\right) \\ & \quad + r^2\left(\theta_1^4-2\left(\theta_1^2-1\right)^2 r-10\theta_1^2+1\right)+\mathcal{G}^4 z^2\Big) \\ & + r\Big(\theta_1^5\left(-(\mathcal{G}rz+r)^2\right)-2\theta_1^3 r(\mathcal{G}-r)(\mathcal{G}z+1)^2 \\ & \quad -\theta_1(\mathcal{G}z(\mathcal{G}+r)+\mathcal{G}(-r)+\mathcal{G}+r)(\mathcal{G}z(\mathcal{G}+r)+\mathcal{G}r+\mathcal{G}+r)\Big). \end{aligned} \tag{3.34}$$

[32] Note that we exchanged minuend and subtrahend for the term in the denominator, explaining changed signs.

Remark: *Pfaffel and Schlemm write about the solution of* (3.33) *in [81, eq. (3.2)] the following: »This is a quartic equation in* $m_z \equiv m(z)$*, which can be solved explicitly.« ([81, p. 11])*[33] *Clearly, our distribution generating polynomial* (3.34) *is quintic and not quartic. Our rescaling should not change the order of the polynomial and does not. Carrying out their calculations one also gets to a quintic polynomial. Pfaffel and Schlemm probably mixed up the polynomial order with the one they got after setting* $\phi_1 = 1$ *and* $\theta_1 = 1/2$*. For this special case the polynomial is, in fact, a quartic and therefore can be solved explicitly in closed form.*

For VMA(1) resp. VAR(1) we get, again, quartic polynomials. They can be deduced by setting either $\theta_1 = 0$ or $\phi_1 = 0$ in (3.34). The spectral densities, generated by the inverse Cauchy-transform theorem, are indistinguishable to the ones generated by the method used in section 3.1.

As we now understand how to utilize this method, we want to apply it to new linear stochastic processes. Up to now, we only investigated VARMA(p, q) processes. One common characteristics of the ARMA(p, q) process family is, that its autocorrelation function decreases exponentially for growing lags τ, that is we have $\left|\gamma_{\mathrm{ARMA}(p,q)}(\tau)\right| \leq Cr^{\tau}$ for some constants $C > 0$ and $0 < r < 1$.[34] Stochastic processes with exponential decaying auto-correlation function are called »short-range dependent« resp. exhibiting »short memory«. On the other hand, a process whose autocorrelation function decays slower than exponentially is said to exhibit »long-range dependence« resp. having »long memory«. We will introduce a prominent class of linear stochastic processes with *long memory* in the next sections.

3.2.2 Short Introduction to *Long-Range Dependence*

Box and Jenkins introduced and popularized in [19] the class of so-called *integrated* ARMA resp. ARIMA(p, d, q) processes with integration order $d \in \mathbb{N}_0$. The linear stochastic process (Y_t) is ARIMA(p, d, q) if the process $(1 - \mathrm{B})^d Y_t = X_t$ and (X_t) is ARMA(p, q). Granger and Joyeux in 1980 and Hosking in 1981 proposed to allow for non integer-valued d and realized that the resulting autocorrelation function exhibits *long memory*. The fact that d is non-integer gave these kind of processes the name »fractional integrated« ARMA processes of (fractional integration) order $d \in (-2^{-1}, 2^{-1})$.

Taqqu discusses in [101, section 4] different definitions of *long-range dependence*. We will only use this term in a broad sense meaning that the auto-correlation function decays slower than exponentially.

In econometrics, [34] and [42] are among the first to introduce and apply fractional integration methods. Henry and Zaffaroni discuss in [41] long range dependence in macroeconomics and finance.

[33] Their $m(z)$ is, in our notation, the $\mathcal{G}$-transform $\mathcal{G}(z)$.
[34] See [20, section 3.3 and p. 520], see also [38, pp. 448–449] and [43, p. 169].

In [89, section 10], Sewell gives empirical evidence and list research that implies that stock return time series sometimes exhibit long range dependence. Baillie et al. is one of the first in 1996 to apply fractional integrated autoregressive–moving-average (ARFIMA) methods to economics, though he analyzed inflation and also used a GARCH model. Willinger et al. find some, slightly inconclusive, evidence for long range dependence in daily stock return time series.

Lillo and Farmer research consequences of empirical identifiable long dependence patterns in order processes, but argue that the market remains efficient due to also long memory in »*anti-correlated fluctuations in transaction size and liquidity*« ([58, abstract]).

Already in 1980 Granger argued in [33] that long memory in economic time series could be induced by aggregation of time series with different levels of persistence. This is essentially the well-known argument of [6] for volatility clustering. Suárez-García and Gómez-Ullate estimate long memory being present in high-frequency return series of Madrid's Stock Exchange Ibex35 index and also give evidence for the above made claim by Granger that this long memory might be caused by a superposition of a high-frequency component and a slow-varying one.

On the other hand, Mikosch and Stărică argue in [69] and [70] that non-stationary of return time series induces the long range dependence type behavior of its empirical auto-correlation function. Overall, Graves et al. give an extensive history of long range dependence, beginning with Hurst[35] and Mandelbrot[36] to fractional integrated models of Hosking.

3.2.3 ARFIMA(p, d, q) processes

We begin by defining what we understand by *fractional integration* in terms of the usual backshift operator B.

Definition (Fractional Difference Operator)
For the backshift operator B *defined in* (3.1)*, for* $d > -1$*, define the* fractional difference operator *by*[37]

$$(1 - \mathrm{B})^d := \sum_{j=0}^{\infty} \binom{d}{j} (-\mathrm{B})^j = \sum_{j=0}^{\infty} \prod_{k=1}^{j} \frac{k - 1 - d}{k} \mathrm{B}^j.$$

Having defined what is meant by *fractional* integration we can now define *fractional integrated* ARMA processes.

[35] Hurst invented the *Rescaled Adjusted Range Statistic* in a series of papers [46, 47, 48, 49].
[36] See [61, 62].
[37] See [43, eq. (2.1)] for the definition and [20, eq. (13.2.2)] for the second equality.

Definition (ARFIMA(p, d, q) processes)
The stochastic process (X_t) *is called an* ARFIMA(p, d, q) *process with* $d \in (-2^{-1}, 2^{-1})$, *if there exists a white noise process* $(Z_t) \sim \mathcal{WN}(\sigma^2)$ *such that*

$$\theta(\mathrm{B})(1-\mathrm{B})^d X_t(\omega) = \phi(\mathrm{B}) Z_t(\omega),$$

holds almost surely and $\theta(\mathrm{B})$ *resp.* $\phi(\mathrm{B})$ *are* AR *resp.* MA *polynomials of order* p *resp.* q.

The auto-correlation function of an ARFIMA(p, d, q) process is given by[38]

$$\gamma_{\mathrm{ARFIMA}(p,d,q)}(\tau) = \prod_{k=1}^{\tau} \frac{k-1+d}{k-d}. \tag{3.35}$$

Noting that a product of the fraction in (3.35) can be written as a product of factorials resp. the Gamma function Γ, an application of Sterling's formula gives the asymptotic behavior of the auto-correlation function as

$$\gamma_{\mathrm{ARFIMA}(p,d,q)}(\tau) \sim \frac{\Gamma(1-d)}{\Gamma(d)} \tau^{2d-1}, \quad \text{for } \tau \to \infty, \tag{3.36}$$

where the Gamma function for negative arguments $x \in [-1, 0)$ is defined to be $\Gamma(x) \coloneqq x^{-1}\Gamma(1+x)$. Clearly, the auto-correlation of (3.36) decays polynomially and such much slower than the exponential decay of ARMA(p, q) auto-correlation functions.

Hosking proves in [43, Theorem 1] that for $d < 1/2$ an ARFIMA$(0, d, 0)$ process (X_t) is stationary and has a MA(∞) representation and for $d > -1/2$ it is invertible and has an AR(∞) representation. The spectral density of an ARFIMA$(0, d, 0)$ process is given by

$$\tilde{\gamma}_{\mathrm{ARFIMA}(0,d,0)}(\omega) = \left(2\sin\left(\frac{1}{2}\omega\right)\right)^{-2d} = (2-2\cos(\omega))^{-d}, \quad \omega \in [0, \pi], \tag{3.37}$$

where the last equality follows from the *half angular formula* $2\sin^2\left(\frac{\omega}{2}\right) = 1 - \cos(\omega)$.

In general, the spectral density of an ARFIMA(p, d, q) process can be calculated to obey

$$\tilde{\gamma}_{\mathrm{ARFIMA}(p,d,q)}(\omega) = \frac{\sigma^2}{2\pi} \left|\frac{\phi(\mathrm{e}^{-\mathrm{i}\omega})}{\theta(\mathrm{e}^{-\mathrm{i}\omega})}\right|^2 \left|1-\mathrm{e}^{-\mathrm{i}\omega}\right|^2. \tag{3.38}$$

Comparing (3.38) with the spectral density of an ARMA(p, q) process (3.18) one notices that the fractional integration part is given by multiplying the spectral density with

$$\left|1-\mathrm{e}^{-\mathrm{i}\omega}\right|^{-2d} = \left(\sqrt{1-\cos(\omega)^2+\sin(\omega)^2}\right)^{-2d} = (2-2\cos(\omega))^{-d}, \tag{3.39}$$

[38] See [20, eq. (13.2.9)].

where the second equality follows from the Pythagorean trigonometric identity and subsequent application of de Moivre's formula.

In general, if (Y_t) is a stochastic process with an absolutely summable autocorrelation function $\gamma_{(Y_t)}(\tau)$, it therefore has a continuous spectral density $\tilde{\gamma}_{(Y_t)}(\omega)$. Let (X_t) be the stochastic process obtained by fractionally integrating process (Y_t), then the spectral density of (X_t) is given by[39]

$$\tilde{\gamma}_{(X_t)}(\omega) = \left|1 - \mathrm{e}^{-\mathrm{i}\omega}\right|^{-2d} \tilde{\gamma}_{(Y_t)}(\omega), \quad \omega \in [0, 2\pi).$$

In subsection 3.1.6 we argued that ARMA(p, q) processes can approximate any continuous spectral density, so why does one need *fractional integrated* ARMA(p, q) processes? The answer is twofold.

If a stochastic process is integrated, fractionally or integer-valued, it has infinite variance and one has to difference in order to transform the process to one with finite variance. If the stochastic process (Y_t) was fractionally integrated with, say, $d = 1/3$, differencing of order 1 would give a stochastic process (X_t) with integration order $-2/3$. This is completely legitimate as we defined fractional integration and, in particular, fractional integrated ARFIMA(p, q) processes. So for the sake of argument take $(Y_t) \sim$ ARFIMA(p, d, q). But according to [43, Theorem 2, (1) & (2)], an ARFIMA(p, d, q) (Y_t) process is

1. *stationary*, if $d < 1/2$ and $\theta(z)$ has only roots outside the unit circle,

2. *invertible*, if $d > -1/2$ and $\phi(z)$ has only roots outside the unit circle.

So the fractional integration order is $d - 1 = -2/3 < -1/2$ for the stochastic process (X_t) and it thus cannot possess an invertible AR(∞) representation. In addition, by differencing »too much« the zero frequency component will be removed.[40]

Second, as it is possible to approximate a given spectral density with an ARMA(p, q) process whose spectral density is close in norm to the first, the polynomial orders p, q might be very high. This induces numerous problems ranging from parameter identification to problems when predicting. Following the epistemic principle of Occam's razor a more parsimonious model should be used that captures the essential characteristics, i.e. the long range dependence. This justifies the class of ARFIMA(p, d, q) models.

A closed form expression for the auto-covariance function and the spectral density for an ARFIMA(p, d, q) process with only Gaussian white noise in terms of hypergeometric functions is given in [95].

[39] See [87, eq. (6.13)].
[40] See [35, p. 16].

3.2.4 VARFIMA$(0, d, 0)$ processes

Let $(X_t) \sim \text{ARFIMA}(p, d, q)$ be a stochastic process with $d \in (-2^{-1}, 2^{-1})$. Construct the sample matrix $\mathbf{X} \in \mathbb{R}^{T \times N}$ as the collection of N independent samples of (X_t), all with length T. So $X_{t,i}$ is the t^{th} realization of the i^{th} copy of process (X_t). As before, we want to examine spectral properties of the sample covariance matrix $\hat{\mathbf{C}} \coloneqq T^{-1}\mathbf{X}^\intercal\mathbf{X} \in \mathbb{R}^{N \times N}$. By Theorem 3.1 this amounts to calculating the integral in (3.29) and subsequently solve the equations for $\mathcal{G}_{\hat{F}}(z)$ with $\hat{F}$ the spectral measure of $\hat{\mathbf{C}}$.

By (3.38) the spectral density of an ARFIMA$(0, d, 0)$ process is given by (3.37), for $d \in (-2^{-1}, 2^{-1})$. Following the same steps as in the derivation of the spectral measure for VARMA$(1, 1)$ processes in subsection 3.2.1, the inverse spectral density is then given by the principal branch of

$$\tilde{\gamma}^{(-1)}_{\text{ARFIMA}(0,d,0)}(\lambda) = \arccos\left(\frac{2 - \lambda^{-\frac{1}{d}}}{2}\right), \quad \text{for } \lambda > 0 \text{ and } d \neq 0. \tag{3.40}$$

(3.40) is a strictly monotone increasing function for $d > 0$ resp. strictly monotone decreasing for $d < 0$. The range of (3.37) is in the interval $[\lambda_-, \lambda_+]$ with boundaries $\left\{0^{-d}, 4^{-d}\right\}$, where $4^{-d} \in (2^{-1}, 2)$ for $d \in (-2^{-1}, 2^{-1})$ and we set $0^{-d} = \infty$ for $d > 0$.

Next, we have to calculate

$$\frac{1}{\left|\frac{\mathrm{d}}{\mathrm{d}\omega}\tilde{\gamma}(\omega)\right|} = \frac{(2 - 2\cos(\omega))^d(\cos(\omega) - 1)}{\sin(\omega)d},$$

to substitute $\omega \coloneqq \tilde{\gamma}^{(-1)}_{\text{ARFIMA}(0,d,01)}(\lambda)$ and ensure non-negativity by introducing an appropriate sign factor. Doing so and simplifying further, one arrives at

$$g(\lambda) \coloneqq \operatorname{sign}(d)\frac{\lambda^{-\frac{d+1}{d}}}{\pi d\sqrt{\lambda^{-\frac{2}{d}}\left(4\lambda^{\frac{1}{d}} - 1\right)}}. \tag{3.41}$$

Remark: *Pfaffel and Schlemm only mention* ARFIMA *processes at the end of [81], but they never tried to obtain or approximate its density. We are also not aware that anybody else tried or succeeded. To the best of our knowledge, we are the first to show the shape of resulting LSDs.*

The integral in (3.29)

$$\int_{\lambda_-}^{\lambda_+} \frac{\lambda}{1 + \lambda\alpha^{-1}\mathcal{G}_{\hat{F}}(z)} g(\lambda)\,\mathrm{d}\lambda, \tag{3.42}$$

is approximated numerically, as we were not able to get closed form solutions for all $d \in (-2^{-1}, 2^{-1})$. For specific rational $d \in \pm\{1/4, 1/3\}$, with the aid of Mathematica we got lengthy expressions involving elliptic integrals and their inverses, elliptic functions, and trigonometric and hyperbolic functions. For instance, the integrals

of the two positive d involved the *incomplete elliptic integral of the third kind* Π and arcsinh, whereas the integrals for the two negative d involved *incomplete elliptic integral of the first kind* F and the arcsin.

Approximation of the complex-valued integral of the complex integrand is challenging but feasible. After all, we want to solve (3.29) for $\mathcal{G}_{\hat{F}}(x + \mathrm{i}\varepsilon)$ and trace out the spectral density $-\pi^{-1} \operatorname{Im} \mathcal{G}_{\hat{F}}(x + \mathrm{i}\varepsilon)$ along $x \in [0, \infty)$. In practice, $x \in [0, x_{\max}]$ suffices for some large enough value of $x_{\max}$ that makes this interval enclose the domain of the spectral measure. We numerically approximate integral (3.42) and search for the argument $\mathcal{G}_{\hat{F}}(x + \mathrm{i}\varepsilon) = \operatorname{Re} \mathcal{G} + \operatorname{Im} \mathcal{G}$ that minimizes the expression

$$\left\| -\frac{\alpha}{\mathcal{G}_{\hat{F}}(z)} - \alpha z + \frac{\alpha}{2\pi} \int_{\lambda_-}^{\lambda_+} \frac{\lambda}{1 + \lambda \alpha^{-1} \mathcal{G}_{\hat{F}}(z)} \sum_{\omega \in [0, 2\pi] : \tilde{\gamma}(\omega) = \lambda} \frac{1}{\left| \frac{\mathrm{d}}{\mathrm{d}\omega} \tilde{\gamma}(\omega) \right|} \, \mathrm{d}\lambda \right\|_2^2 . \tag{3.43}$$

For values of (3.43) close to 0 the equality of (3.29) holds approximately and we can solve, with the inverse Cauchy-transform theorem, for $-\pi^{-1} \operatorname{Im} \mathcal{G}$ to get the value of the spectral measure at x.

Clearly, this method is computationally expansive, as for the search for an argmin of (3.43) we have to approximate integral (3.42) for every search step. Secondly, one has to repeat this procedure for every grid point x_k in $[x_{\min}, x_{\max}]$ to approximate the spectral measure. Utilizing already calculated nearby solutions as starting points for the minimization algorithm reduces its time and increases its stability. Nevertheless, if at one step the minimization fails to find arguments that are in a small neighborhood of the true argmin, there might occur an artifact in the estimated density. Of course one could smooth those out by applying, for instance, a median filter for the density values. But this smoothens the very sharp ascent typically seen for the left-sided tail of the density. This effect would be mitigated by refining the grid points (x_k), which in turn increases computational time.

In Figure 3.6 we show some simulated VARFIMA$(0, d, 0)$ eigenvalue densities and compare them to the Marchenko–Pastur law one would get for $d = 0$ and Gaussian white noise.

One can identify different shapes for positive resp. negative fractional integration orders. We also chose a parameter value of $d = 0.3$ where the simulated eigenvalue density functions has some clearly visible artifacts. Also note the small artifacts present for small values of x.

The estimated eigenvalue densities have to be used with some caution. Different from previous numerical simulations, the eigenvalue densities do not integrate to 1 but rather to values about 0.78. In fact the eigenvalue densities for the fractional integration orders $d \in \{-0.4, 0.1, 0.3\}$ sum to 0.783, 0.786 resp. 0.89. The last value for $d = 0.3$ is distinctly larger than about 0.78 because all the artifacts also contribute to the sum. For the plot and subsequent use we normalized each eigenvalue density so that it integrates to 1. But obviously this procedure lowers all values of eigenvalue densities with artifacts present. Following this reasoning the eigenvalue density for $d = 0.3$ should be depicted a bit higher than it is.

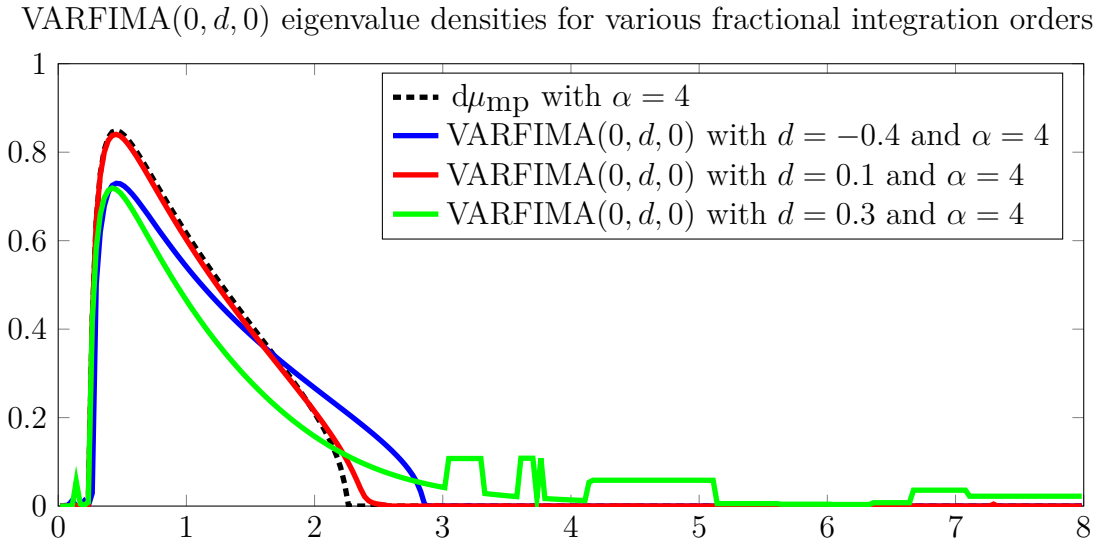

Figure 3.6: Comparison of VARFIMA$(0, d, 0)$ eigenvalue densities for various fractional integration orders d, all with $\alpha = 4$. Estimated with algorithm 4.1.

Remark: *One can detect the presence of artifacts quite well by looking at the residues* (3.43) *of the* argmin. *Almost all artifacts have abnormal high residues of more then* 5 *orders of magnitude larger in* $\log_2$. *The problem remains to find the true values. One could try other minimization algorithms for grid values* x_k *that produced abnormal high residues in the first minimization run. Unfortunately this makes it hard to automate the generation of* VARFIMA$(0, d, 0)$ *eigenvalue densities.*

3.2.5 VARFIMA$(1, d, 1)$ processes

The spectral density of an ARFIMA$(1, d, 1)$ process (X_t) is because of (3.38) in conjunction with (3.39) and (3.31) given by

$$\tilde{\gamma}_{\text{ARFIMA}(1,d,1)}(\omega) = \frac{1}{2\pi} \frac{1 + \phi_1^2 + 2\phi_1 \cos(\omega)}{1 + \theta_1^2 - 2\theta_1 \cos(\omega)} (2 - 2\cos(\omega))^{-d}, \quad \omega \in [0, \pi]. \tag{3.44}$$

The spectral density function (3.44) can assume many different shapes. This flexibility is great for modeling purposes as it permits a very parsimonious representation for a wide variety of situations. Unfortunately, due to the many shapes possible it is very tedious to find a closed form inverse spectral density. The spectral density in $\tilde{\gamma}_{\text{ARFIMA}(1,d,1)}(\omega)$ is in general not anymore monotone on $[0, \pi]$, so the inverse has to be defined piecewise. At most, except for possible degeneration, the inverse consists of 3 different piecewise functions. This amounts to saying that the spectral density $\tilde{\gamma}_{\text{ARFIMA}(1,d,1)}(\omega)$ has some value(s) λ in its range which is the image of 3 different arguments $\omega_1, \omega_2, \omega_3$. It is a tedious task to calculate the domain of each of the piecewise functions, identify which of the endpoints of the domain is smaller and determine the right permutation the piecewise inverse functions have to be in.

We opted for numerical inversion, then searched for the appropriate real inverse piecewise functions and subsequently numerically differentiated $\tilde{\gamma}_{\mathrm{ARFIMA}(1,d,1)}(\omega)$ in order to numerically approximate the function g.

In Figure 3.7 we compare the eigenvalue density of a VARMA$(1, 1)$ process with the eigenvalue density of a VARFIMA$(1, d, 1)$ process, where only the fractional integration parameter d changes from $d = 0$ to $d = 0.4$.

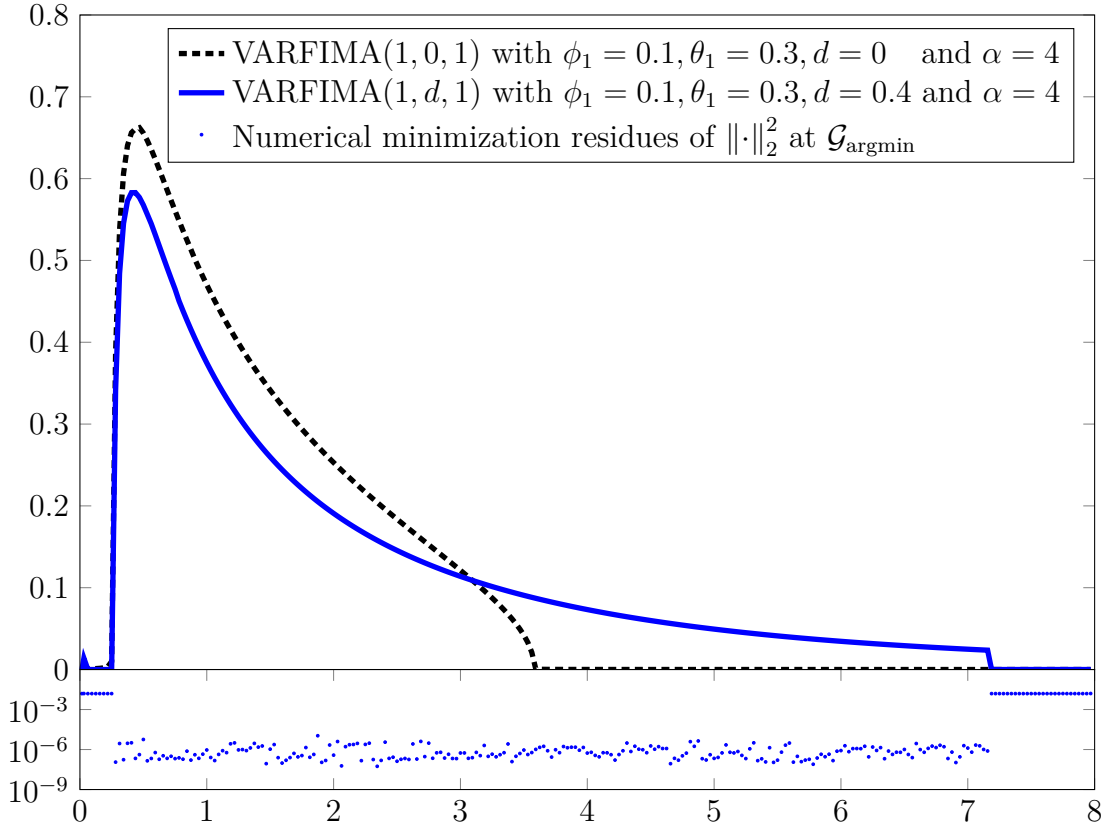

Figure 3.7: Comparison of VARFIMA$(1, d, 1)$ eigenvalue densities with fractional integration orders $d \in \{0, 0.4\}$, both with $\alpha = 4$.

The VARFIMA$(1, 0, 1)$ = VARMA$(1, 1)$ process is the one depicted in Figure 3.5. The positive fractional integration order is with $d = 0.4$ a relatively high value of all admissible values $d \in (-2^{-1}, 2^{-1})$ and visibly prolongs the right tail. We also depicted the residues from the numerical minimization for each grid point x_k in the semi-log plot beneath the densities. There is a distinct pattern visible. Residues have values of about $2^{-6} \approx 0.016$ if the minimization algorithm did not converge properly, otherwise the residues are in $[5.519 \cdot 10^{-8}, 1.085 \cdot 10^{-5}]$.

Remark: *We deliberately chose to depict a parameter combination where the minimization algorithm for grid points x_k of the far right tail does not converge to caution the reader and underline the relevance of some more in-depth remarks on numerical methods used and their limits.*

3.3 Numerical Methods and Algorithms used

3.3.1 Calculating the DGP for VARMA(p, q) processes

For generation of the distribution generating polynomial we opted to re-write the VAR-process as VMA-process with the parameter transformations already explained on p. 43 and p. 46.

Formula (3.24) only gives a general function of nested square roots. One can try to expand and combine these square roots and transform to a rational function, that is a function that can be written as quotient of two polynomials. To convert a rational function into a polynomial, we have to expand by the denominator polynomial. The much harder part is to power-expand the nested square roots simultaneously making sure that the solution set of the equation does not change.

In Mathematica, we defined patterns to combine terms and applied this patterns in an recursive way locally. Without the aid of Computer Algebra Systems like Mathematica combining terms would have been quite futile. For instance, Mathematicas `LeafCount` is 4 227 for the VMA(2) DGP. To get a feeling on how to interpret Mathematicas `LeafCount`, the DGP for the VMA(2) as shown in subsection A.3.1 with `LeafCount` of 4 227 amounts to a formula about 5 pages long.

With tedious and lengthly calculation we were able to give the DGPs for both, the VMA(2) and the VAR(2) processes. Both are polynomials of degree 9. The VMA(2) DGP is shown in subsection A.3.1 and the VAR(2) DGP is shown in subsection A.3.2.

The Mathematica Code can be found in subsection B.1.1.

Remark: *To the best of out knowledge, we are the first to derive the distribution generating polynomials for general* VMA(2) *and the* VAR(2) *processes. This lays the foundation for the subsequent calculation of their spectral densities.*

3.3.2 Determining the right root of the DGP for VARMA(p, q) processes

Let a DGP be given as calculated in subsection 3.3.1. We use the relationship between $\mathcal{M}$-transform and $\mathcal{G}$-transform given in (2.10) to transform back to the $\mathcal{G}$-transform and subsequently apply the inverse Cauchy-transform theorem to arrive at a possible spectral density — if one chose the right root. Thus, for given parameters and rectangularity ratio α every root $(c_k)_{k\in\{1,\dots,9\}}$ remains a function of the argument of the $\mathcal{G}$-transform $z \in \mathbb{C}$. i.e. $c_k : \mathbb{C}_+ \to \mathbb{C}$, $z \mapsto c_k(z)$.

If the DGP has degree up to 4, there exists an, albeit complicated, closed form

solution for each root. The real difficulty arises for DGPs exceeding polynomial degree 4. There exist special parameter constellations which lower the degree of the DGP. But as we want to obtain the spectral density for all possible parameter constellations we have to numerically solve for the right root.

As every root of the DGP, depending on in which domain it is defined, can be interpreted as a $\mathcal{M}$-transform or $\mathcal{G}$-transform, we trivialy transform to the implied $\mathcal{G}$-transform. Every valid $\mathcal{G}$-transform has to obey $\lim_{|z|\to\infty} z\mathcal{G}(z) = 1$, so we check whether this limit holds numerically. If it fails to do so, we discard the explored root and move to the next root.

Assuming for now that c_1 is the right root, $c_1(z)$ gives the spectral density at x by the formula $-\pi^{-1} \operatorname{Im} c_1(x + \mathrm{i}\varepsilon)$, for $\varepsilon \searrow 0$. So in order to arrive at the values for the spectral density, one has to find the numeric root for $c_1(x + \mathrm{i}\varepsilon)$ along $x \in \mathbb{R}_+$ and ε small. We usually set $\varepsilon := 2^{-20}$ and have to trace out the spectral density for many values of x.[41] Mostly, the density has its domain contained in $[0, 8]$ and we sample x_i with $|x_{i+1} - x_i| = 2^{-4}$.

So from the perspective of the root finding algorithm given the parameters $(\boldsymbol{\theta}, \alpha, \varepsilon, x)$ one has to find the right of all possible roots. Calculating the numerical roots one has to move along the same branch of root and must not unintentionally switch. But this could happen if for some parameter constellation two roots collapse to the same value in $\mathbb{C}$ and the root finding algorithm does not know which branch of both roots to follow next. In practice, the roots do not have to coincide, it suffices for two roots being close together to distort the root finding algorithm.

In Figure 3.8 we depict the described challenges. The different roots imply each, by means of the inverse $\mathcal{G}$-transform, a »density« function. Only one of the true roots generates the searched for spectral density. This spectral density is a *real* probability density function in the sense that it integrates to 1.

[41] $\varepsilon = 2^{-20}$ is just 3.5 orders below the square root of the machine epsilon for *binary64* `doubles` of 2^{-53}, see »754-2008 IEEE Standard for Floating-Point Arithmetic« in [50]. This should provide enough leeway for numerical stability and is also sufficiently small.

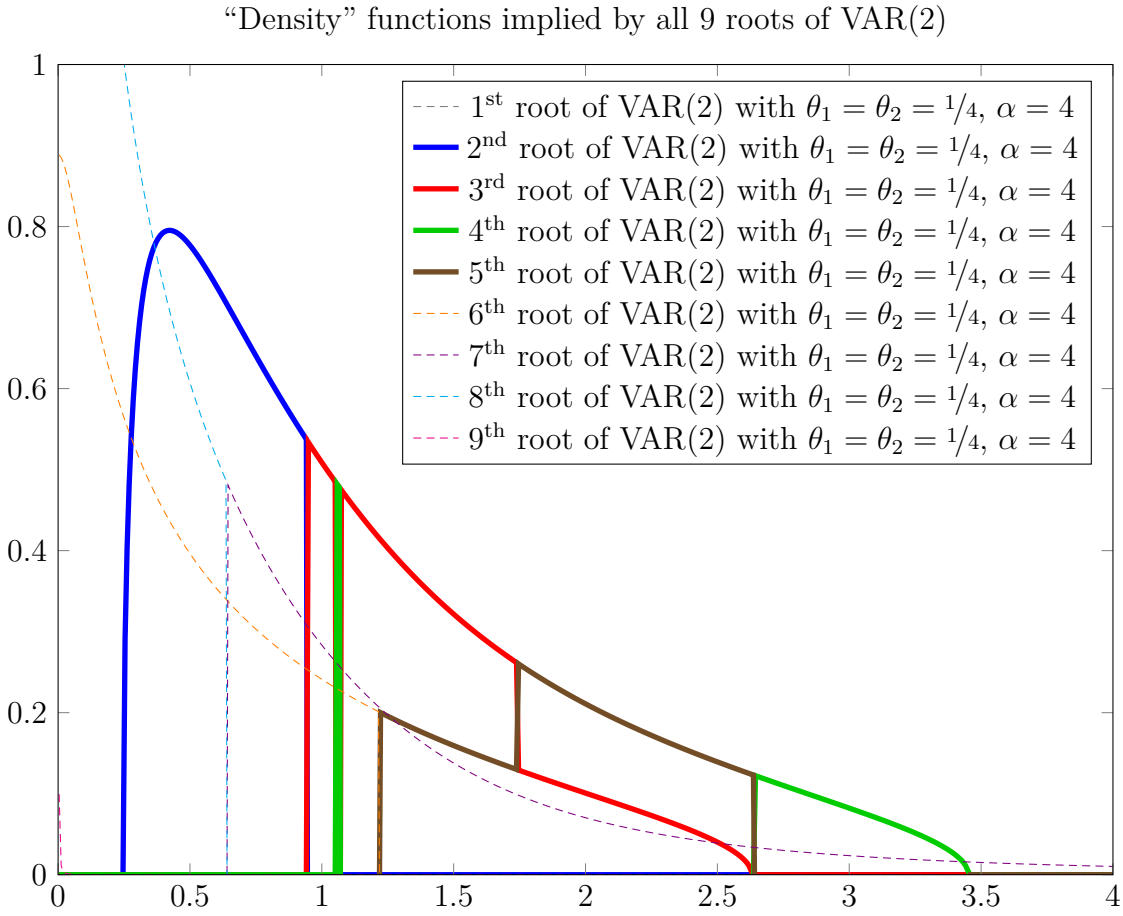

Figure 3.8: »Density« functions implied by all 9 roots of VAR(2) with $\theta_1 = \theta_2 = 1/4$ and $\alpha = 4$. Functions corresponding to the true spectral density in bold, all other dashed.

It is clearly visible that none of the roots on its own generates the searched for spectral density of the VAR(2) process. In fact, none of the »density« functions implied by the roots integrates to 1. The integrals range from -3.784 over 0.015 to 3.37181. Negative integrals imply that the corresponding root assumes negative values, which also dominate the integral. The negative part of the roots is not shown in Figure 3.8 as we are only interested in the root leading to the spectral density.

By closer inspection one also notices that a piecewise combination of roots $\{2, 3, 4, 5\}$ gives a density that looks like the one we are searching for. In fact, the piecewise defined function integrates numerically to $0.999\,37 \approx 1$. This is evidence that the numerical root finding algorithm sometimes switches between different root branches. By visual inspection, when branches of different roots cross the numerical root finding algorithm might choose a different branch. So we have to correct this shortcoming and manually reconstruct the complete branch of the true root that generates the spectral density distribution. But this is computationally expansive, as we have to check at every possible, positive-valued furcation which branch of root to follow. For the parameter values $\theta_1 = \theta_2 = 1/4$ and $\alpha = 4$ depicted in Figure 3.8 the true spectral density consists of 4 different numerical root branches.

The complicacy of the 9 roots of a typical VAR(2) distribution generating polynomial in contrast to the simple polynomial $M^9 + (x + \mathrm{i}y)$ is depicted in subsection A.3.3, plots Figure A.1 and Figure A.2. The complex structure any numerical root finding algorithm has to cope with is obvious.

Numeric and thus algorithmic problems like the one occurring for calculating the spectral density of VAR(2) processes will be likely problems for more complicated VARMA(p, q) processes.

3.3.3 Numerical Integration for VARFIMA(p, d, q) processes

For numerical approximation of the integral in (3.42) with an appropriate function g we used a *global adaptive* integration strategy to recursively bisect integration intervals based on its specific error. This ensures parsimony of the number of integrand evaluations as the integration interval is not sampled uniformly, but where the global integration error is reduced most.

We use Gaussian quadrature rules[42] for integration of the specific subintervals. The main idea of those is to approximate the integrand locally as weighted sum of polynomials. Picking n *sampling points* also implies n weights, which makes the Gaussian quadrature for n sampling points exact for polynomials up to order $2n-1$.[43] Interpolation schemata with polynomials at equidistant spaced points belong to the class generated by the Newton–Cotes formula. A convenient property of quadrature rules with equidistant evaluation points is that they *nest*, i.e. a refinement of sampling points in the integration interval re-uses all previously evaluated points. This is not true in general for non-equidistant sampling points, but those methods sample more appropriate locally in regions of greater absolute variation of the integrand. Also, Newton–Cotes rules can suffer from so-called *Runge's phenomenon*, that is the integration error at the edge of the interval increases exponentially with growing n. This is because a polynomial of order $n - 1$ can always be made to go through n points, but it might oscillate quite heavily in accomplishing this task.

The class of so-called *Gauss–Kronrod quadrature rules*[44] combines both afore mentioned advantages as it does not require to use equidistant sampling points but does nest for subsequent refinements. It reuses $n + 1$ previously calculated optimal sampling points and in doing so is exact at step n, for accumulated $n + (n + 1)$ integrand evaluations, for polynomials of order $3n + 1$. The next Kronrod–Points can be efficiently calculated by a 5 term recurrence relation as the zeros of the orthogonal family of Stieltjes polynomials.[45] The nesting property also comes handy in combination with the global adaptive integration strategy we use to subdivide

[42] See, for instance, the classic [84, section 4.6].
[43] We lose one degree because we need 2 points to parametrize a linear-affine monomial, 3 points for a quadratic polynomial and so on.
[44] See [56].
[45] See [57, eq. (18), see also eqs. (19)–(23)].

the integration domain into subintervals. Both strategies ensure that not too many costly integrand evaluations are required to reach some convergence criterion.

If the function g is defined piecewise, it often has singularities at the boundary points or exhibits singularity-near behavior. This issue has to be addressed to efficiently calculate integrals of integrands where the function g is the main part. Otherwise the global adaptive algorithm keeps bisecting subintervals, because new integrand evaluations approaching a singularity keep changing the integration value and prevent the integration error estimate from decreasing beneath a pre-defined threshold to terminate the algorithm. We chose a number of 4 bisections to indicate a possible singularity in the subintervals where the local error estimate is not below the stopping threshold of the algorithm. We follow Iri et al., who 1970 proposed in [51] a variable transformation method for quadrature rules in such a way that all odd-order derivatives of the transformed integrand vanishes near the singularity. This lets the integration error decrease at a much faster rate than with about $\mathcal{O}\big(N^{-2(m+1)}\big)$, where N is the number of subintervals and m is the order of the remainder term in the Euler–Maclaurin approximation.[46]

Numerical Minimization

After integral (3.42) is calculated, equation (3.29) implies solutions for $\mathcal{G} = \mathcal{G}_{\hat{F}}(z)$, which can be found as roots to a form like the one in (3.33). Unfortunately, searching for a root $\mathcal{G}_{\hat{F}}(z)$ changes the value of the integral, which is part of the equation. In addition, since we only approximated the integral we cannot hope to really find an exact root; it may happen that one only finds a local minimum with an absolute value near zero. Because of the dependence of (3.29) on the integral on the complex argument $\mathcal{G} = \operatorname{Re}\mathcal{G} + \operatorname{Im}\mathcal{G}$, we essentially treat minimization of the norm of form (3.43) as a black-box problem. We can only evaluate the norm for different values of $\operatorname{Re}\mathcal{G} + \operatorname{Im}\mathcal{G}$ and search for a local minimum. The closeness of the imaginary value of the found numerical minimum to the exact one determines the accuracy of the final spectral density estimate at point $z = x + \mathrm{i}\varepsilon$.

The step of numerical minimizing the squared norm of the difference of LHS and RHS of (3.29) is conceptually detached from the numerical integration algorithm, as there do not seem to exist any obvious strategies on how to utilize intermediary results of the numerical integration for the minimization algorithm. Not having to build a combined
error.

For the described minimization problem the derivative of the function is not known and function evaluation is, because of the numerical integration necessary, quite expensive. The objective function is the squared norm (3.43), which is a real-valued

[46] See [51].
[47] See [83].
[48] See [77].

Algorithm 3.1: Norm Minimization for Density

Input: α rectangularity parameter; d integration order; ε small positive parameter; $x_{\max}$ range from 0 to $x_{\max}$; bw spacing between successive x-values

Output: vector $\mathbf{y}$ of equal spaced density values

Function `NumIntGaussKronrod`$(f(\lambda), \lambda_-, \lambda_+)$ **{**

  use Gauss–Kronrod rule[47] to integrate

  $(x + \mathrm{i}y) \coloneqq \int_{\lambda_-}^{\lambda_+} f(\lambda)\,\mathrm{d}\lambda$

  return $(x + \mathrm{i}y)$

}

Function `MASTER`$(x, y, k, \alpha, d, \varepsilon)$ **{**

  if $d > 0$ **then**

    $\lambda_- \coloneqq 4^{-d}$

    $\lambda_+ \coloneqq \infty$

  else

    $\lambda_- \coloneqq 0$

    $\lambda_+ \coloneqq 4^{-d}$

  end

  define $g(\lambda, d)$ either by (3.41) (ARFIMA$(0, d, 0)$)

  or by $g(\lambda, d) \coloneqq \sum_{\omega \in [0,2\pi]:\tilde{\gamma}(\omega)=\lambda} \left| \frac{\mathrm{d}}{\mathrm{d}\omega} \tilde{\gamma}(\omega) \right|^{-1}$ with $\tilde{\gamma}(\omega)$ given by (3.44) (ARFIMA$(1, d, 1)$)

  $f(\lambda) \coloneqq \int_{\lambda_-}^{\lambda_+} \frac{\lambda}{1 + \lambda\alpha^{-1}(x + \mathrm{i}y)} g(\lambda, d)\,\mathrm{d}\lambda$

  $(a + \mathrm{i}b) \coloneqq$ `NumIntGaussKronrod`$(f(\lambda), \lambda_-, \lambda_+)$

  $z \coloneqq \left(-\frac{\alpha}{(x + \mathrm{i}y)} - \alpha(k + \mathrm{i}\varepsilon) + \frac{\alpha}{2\pi}(a + \mathrm{i}b) \right)^2$

  return z

}

Function `minMASTER`$(k, \alpha, d, \varepsilon)$ **{**

  use Nelder–Mead algorithm[48] to minimize over $(x, y) \in \mathbb{R}^2$

  $(x + \mathrm{i}y) \coloneqq \operatorname*{argmin}_{x,y}$ `MASTER`$(x, y, k, \alpha, d, \varepsilon)$

  return y

}

Function `getDensityValues`$(d, \alpha, binwidth, x_{\max}, \varepsilon)$ **{**

  let $\mathbf{x}$ contain the equaly spaced sequence from 0 to $x_{\max}$ with $|x_{i+1} - x_i| = binwidth$

  foreach *entry* x_i *of* $\mathbf{x}$ **do**

    $y_i \coloneqq$ `minMASTER`$(x_i, \alpha, d, \varepsilon)$

  end

  return $\mathbf{y}$

}

$\mathbf{y} \coloneqq$ `getDensityValues`$(d, \alpha, binwidth, x_{\max}, \varepsilon)$

---

function of the complex argument $\mathcal{G} = \operatorname{Re}\mathcal{G} + \operatorname{Im}\mathcal{G}$. Ultimately we are interested in the imaginary part of the argmin of $\mathcal{G}$.

The $\mathcal{G}$-transform maps arguments $z = x + \mathrm{i}\varepsilon \in \mathbb{C}_+$ to the positive complex half-plane $\mathbb{C}_+$, so we have to restrict the domain for the minimization algorithm to $\mathbb{C}_+$. We therefore have to choose a minimization algorithm capable of coping with constrained optimization.

We opted for the robust search heuristic from Nelder and Mead for multi-dimensional non-linear minimization problems without known closed form derivatives. It essentially constructs a simplex of $d+1$ points for a d-dimensional minimization problem. At each iteration step the $d+1$ points form a simplex in dimension d. The Nelder–Mead algorithm[49] tries to minimize the volume of the simplex in each iteration step by reflecting from the centroid of the simplex in the opposite direction of its largest function value and shrinking the simplex. The algorithm tries to enclose the true argmin inside the simplex and returns, for the volume very small beneath a threshold, the argmin of the current $d+1$ simplex points. Since the squared norm depends only on the complex-valued $\mathcal{G}$ for fixed $z = x + \mathrm{i}\varepsilon$, we have $d = 2$ real arguments and as simplex a triangle. Intermediary expansion and contraction steps try to ensure validity resp. faster convergence. We chose the reflection and expansion parameters to be both 2 and the contraction and shrinking parameters to be both 2^{-1}.

The Nelder–Mead algorithm can cope with constrained optimization by incorporating a penalty function indicating constraint violations. As the constraint is simple and therefore the sets of admissible and inadmissible solutions are connected partitions of the whole domain $\mathbb{C}$, the Nelder–Mead algorithm can simply ensure in every step that at least one of the simplex nodes x_k does not violate the constrained and is thus in the complex upper half-plane. This implies that the closure of the interior of the simplex contains at least an admissible point. The algorithm has to check before termination that the obtained candidate x_{argmin} does not violate the constrained, too.

Remark: *Sometimes we had to adjust the parameters mentioned in the paragraph above to greatly boost chances for the algorithm to converge to the global minimum and not just a local one. We also tried to incorporate any additional information we had by choosing appropriate starting points for the simplex.*

Since functional evaluations are so computational expensive we could not rely on *Monte Carlo* resp. *Quasi Monte Carlo* methods or grid search, as to make the computation practicable the running time of the algorithm has to be bounded in the order of magnitude of minutes, not hours.

[49] See [77].

Parallelization and further Speed-Up

Only the numerical integration algorithm can be easily speed up by parallelization, as the integrand can be evaluated in parallel without causing any interdependence to other calculations. The numerical minimization is not easily parallelized, but the essential computational cost are its function evaluations anyway.

The final combined algorithm of numerical minimization and numerical integration traces the spectral density on a grid (x_k) of equidistant nodes for $z_k = x_k + \mathrm{i}\varepsilon$. This task parallelizes, if one does not wish to use the information from the predecessor $\mathcal{G}_{\hat{F}}^{\mathrm{argmin}}(x_{k-1} + \mathrm{i}\varepsilon)$ for determination of the starting points for the Nelder–Mead algorithm. We chose to utilize this information, because we did not need to parallelize on this level. The ultimate task was to fit spectral densities to different empirical density functions corresponding to different trading periods — and this outermost loop is *pleasingly parallel*, too.

In addition, we tried to port bottleneck portions of the source code to `C` for speed. `C` is a low-level programming language, its source code is compiled to machine code and the compiler[50] fine-controls CPU and memory usage.

[50] We use the `C++` compiler from *Visual Studio Community 2015*, [68].

4 Overview of Data and Descriptive Statistics

We introduce the basic methodology used for the construction of raw data sets in section 4.1. In section 4.2 we describe the raw data sets and preliminary transformation necessary to coherently analyze the data series using descriptive statistics. In section 4.3 we give an overview of their main characteristics. We use descriptive statistics to select suitable subsets of data for refined data analysis. We especially concentrate on the empirical spectral distribution in section 4.4, as the ESD is the empirical counterpart to the theoretical spectral density. In chapter 5, we want to fit the theoretical spectral densities from the previous chapters to the ESDs of covariance matrix estimates from data sets composed in this chapter.

4.1 Basic Methodology for Constructing Raw Data Sets

Our aim is to construct several data sets, each comprised of logarithmic return data suitable for covariance matrix estimation. First, we have to investigate the basic building blocks of logarithmic returns, stock prices. We can observe stock prices as a pair of a transaction time t_i and a price p_{t_i} agreed upon by both parties, seller and buyer. These transaction prices, also called *tick data*, happen naturally neither evenly spaced in time nor synchronous across all stocks. As we are mainly interested in covariance estimates, we have to pay attention to the so-called *Epps effect*, introduced by Epps [30]. The *Epps effect* is the observable phenomenon that the empirical correlation between the returns of two different stocks decreases in absolute value as the sampling frequency of data increases. Münnix et al. [73] argue that this idiosyncratic characteristic of stock returns is caused by non-synchronous trading and microstructure noise. As Merton notes in [67, p. 357], in an ideal situation where prices can be observed continuously without any measurement error squared intraday returns would yield a variance estimate without any error for every finite interval. The squared intraday return variance estimator, commonly called the »realized variance (RV)« estimator, suffers from a well-known estimation bias that increases with intraday sampling frequency, see e.g. [32, 7, 80]. Hansen and Lunde define market microstructure as the source of this bias problem [39]. According to them, »*[. . .] there is a trade-off between bias and variance when choosing the sampling frequency, as discussed by Bandiy and Russell [11] and Aït-Sahalia et al.*

[1]. This trade-off is the reason that the RV is often computed from intraday returns sampled at a moderate frequency, such as 5*-minute or* 20*-minute sampling.«* ([39, p. 128])

As we want to study theoretical properties of covariance matrix estimators and infer from these *ex ante* properties and knowledge of their empirical *ex post* counterparts about the underlying market structure, we opt for a simple structure of covariance matrix estimators. Since we chose to not focus our research on microstructure noise[1], we opt for circumventing the associated problematics by using aggregated transaction prices, synchronized in time across different stocks.

This decision comes in handy to reduce the size of data sets to levels suitable to process with personal computers and up to 16 GB RAM.

To make analysis of subsamples easy, we require equitemporal observations. This requirement is closely linked to the decision to work with synchronized data. A priori synchronization of transaction prices just means that we are able to assign to each observation time t_k a price $p_{t_k}^{(j)}$ for every stock j. We could choose a set of observation times $\{t_1, t_2, \ldots, t_n\}$ with varying time interval lengths $|t_{k+1} - t_k|$. This would be a sensible choice if we took t_{k+1} to be the time after which every stock j has a new transaction price $p_{t_{k+1}}^{(j)}$. But this approach is not feasible, because we already examine the most liquid stocks available and we are still in a situation that there exist periods in time where liquidity dries out and a specific stock is not traded once for hours. Dacorogna et al. state that »*[. . .] market ticks arrive at random times*« ([28, p. 34]). They go on and refer to, e.g. [29], for »*[. . .] models for the stochastic nature of these times.*« ([28, p. 34]). Due to this stochastic nature there is some self-averaging of the expected number of stocks with at least one transaction price in a given time interval. This supports the stipulation for exogenously provided observation times t_i detached from actual trading times. This has the further benefit of not having to discard massive amounts of the available price data. A drawback of this method is that we cannot ensure that in a given time interval $(t_k, t_{k+1}]$ every stock j has a new transaction price. For every stock and for time t_k, we used the last observation in the time interval and set accordingly price $p_{t_k} \coloneqq p_{t_i}$ for $\max\{i \mid k-1 < i < k\}$. If the interval $(t_{k-1}, t_k]$ was empty, we set p_{t_k} to the last observed price on that day[2] instead. [28, p. 37] call this procedure self-explanatory the *previous-tick-interpolation*, taking the most recent value. This procedure is in agreement with how trading prices are calculated. The current price shown is the last price at which trading occurred. As we want to model stock prices, and via a transformation stock returns, as stochastic variables, the property of only utilizing information up to the present is closely linked to the notion of *adapted processes*[3] in classical probability theory. This property is important for a realistic model that is operational and can be readily utilized by practitioners.

[1] For a general treatise on noise in economics and finance, see [15].
[2] If there was no previous transaction for that stock on this day, we set $p_{t_i} \coloneqq$ `NA`.
[3] See [92, p. 5, 1.20 Definition a)].

4.2 Overview of Raw Data

We study different data sets of price data of stocks from various countries, traded at several different exchanges and with temporal granularity ranging from milliseconds to days.

4.2.1 NASDAQ Intraday Data

We use NASDAQ's Historical TotalView-ITCH data, which contains historical order and trade transactions from the NASDAQ exchange.

We downloaded NASDAQ's Historical TotalView-ITCH data from LOBSTER[4], a data vendor for academic research. Since we are only interested in trading prices we only used the order book of depth 1, best bid and ask orders. From this, we extracted all transaction orders and collected the corresponding prices and time stamps in milliseconds. Because we aggregated the order book data to transactions only, the information about bid/ask is irrelevant.

The NASDAQ Stock Market trading hours[5] for the regular market are between 9:30am–4:00pm, 6.5 hours in total.

NASDAQ-100 Intraday Data from 2007-06-27 to 2015-01-16

At January 23, 2020, the NASDAQ-100 has, despite its name, 107 constituents[6]. At December 13, 2014, NASDAQ OMX Group announced their annual changes[7] to the NASDAQ-100 index:

> "*The following three companies will be added to the Index: American Airlines Group, Inc. (Nasdaq:AAL), Electronic Arts Inc. (Nasdaq:EA) and Lam Research Corporation (Nasdaq:LRCX).*
> *[...]*
> *As a result of the re-ranking, the following three companies will be removed from the Index: Expedia, Inc. (Nasdaq:EXPE), F5 Networks, Inc. (Nasdaq:FFIV) and Maxim Integrated Products, Inc. (Nasdaq:MXIM).*" ([75])

We included the three companies that were excluded from the NASDAQ-100 at this day in our list of NASDAQ-100 companies, because the overwhelming majority of

[4] LOBSTER: Limit Order Book System — The Efficient Reconstructor at Humboldt Universität zu Berlin, Germany. http://LOBSTER.wiwi.hu-berlin.de
[5] See [76]: http://www.nasdaqomx.com/aboutus/market-information/market-calendar
[6] The official list can be found here: https://indexes.nasdaqomx.com/Index/Weighting/NDX
[7] http://www.nasdaqomx.com/newsroom/pressreleases/pressrelease?messageId=1320684

observations are from the observation period from 2007-06-27 to 2015-01-16. These are 1900 trading days.

To reduce data size, we aggregated all prices to 10 second intervals. We choose 10 seconds as a compromise to fit the whole data set into 16 GB of RAM, with enough memory left to process the data with GNU R. Dividing 6.5 hours by 10 second intervals gives 2 340. For the sake of completeness, we also include the first time point 9:30:00. Since at this point in time no trading could have occurred, the price values are all NA. In total, this makes 2 341 observations per day and $1900 \cdot 2\,341 = 4\,447\,900$ observations per stock.

NASDAQ Intraday Data from 2007-06-27 to 2014-11-28

We downloaded the company list from [74, csv link]. It encompasses 2 954 companies traded at the NASDAQ exchange. Only 7 of those companies were not available (NA) on LOBSTER, so there are 2 947 companies left. Of those, 1 618 companies were listed the complete time period under investigation from 2007-06-27 until 2014-11-28.

The observation period is comprised by the 1867 trading days between 2007-06-27 and 2014-11-28. Due to memory constraints and data quality we aggregated all prices to 60 second intervals, similar to subsection 4.2.1. That makes 60 seconds times 6.5 hours plus the first time point, so in total 391 observations per stock and trading day. In total, this are $1\,867 \cdot 391 = 729\,997$ observations per stock.

4.2.2 S&P 500 and DJIA Intraday Data

We utilize the historical intraday stock data provided by the data vendor pitrading.com.[8] The so-called »stock data« data set consists of all constituent stocks of various indices and exchange traded funds from the United States (U.S.). We are only interested in indices comprised of at least a medium number of sufficiently liquid stocks.

- »S&P 500 Index Component Stocks (S&P 500)« with 500 large publicly owned U.S.-based companies
- »Dow Jones Industrial Average Component Stocks (DJIA)« with 30 large publicly owned U.S.-based companies

Both index data sets span the time period of 3454 trading days between 2002-12-30 and 2016-09-16, with intraday »OHLC«[9] stock prices available for every minute from 9:30 to 15:59. The closing price for e.g. minute 9:30 is the last traded price between 9:30:00 and just before 9:31:00. For these two data sets we have 60 observations per

[8] [82], pitrading.com with historical intraday stock data up to 2016-09-16.
[9] The acronym »OHLC« stands for the prices »Open«, »High«, »Low« and »Close«.

trading hour times 6.5 trading hours per day, i.e. in total 390 observations per stock and trading day. For stocks traded over the whole period under investigation this makes in total $3\,082 \cdot 390 = 1\,201\,980$ observations per stock. In both data sets a vast majority of stocks, which were an index member at the end of our observational period at 2016-09-16, have a trading history available from at least the start of our observational period 2002-12-30. Therefore, we discard all stocks which were not traded over the full observational period.

S&P 500 Intraday Data from 2002-12-30 to 2016-09-16

There are 405 stocks that have a complete trading history between 2002-12-30 and 2016-09-16. This does not mean that all 405 of this stocks were in the S&P 500 for the whole observational period. For instance the Office REIT[10] »SL Green Realty Corp. (NYSE:SLG)« from the finance sector only became a S&P 500 constituent 2015-03-20.[11] Nevertheless, its stock was publicly traded over the whole observational period, therefore we utilize its price history.

DJIA Intraday Data from 2002-12-30 to 2016-09-16

As it was the case for some S&P 500 stocks, for instance »Apple Inc. (AAPL)«, besides »Microsoft Corp. (MSFT)« and »Cisco Systems (CSCO)«, only one of three NASDAQ listed companies, was added 2015-03-19 to the DJIA. Of course its trading history is known and will be used. Of all 30 DJIA stocks as of 2015-03-27, the consumer bank »Visa Inc. (V)« had its IPO 2008-03-18 and therefore cannot be included in the data set. All remaining 29 constituents of the DJIA have a complete trading history and will be used.

4.3 Descriptive Statistics and Data Sets

We want to give an overview and brief understanding of certain characteristics of the data sets. Our goal is to retrieve a subsample of data as large as possible, which is suitable for further investigation. Our main application is the estimation of covariance matrices from historical returns. There are two main obstacles to this goal when viewing the raw data.

First, data completeness might be problematic if not all stocks were traded during the whole time period, indicated by days with only `NA` values. Intraday data completeness might also be an issue, as we want to apply standard covariance matrix estimators which cannot handle any `NA` values.

[10] real estate investment trust (REIT).

[11] »*S&P MidCap 400 constituent SL Green Realty Corp. (NYSE:SLG) will replace Nabors Industries Ltd (NYSE:NBR) in the S&P 500, [...]* « ([86, p. 1]).

Second, as covariances are the second central moments, covariance estimation requires sufficient variation around the mean. This is not in itself a problem of the covariance estimator, but with many observations equal to the population mean the spectrum of such an estimator might be distorted. Since we want to analyze this very spectrum, we have to aggregate returns in time to ensure that enough returns are different from 0. This aggregation in time is equivalent to choosing a time interval length so that we have at least one observed price in (almost) each interval.

All raw price data series are transformed to logarithmic return data series by

$$r_{t_k} := \log\left(\frac{p_{t_k}}{p_{t_{k-1}}}\right).$$

If the prices are intraday, returns are only calculated with intraday prices from the same day. Thus, in case of intraday data the return series lose one observation for every trading day.

Descriptive statistics about data completeness will provide the necessary information to choose appropriate time aggregation and subsamples of stocks.

4.3.1 NASDAQ Intraday Data

NASDAQ-100 Intraday Data from 2007-06-27 to 2015-01-16

The raw NASDAQ-100 data set consists of intraday returns of 110 stocks for the trading days between 2007-06-27 and 2015-01-16. Since we are interested in variations of stock returns and in order to be able to compute proper empirical variances and not merely squared returns, we need at least two returns per day. In Figure 4.1 we depict the portion of days that have at least two returns.

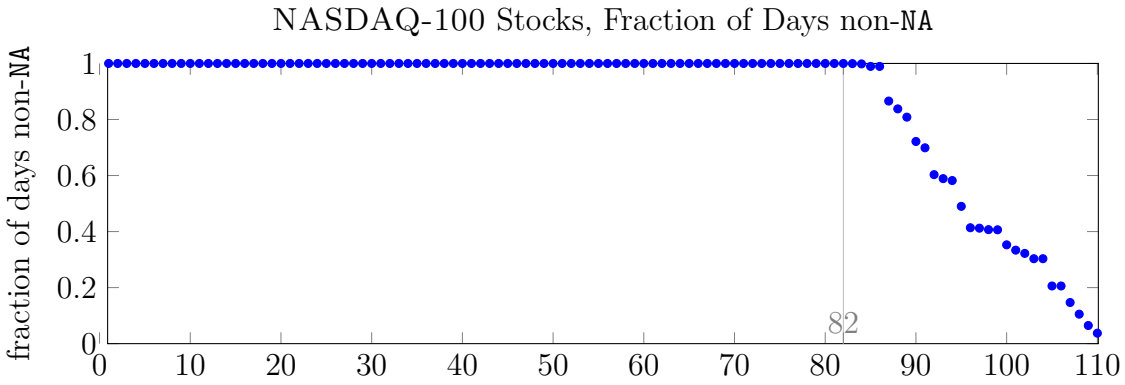

Figure 4.1: 110 NASDAQ-100 stocks, portion of days non-NA and with at least two non-0 returns in trading days between 2007-06-27 and 2015-01-16.

We see that 82 stocks have at least two returns per day. Since for empty intraday intervals $(t_{k-1}, t_k]$ we used the last price on that day, returns r_{t_k} belonging to such

consecutive intervals are zero. A stock is said to have a high liquidity, if one is able to trade it at any given time. Therefore a stock is said to be (more) liquid, if there are transactions in almost every (smaller) time interval. The statistic of non-zero returns is such a measure of a stock's liquidity. Next, we want to investigate the percentage of returns different from zero for every trading day. We do so for the 82 stocks with all non-`NA` returns in Figure 4.2.

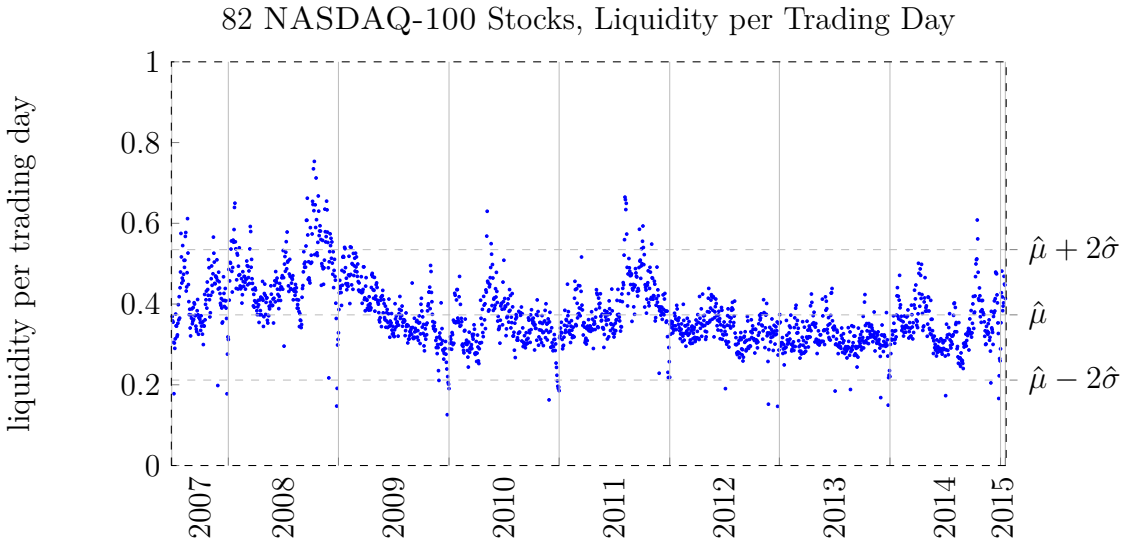

Figure 4.2: 82 NASDAQ-100 stocks, fraction of non-zero returns per trading day between 2007-06-27 and 2015-01-16.

On average, a fraction of $\hat{\mu} = 0.373$ returns is non-zero with an estimated standard deviation of $\hat{\sigma} = 0.081$. We see some seasonality at the change of year, but in general nearly all days have a similar portion of non-zero returns.

Second, we examine the percentage of returns different from zero intraday. We therefore average each intraday time interval over all trading days. The result is depicted in Figure 4.3.

The distribution of transactions and subsequent non-zero returns is much smoother intraday along time than along trading days. A seasonal pattern is evident. It has a higher fraction of non-zero returns at the beginning of every new half-hour. This increased trading volume microstructure might be due to automated trading algorithms that execute orders at predetermined points in time. One also sees increased activity at the very end of the trading day, shortly before 16:00. The fraction of non-zero returns is, of course, also $\hat{\mu} = 0.373$.

Over all 82 stocks, intraday and per trading day, we observe that only a fraction of $\hat{\mu} = 0.373$ returns are non-zero. This low fraction would distort the spectrum of a covariance matrix estimator, so we need to aggregate returns in order to reach much higher levels of non-zero returns.

First, we will concentrate on the 82 stocks that have all non-`NA` days. For this

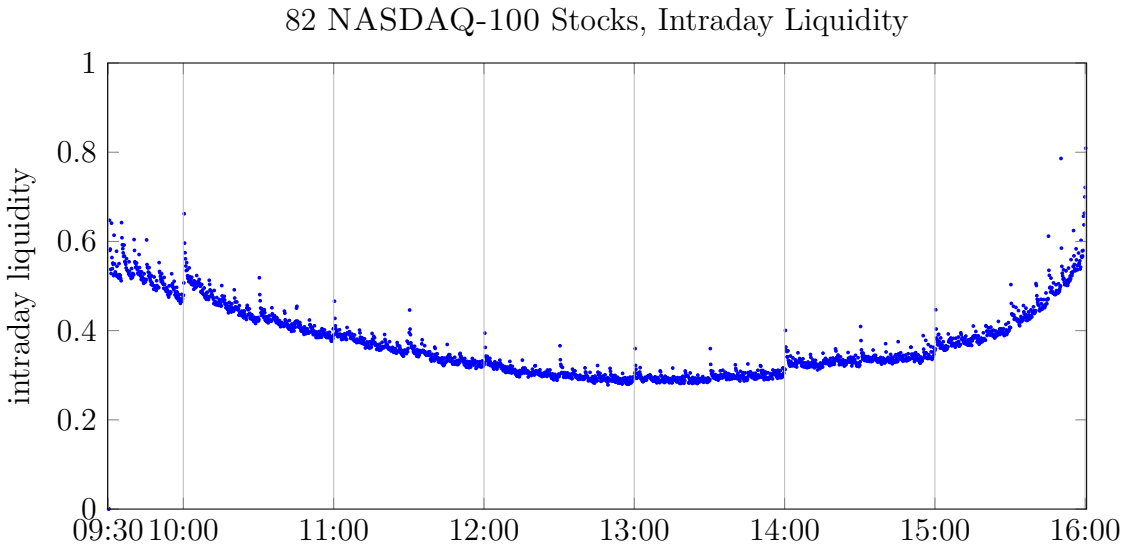

Figure 4.3: 82 NASDAQ-100 stocks, fraction of non-zero returns per stock intraday between 2007-06-27 and 2015-01-16.

82 stocks, Figure 4.4 depicts the sorted fraction of non-zero returns for various aggregation levels. We only choose such aggregation levels so that the total trading time per day of 6.5 hours is divided by this time without remainder.

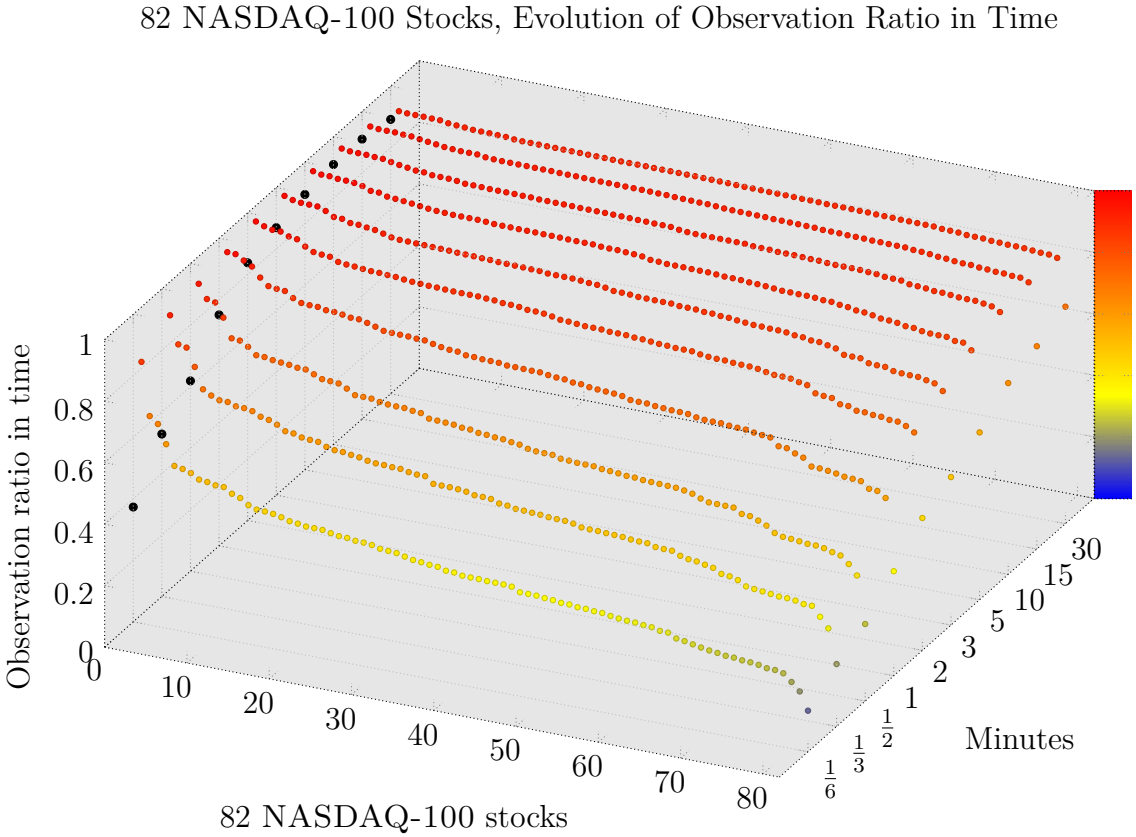

Figure 4.4: 82 NASDAQ-100 stocks, change of liquidity per stock for aggregated returns, sorted. In black the average for each series.

For longer aggregation time levels the fraction of non-zero returns increases. It only starts decreasing from a very high level for the longest time intervals, because the first interval has by definition a return of `NA` and this single `NA` dominates for large aggregation times. Because with every aggregation step the resulting time series shrinks, we only want to aggregate up to a level necessary, but not beyond.

As can be seen from the averaged liquidity statistics depicted as black dots in Figure 4.4, from time intervals of length 2 minutes and longer, aggregation does not enhance liquidity perceptibly. Over the considered liquidity levels, the stocks with highest liquidity are »Apple Inc. (AAPL)« before »Amazon.com, Inc. (AMZN)« and »Google Inc. Class C (GOOG)«.

In Table 4.1 we show a comprehensive summary of key statistics depending on increasing aggregation levels. At aggregation level of 3 minutes and more the number of stocks with liquidity at least 0.75 saturates and average liquidity is well above 0.85.

Time [min]	Number of Observations	Number of Stocks	Liquidity
1/6	4 447 900	1	0.850
1/3	8 899 600	4	0.835
1/2	10 387 300	7	0.842
1	32 687 600	44	0.815
2	**26 440 400**	**71**	**0.858**
3	20 160 900	81	0.874
5	12 158 100	81	0.899
10	6 156 000	81	0.914
15	4 155 300	81	0.913
30	2 154 600	81	0.894

Table 4.1: NASDAQ-100 key statistics for different aggregation levels.

We demand for liquidity the following two requirements:

Requirement 4.1 (Liquidity)

L1 Liquidity for every single stock chosen has to be at least 0.75.

L2 Average liquidity over all stocks chosen has to be at least 0.85.

We will adopt both requirements L1 and L2 for all data sets. For the NASDAQ-100 data set we chose an aggregation level of 2 minutes, which amounts to an average liquidity of 0.858. Our final data set NASDAQ-100 consists of 71 stocks between 2007-06-27 and 2015-01-16 with 2 minute continuous returns and a total of 26 440 400 observations.

NASDAQ Intraday Data from 2007-06-27 to 2014-11-28

The raw NASDAQ data set is comprised of intraday returns of 1618 stocks for the trading days between 2007-06-27 and 2014-11-28. We will analyze the raw NASDAQ data set in the same vein as the we did with the raw NASDAQ-100 data set in subsection 4.3.1.

In Figure 4.5 we depict the portion of days that have at least two returns.

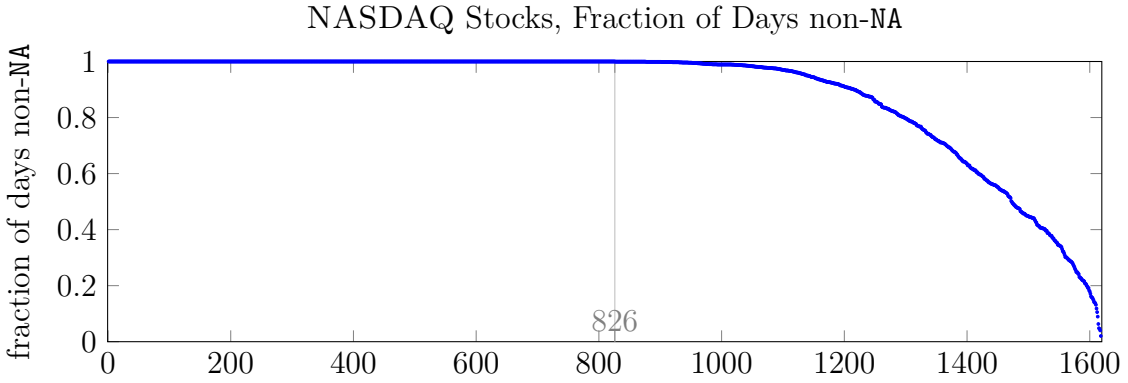

Figure 4.5: 1618 NASDAQ stocks, portion of days non-NA and with at least two non-0 returns in trading days between 2007-06-27 and 2014-11-28.

As a first step, we restrict the data set to all 826 stocks that have at least two returns every trading day. As before, we analyze average liquidity over all trading days and intraday. The former is depicted in Figure 4.6.

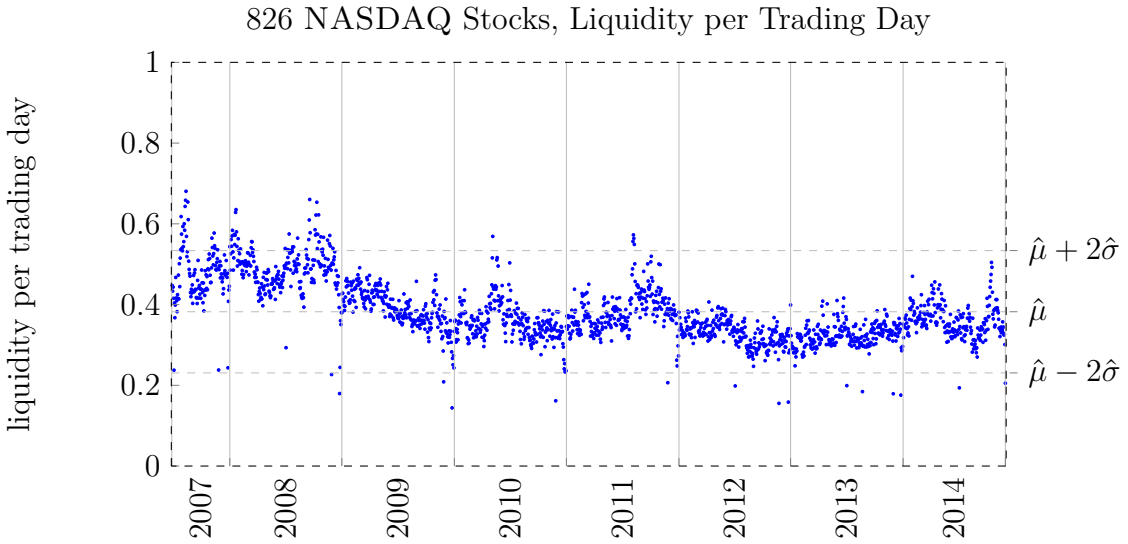

Figure 4.6: 826 NASDAQ stocks, fraction of non-zero returns per trading day between 2007-06-27 and 2014-11-28.

On average, a fraction of $\hat{\mu} = 0.382$ returns is non-zero with an estimated standard deviation of $\hat{\sigma} = 0.076$. We note some slight seasonality pattern at the change of

year, but not as pronounced as in Figure 4.2. The liquidity statistic also remains approximately constant in time, but with a lower estimated liquidity average of $\hat{\mu} = 0.382$.

Intraday liquidity is shown in Figure 4.7.

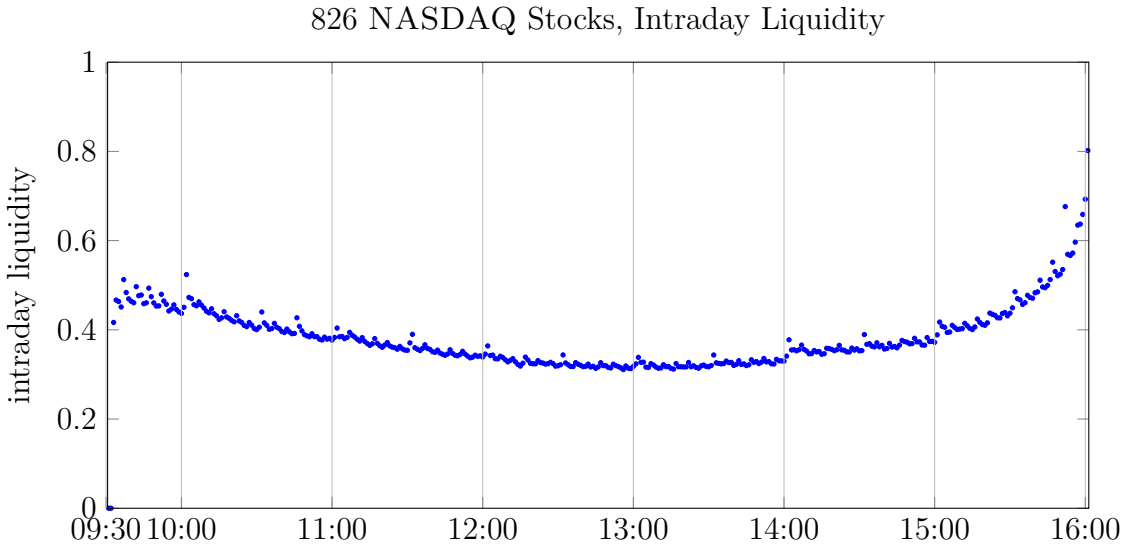

Figure 4.7: 826 NASDAQ stocks, fraction of non-zero returns per stock intraday between 2007-06-27 and 2014-11-28.

The pattern of NASDAQ is similar to NASDAQ-100 depicted in Figure 4.3, but without such a distinct »smile« shape. We also observe an increase in liquidity shortly before trading end. Overall, intraday liquidity is in good approximation constant.

The fraction of non-zero returns of $\hat{\mu} = 0.382$ is too low, so we need to aggregate intraday returns again.

As before, Figure 4.8 confirms our believe that liquidity is increasing with intraday time aggregation of returns. In comparison with the NASDAQ-100 data set the liquidity statistic remains lower. This should not surprise as the NASDAQ-100 stocks comprise the largest companies, but the largest companies are evidently also traded most.

To determine which stocks to include in our final data set we employ requirements L1 and L2, i.e. that every chosen stock has a liquidity of at least 0.75 and the average liquidity is at least 0.85. From all aggregation level and number of stock combinations we want to maximize available observations.

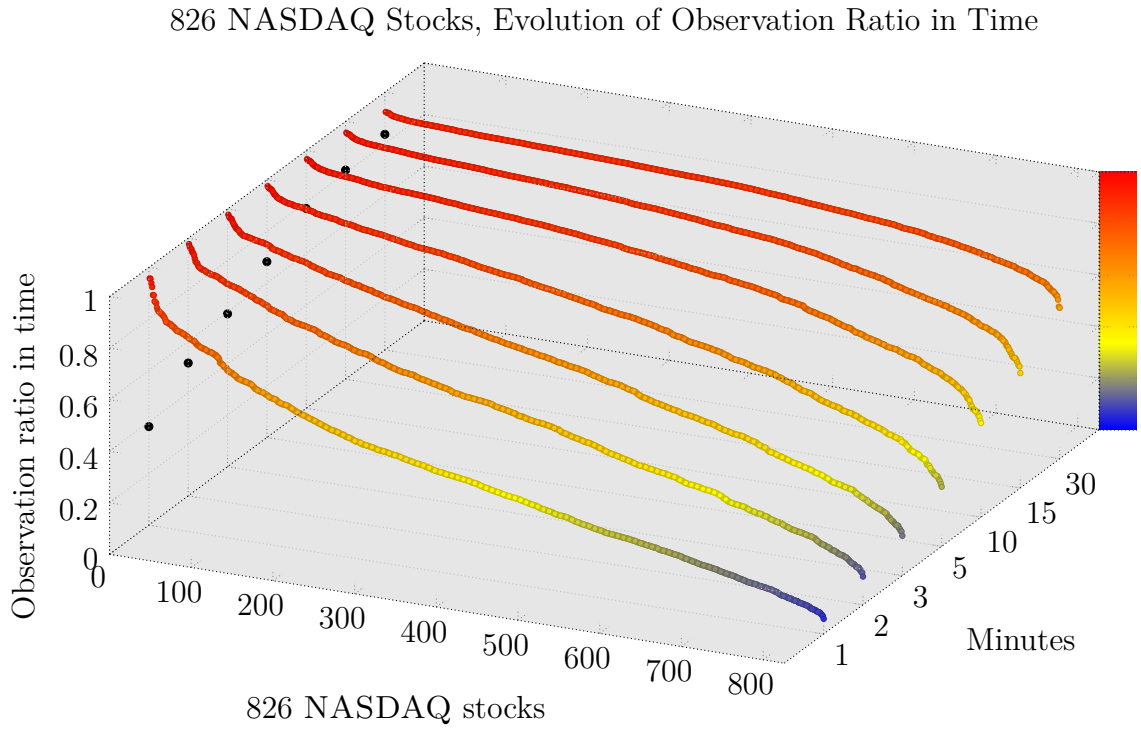

Figure 4.8: 826 NASDAQ stocks, change of liquidity per stock for aggregated returns, sorted. In black the average for each series.

Time [min]	Number of Observations	Number of Stocks	Liquidity
1	35 039 856	48	0.813
2	43 545 908	119	0.837
3	49 893 708	204	0.837
5	51 475 057	349	0.842
10	**40 924 640**	**548**	**0.855**
15	31 051 944	616	0.863
30	19 002 326	727	0.855

Table 4.2: NASDAQ key statistics for different aggregation levels.

Only aggregation levels of at least 10 minutes time intervals fulfill the two requirements L1 and L2 combined.

The NSADAQ-100 data set already contains very many observations in time. To provide more variety, we therefore want the NASDAQ data set to have a large number of stocks compared to the number of intraday observations. Aggregation levels of 10 or 15 minutes seem appropriate. Noting that the number of observations increases by 31.4 % when going from aggregation level 15 minutes back to 10 minutes, compared to the 11 % decrease in number of stocks, this seems like a decent trade-off.

We will therefore use the NASDAQ data set with an aggregation level of 10 minute returns, 548 stocks and a total of 40 924 640 observations.

4.3.2 S&P 500 and DJIA Intraday Data

S&P 500 Intraday Data from 2002-12-30 to 2016-09-16

The raw S&P 500 data set encompasses 502 stocks with intraday returns for the trading days between 2002-12-30 and 2016-09-16. We will analyze the raw S&P 500 data set in the same vein as we did with the previous ones.

First, we show the summary liquidity statistics for every of the 502 stocks in Figure 4.9.

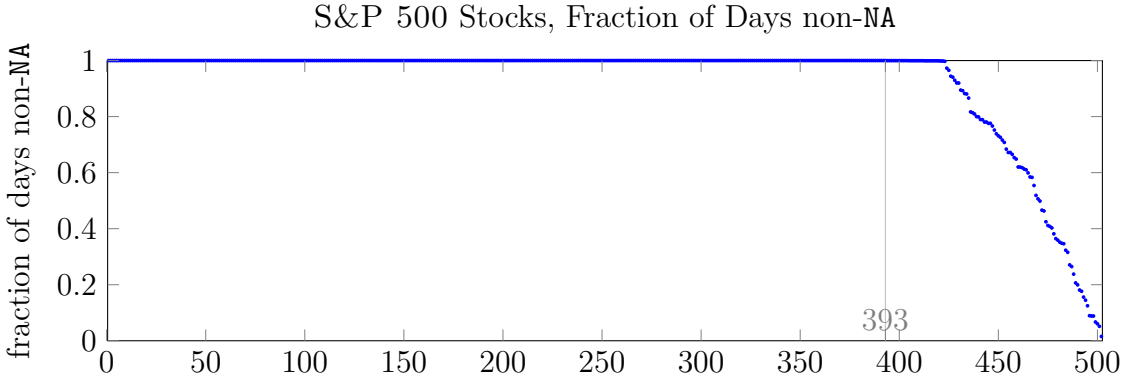

Figure 4.9: 502 S&P 500 stocks, portion of days non-NA and with at least two non-0 returns in trading days between 2002-12-30 and 2016-09-16.

Analog to subsection 4.3.1, we only use stocks that have at least two returns at every single trading day between 2002-12-30 and 2016-09-16. There are 393 stocks that fulfill this requirement. We look at average liquidity over all trading days and intraday, which is plotted in Figure 4.10 resp. Figure 4.11.

In contrast to the two NASDAQ data sets which each start 2007-06-27, this S&P 500 data set goes back to 2002-12-30. One can see a pronounced increase in average liquidity, starting in the year 2003 at about 0.4 and steadily increasing up to around 0.7 right about the 2007–2008 financial crisis. The short time periods of high average liquidity roughly correspond to times of highest TED spread[12], i.e. ICE LIBOR[13] over T-bills[14]. Incidentally, the two minor liquidity spikes around June 2010 and

[12] The acronym »TED« stands for »Treasury Bill Eurodollar Difference«. »TED spread« is the difference between ICE LIBOR[13] and T-bill[14] interest rate, both for 3 month.

[13] Intercontinental Exchange LIBOR, where »LIBOR« is an acronym for »London Interbank Offered Rate«.

[14] »T-bills«, long »Treasury bills«, are short term U.S. government debt obligations with a maturity of less than a year.

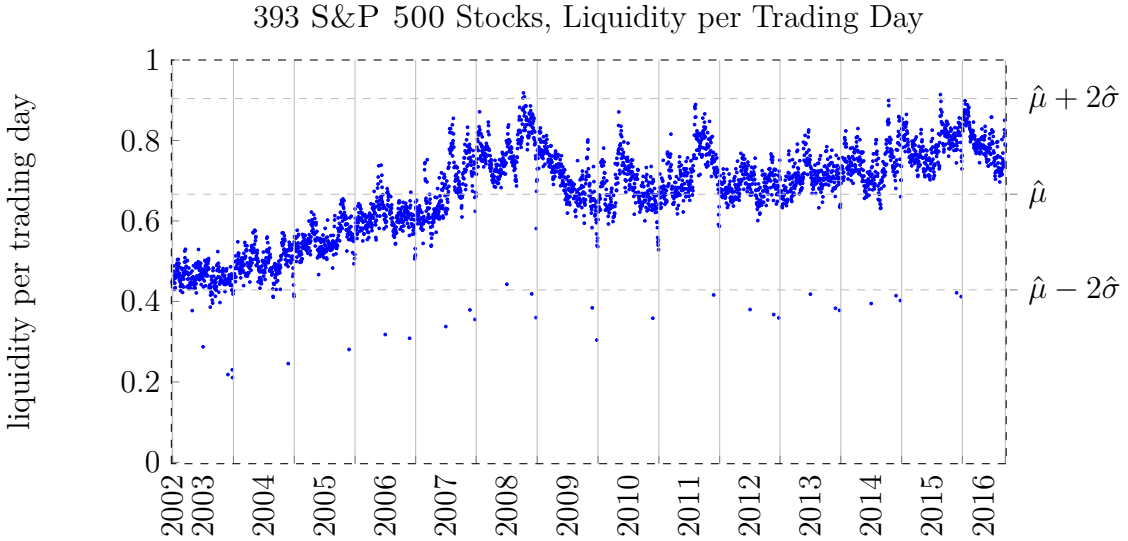

Figure 4.10: 393 S&P 500 stocks, fraction of non-zero returns per trading day between 2002-12-30 and 2016-09-16.

October 2011 correspond to increasing TED spreads.[15]

On average, a fraction of $\hat{\mu} = 0.666$ returns is non-zero with an estimated standard deviation of $\hat{\sigma} = 0.119$. Part of the reason behind the distinct higher estimated standard deviation, compared to the corresponding estimates for the NASDAQ-100 and NASDAQ, is the visible violation of mean stationary in the first part of the time period roughly between 2003–2007. In addition, from visual inspection there seems to be a distinct seasonal pattern of outliers with roughly half the liquidity compared to other nearby trading days. These outliers were also present in the both previous data sets for the NASDAQ, but are more pronounced here. By closer inspection the postulated fixed seasonal pattern can be confirmed, almost all low liquidity outliers are either on Black Friday[16], during the Christmas holidays and sometimes the day before Independence Day[17]. This also explains the seasonal low liquidity outliers present in both NASDAQ data sets. Overall, their contribution is negligible.

Intraday liquidity for the S&P 500 is shown in Figure 4.11.

[15] `https://research.stlouisfed.org/fred2/series/TEDRATE`, best seen with log scale.
[16] Black Friday is the day after Thanksgiving Day in the United states, which in turn is every fourth Thursday of November and a federal holiday.
[17] Independence Day is a federal holiday in the United States, celebrated every July 4$^{\text{th}}$.

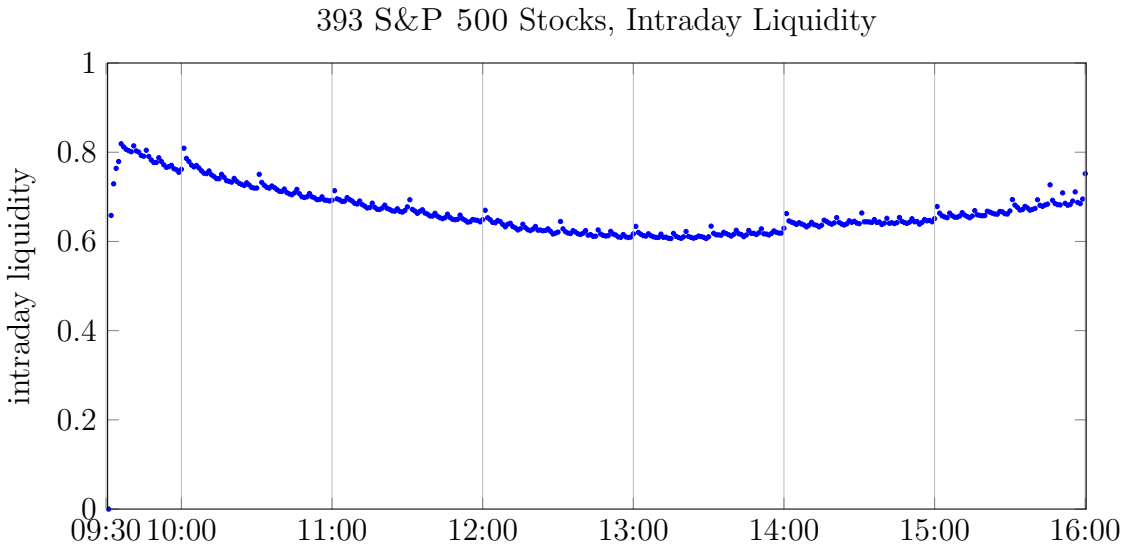

Figure 4.11: 393 S&P 500 stocks, fraction of non-zero returns per stock intraday between 2002-12-30 and 2016-09-16.

The liquidity »smile« is visible again, this time with slightly higher liquidity at the beginning of the trading day. Looking at singly stock liquidity subject to different aggregation levels in Figure 4.12, we observe that many stocks in the S&P 500 have a relatively high average liquidity.

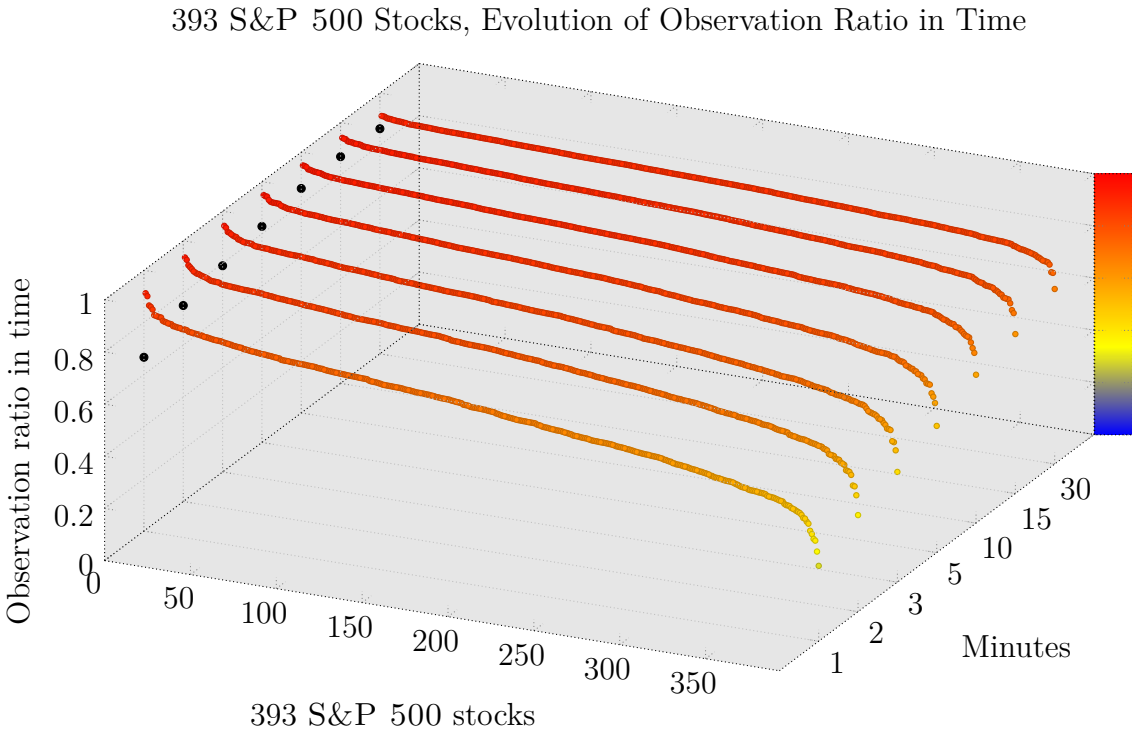

Figure 4.12: 393 S&P500 stocks, change of liquidity per stock for aggregated returns, sorted. In black the average for each series.

To determine which stocks to include in our final S&P 500 data set we again call for requirements L1 and L2 to hold. A summary is given in Table 4.3.

Time [min]	Number of Observations	Number of Stocks	Liquidity
1	99 938 036	74	0.789
2	**151 644 416**	**224**	**0.811**
3	130 312 512	288	0.828
5	94 138 770	345	0.846
10	**52 224 480**	**378**	**0.867**
15	35 904 330	385	0.873
30	18 858 840	390	0.864

Table 4.3: S&P 500 key statistics for different aggregation levels.

Intraday returns with aggregation levels of 10 minutes or more fulfill the two requirements L1 and L2. As maximization of number of observations is the secondary goal, we choose 10 minutes returns for the S&P 500 data set.

The final S&P 500 data sets consists of 3454 trading days between 2002-12-30 and 2016-09-16 with 10 minute returns, 378 stocks and 52 224 480 observation in total.

In addition, we also want to utilize the S&P 500 raw data to construct a second data set with very high number of observations and modest average liquidity. Because the observation count is highest for 2 minute returns, we choose those. This implies a data set with 224 stocks and a total of 151 644 416 observations. We will call the two data sets from the subsection the »long« S&P 500 respectively the »big« S&P 500 data set.

DJIA Intraday Data from 2002-12-30 to 2016-09-16

The raw DJIA data set encompasses 30 stocks with intraday returns for the trading days between 2002-12-30 and 2016-09-16. We will analyze the raw DJIA data set in the same vein as the we did with previous ones.

First, we show the summary liquidity statistics for every of the 30 stocks in Figure 4.13.

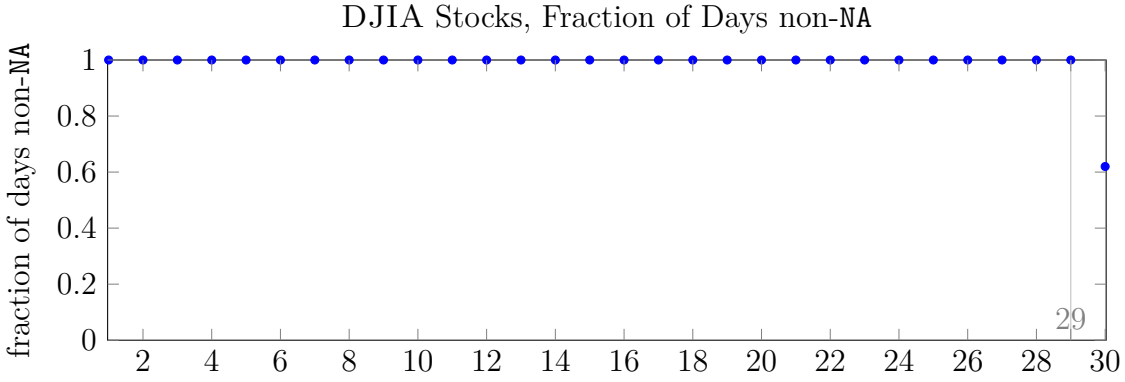

Figure 4.13: 30 DJIA stocks, portion of days non-NA and with at least two non-0 returns in trading days between 2002-12-30 and 2016-09-16.

As we mentioned in subsection 4.2.2, the DJIA data set has all stocks but Apple Inc. and Visa Inc. with at least two returns at every trading day between 2002-12-30 and 2016-09-16, so we analyze the liquidity statistics of the remaining 28 stocks. Average liquidity over days is depicted in Figure 4.14.

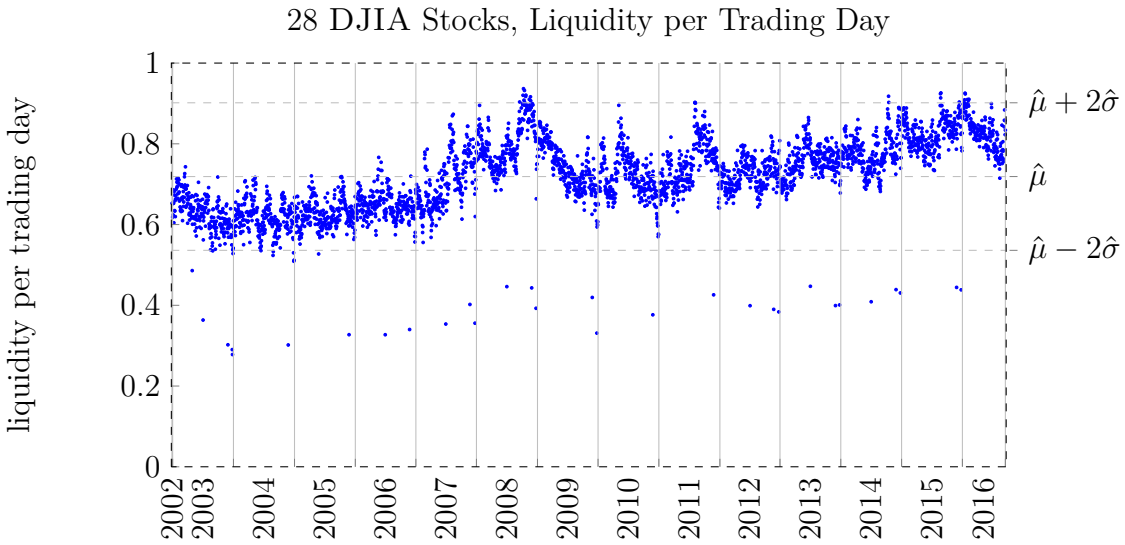

Figure 4.14: 28 DJIA stocks, fraction of non-zero returns per trading day between 2002-12-30 and 2016-09-16.

As the DJIA encompasses 30 of the largest U.S. publicly held companies and large companies tend to be more liquid, there should be no surprise that average liquidity is highest for the DJIA with a fraction of $\hat{\mu} = 0.719$ returns non-zero. The estimated standard deviation of $\hat{\sigma} = 0.091$ is again, as it was the case with the S&P 500 data set, considerably higher than for the much shorter data sets NASDAQ-100 and NASDAQ. By visual inspection we also see a slight increase in liquidity during the first years, but not as pronounced as for the S&P 500. The holiday pattern is also visible as distinct low liquidity outliers. In fact, empirical correlation of the respective average liquidity series for the S&P 500 and DJIA is 0.939 and for the intraday liquidity between the two it is 0.988. On the other hand, the empirical correlation

between, for instance, the intraday liquidity of NASDAQ and of S&P 500 is only about 0.699. This high empirical correlation between S&P 500 and DJIA is also explainable by the fact that all companies from the DJIA are included in the S&P 500. So one presumably only sees a catch-up of liquidity of the smaller S&P 500 companies to the level of the big DJIA ones.

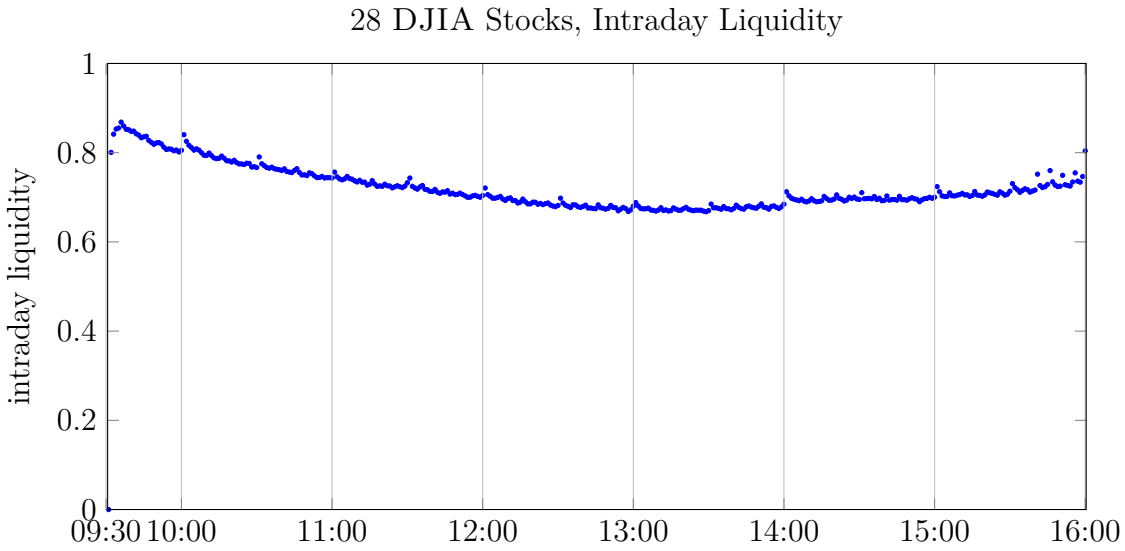

Figure 4.15: 28 DJIA stocks, fraction of non-zero returns per stock intraday between 2002-12-30 and 2016-09-16.

As it was the case with the previous depictions of intraday liquidity, there is a discernible »smile« present. Because average liquidity is comparably high for the DJIA data set, we expect this to be true for different aggregation levels, too. This is displayed in Figure 4.16.

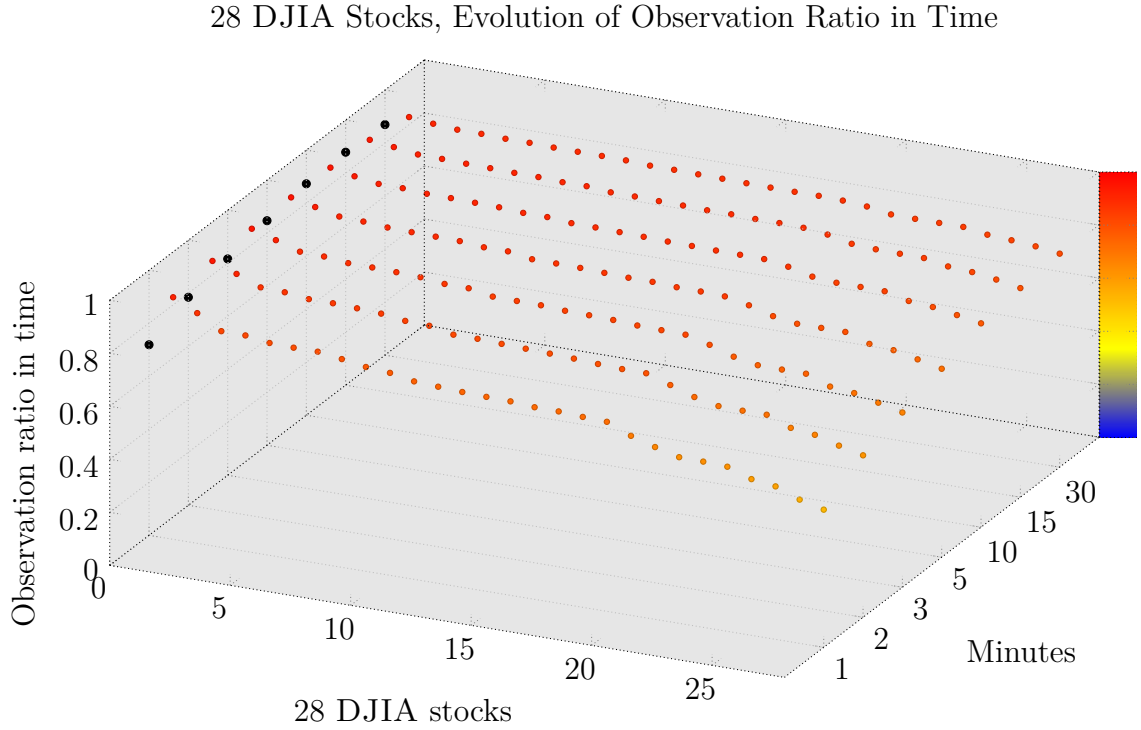

Figure 4.16: 28 DJIA stocks, change of liquidity per stock for aggregated returns, sorted. In black the average for each series.

From Table 4.4 we see that requirements L1 and L2 are fulfilled for all aggregation levels of 5 minutes and more.

Time [min]	Number of Observations	Number of Stocks	Liquidity
1	13 505 140	10	0.813
2	13 539 680	20	0.828
3	9 501 954	21	0.850
5	**7 094 516**	**26**	**0.857**
10	3 868 480	28	0.873
15	2 611 224	28	0.879
30	1 353 968	28	0.870

Table 4.4: DJIA key statistics for different aggregation levels.

We chose according to our secondary goal of observation maximization. The DJIA data set consists of 5 minute returns on 3454 trading days between 2002-12-30 and 2016-09-16 with 26 stocks and a total number of 7 094 516 observations.

4.4 Empirical Spectral Distributions

4.4.1 Motivation

In the previous section 4.3 we gave descriptive statistics of the raw data sets that helped us to choose the final data sets. For every of the final data sets, one can estimate various descriptive statistics. We are interested in the dynamics of multiple stocks in time and between themselves. In addition, we derived parametric families of spectral distributions for covariance matrix estimators in chapter 3. Thus, our main object of study are the empirical spectral distributions of covariance matrix estimates of the data sets.

The ESD encompasses the spectral information of a quadratic and symmetric real-valued matrix, but neglects the information present in its eigenvectors. Any quadratic symmetric real-valued matrix of size $(N \times N)$ can be completely parametrized by its N main diagonal values and its $2^{-1}N(N-1)$ independent values on the lower or upper triangular matrix. Those $2^{-1}N(N+1)$ degrees of freedom (d.o.f.) do not vanish in the spectral representation of a matrix, $2^{-1}N(N-1)$ d.o.f. are hidden in the N eigenvectors, each with N components. Only N d.o.f. are displayed as N eigenvalues.

One can discard the N eigenvectors for rotational invariant matrix ensembles, which are already completely parametrized by its N eigenvalues. But even for non-rotational invariant matrix ensemble it can be a sensible choice to disregard $2^{-1}N(N-1)$ d.o.f. and concentrate only on the eigenvalues or there spectral distribution. The former step reduce the parametrization from $\mathcal{O}(N^2)$ to $\mathcal{O}(N)$. If one could parametrize the empirical spectral distribution with a fixed number of parameters, the parametrization would be of asymptotically optimal size of $\mathcal{O}(1)$. Since every model is an idealized and thus simplified image of reality, one part of modelling is to reduce dimensions. But this compression has to be done in a sensible way to conserve the main characteristics. This perspective motivates the concentration on only the ESD.

Remark: *There are few but promissing works that go beyond the analysis of the spectral distribution and investigate the behavior of eigenvectors of random matrices. In the context of the four moment theorem and central limit theorems* universal behavior *for these eigenvectors was proven 2013 by Knowles and Yin in [55] and 2012 by Tao and Vu in [100]. In 2013, Xia et al. proved convergence rates of eigenvector empirical spectral distribution of large dimensional sample covariance matrix in [119]. In a series of papers 2011, 2012, 2014, Allez and Bouchaud established eigenvector dynamics of certain random matrix models in [2, 3, 4]. In 2014 resp. 2016, Allez, Bun, et al. resp. Bun et al. extend the results in [5] resp. [21].*

Bordenave and Guionnet analized in 2013 the eigenvector behavior of matrices with heavy-tailed entries.

The earlist works on eigenvectors of non-hermitian matrices are in a series of papers

[65, 27, 66] from 1998 to 2000 by Mehlig and Chalker, followed 1999 by Janik et al. in [54]. 2015, Rudelson and Vershynin proved in [85] that random matrices with independent entries habe completely delocalized eigenvalues, i.e. with high probability the coordinates of unit eigenvectors are of magnitude $\mathcal{O}\left(N^{-1/2}\right)$*, modulo logarithmic corrections. Burda, Grela, et al. showed 2015 in [25] a connection of eigenvector and so-called Burger dynamics.*

In [79], one can find a recent survey on eigenvectors of random matrices by O'Rourke et al.

4.4.2 Market Eigenvalue

To facilitate comparability of the ESD of covariance matrix estimators, we normalize to correlation matrix estimators. Thus, the estimated correlation matrices have normed variances of 1 on their main diagonal. Due to the invariance of the trace to the sum of the main diagonal the real eigenvalues also have a sum equal to the number of securities. This normalization is generally applied in the literature when studying ESDs of covariance matrix estimators, so we comply. But note that due to the varying sum of eigenvalues for correlation matrices of different size comparisons are not as easy as with normed total sum, e.g. to unity.

It is a well known characteristic[18] that the ESD of correlation matrix estimators for stock returns have a largest eigenvalue that is much bigger than all the other eigenvalues. For the NASDAQ-100 data set with daily returns there are 1900 trading days. For each trading day we estimated the correlation matrix. The largest eigenvalue λ_1 is at least 1.3 times higher than the second largest eigenvalue λ_2 — but this ratio has a maximal value of 29.6, a mean of 5.6 and a median of 5. In comparison, the ratio λ_2/λ_3 over all trading days has a minimum of 1.006, a maximum of 3, a mean of 1.4 and a median of 1.4. Averaged over all trading days, the ratio of consecutive eigenvalues λ_k/λ_{k+1} is a decreasing function in its argument k until $k = 14$, after that it is still globally decreasing, albeit not monotonically anymore. This characteristics hold true for all the other data sets and the correlation matrices estimated daily.

For daily, monthly and quarterly estimated periods for the correlation matrices, the consecutive eigenvalue ratio function has globally more of a »U«-shape. Nevertheless, the largest eigenvalue is still, in average, 3.3 to 11.7 times larger than the second largest eigenvalue. This ratio is, by far, the largest of all consecutive eigenvalue ratios for all data sets and estimation periods.

In the literature, because of this pronounced characteristics the largest eigenvalue is called the »market eigenvalue«. This name is motivated by the observation that the associated eigenvector $\mathbf{v}_1$ to the largest eigenvalue λ_1 has its components distributed nearly uniformly.

[18] See e.g. [17, p. 148].

Remark: *Eigenvectors give directional information, so only the relative magnitude of the components is of interest, but not their absolute values. Normed to unity in L^1-norm, any eigenvector can be interpreted as a portfolio weight vector. For all components of the eigenvector non-negative*[19]*, the associated portfolio weight vector represents a full investment, i.e. it sums to unity.*[20]

As eigenvectors span the associated eigenspace, depending on the dimension on the latter there might be degrees of freedom on how to choose eigenvectors, independent of their normalization. In practical correlation matrix estimation, almost always the eigenspaces are 1*-dimensional*[21]*. This leaves only* 1 *d.o.f. for the parametrization of the eigenvector, multiplication by* -1*. So the portfolio weight vector associated to an eigenvector contains no long/short information for individual stocks.*

Following the reasoning from the last paragraph, we choose a parametrization of the eigenvector associated to the largest eigenvalue in such a way that preferable most of its components are non-negative. As an empirical characteristic, the eigenvector components of the so-called »market eigenvector« are almost always all positive and nearly uniformly distributed. Therefore, the economic interpretation of this eigenvector is that of a so-called »market portfolio«, which is a portfolio including all assets in a market. The market portfolio is routinely assumed to be *long-only* and non-negative, because of theoretic motivation in financial models such as the capital asset pricing model (CAPM) from Treynor, Sharpe, Lintner and Mossin[22] or the market model from Sharpe[23].

4.4.3 Plots of Empirical Spectral Distributions

We plot the histogram of the average ESD of the correlation matrix estimators for the specified data sets over the estimation periods ranging from daily, to weekly, monthly and quarterly. The market eigenvalue of the average ESD is between 2.9 and 11.5 times larger than the second largest eigenvalue. To ensure a distortion-free plot of the histogram only an excerpt is shown. The histogram over all eigenvalues is shown as inlay. Note that the bin size in the inlay is different from the main plot due to visibility constraints. Every single eigenvalue is marked underneath each histograms.

All the ESDs are non-negativly supported and have a long right tail. In most histograms there exists a clear »bulk« of eigenvalues and therefrom separated some single eigenvalues. As the sample period size increases from days to weeks, month and

[19] Negative portfolio weights correspond to a *short* investment. For a *short* investment the investor sells a security he does not currently own with the goal to buy it back latter to a lower price.
[20] Henceforth, we speak of eigenvectors and their associated portfolio weight vectors interchangeably.
[21] Note that for invertible matrices the algebraic and geometric multiplicity of eigenvalues coincides.
[22] The CAPM was developed independently in a series of papers [103, 104, 91, 59, 71] between 1961 and 1966.
[23] See [90].

quarters, the separation into bulk and isolated eigenvalues becomes more pronounced, see Figure 4.17 – Figure 4.21.

Depending on the granularity of intraday returns, for daily or even weekly sample period size there might be more stocks than observations in the sample period. By construction, such a correlation matrix estimate has eigenvalues of value 0 with frequency given by the difference of number of stocks and number of available observations. This situations happens for the NASDAQ and days or weeks of sample period size, for the S&P 500 with 2 minute returns and days of sample period size and for the S&P 500 with 10 minute returns and days or weeks of sample period size. For all those mentioned, the histogram only show plots truncated in such a way that the remaining eigenvalue density is visible.

The histograms for the NASDAQ and S&P 500 data sets with 10 minute returns and sample period size of one day show very spread out eigenvalues. With 10 minute returns and 6.5 trading hours there are only 39 observations per trading day. For a high number of stocks like 548 resp. 52 224 480, for the NASDAQ resp. S&P 500 data set, there is a high share of 0 eigenvalues by construction. All the remaining 39 eigenvalues are spread out so that sum constrained induced by normalization to a correlation matrix is fulfilled.

Averaged over all observations, the largest eigenvalue accounts for a portion of 24.8 % to 36.1 % of the eigenvalue sum, depending on the specific data set and length of sample period. Over all data sets and length of sample periods, the minimal share of the largest eigenvalue of the sum of all eigenvalues is 6.5 %, the maximal share is 83.2 %.

There is a clear resemblance of the ESD implied by the bulk eigenvalues of the previous histograms and the empirical spectral distributions from section 3.1 and section 3.2. We investigate this phenomenon thoroughly in the next chapter.

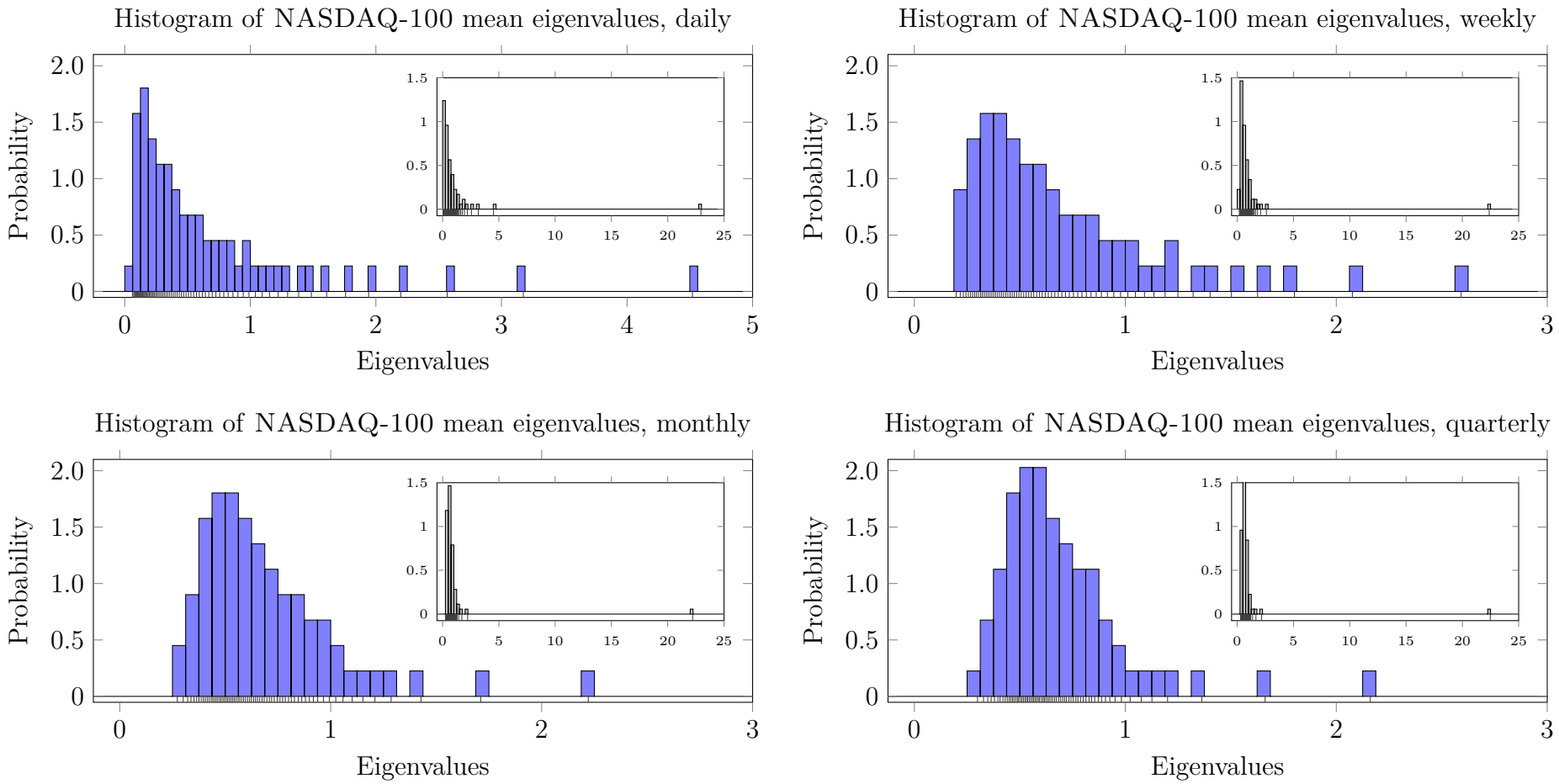

Figure 4.17: Histograms of averaged eigenvalues of the NASDAQ-100.

NASDAQ

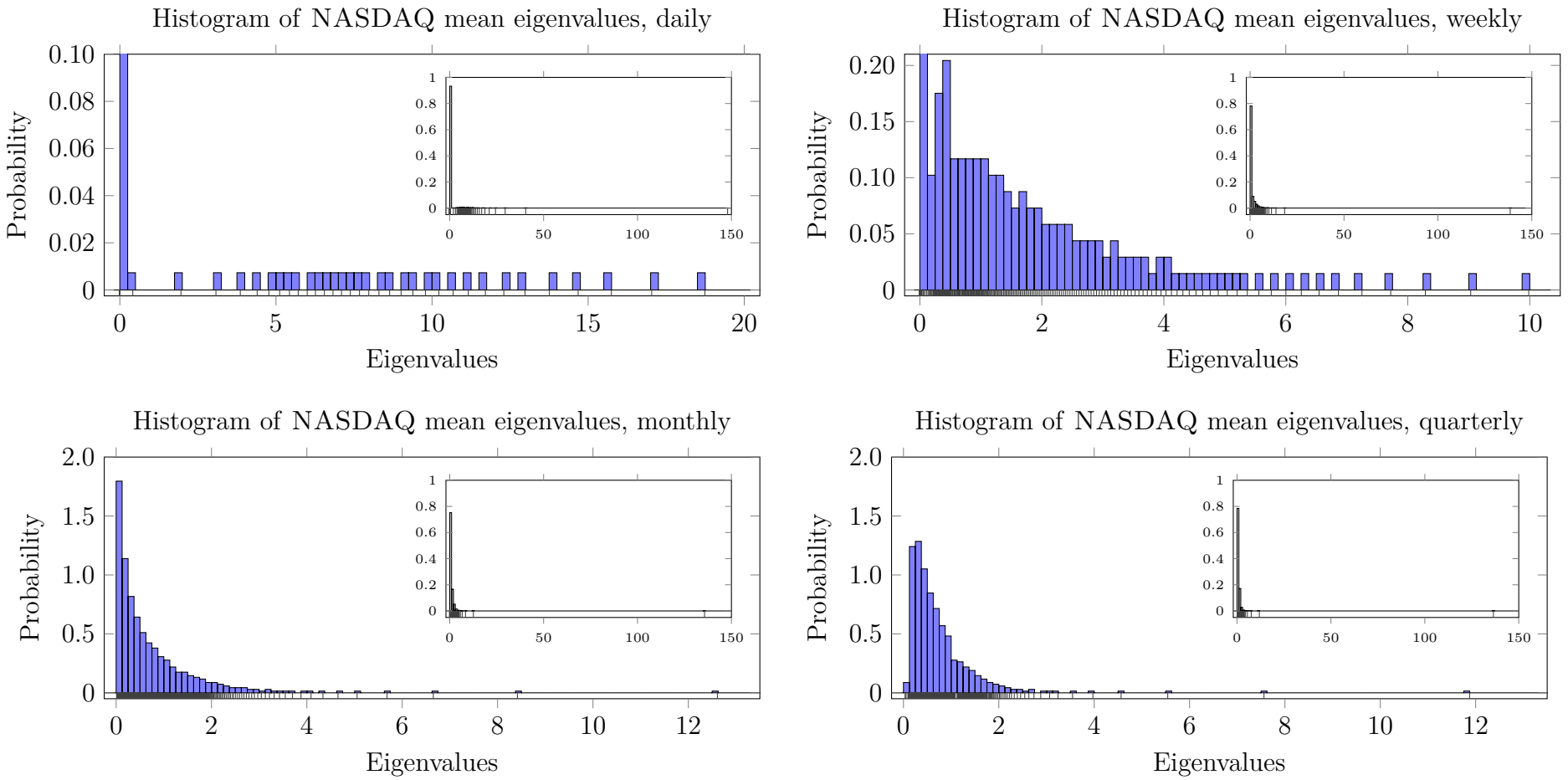

Figure 4.18: Histograms of averaged eigenvalues of the NASDAQ.

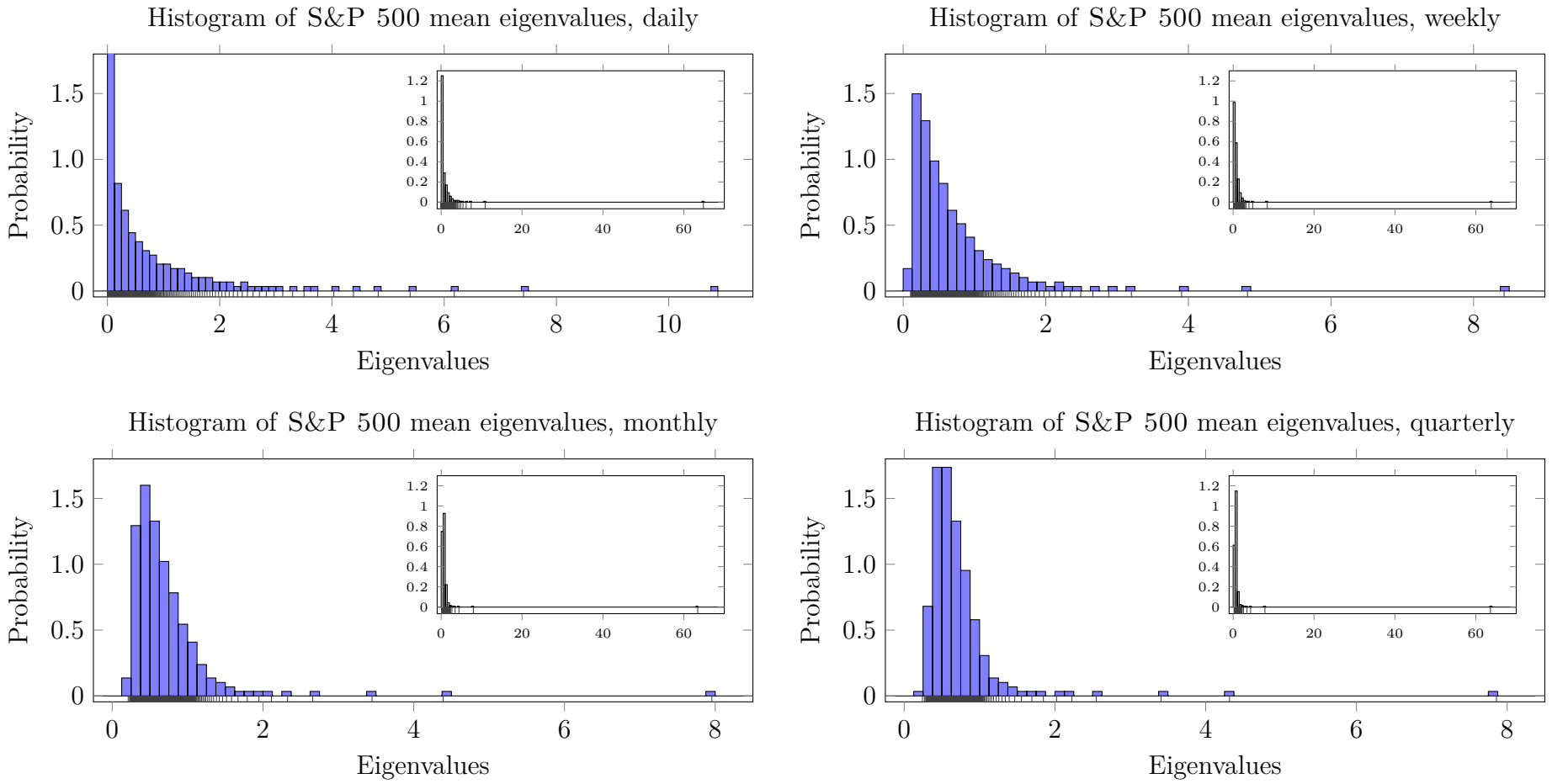

Figure 4.19: Histograms of averaged eigenvalues of the S&P 500.

S&P 500 with 10 **minute returns**

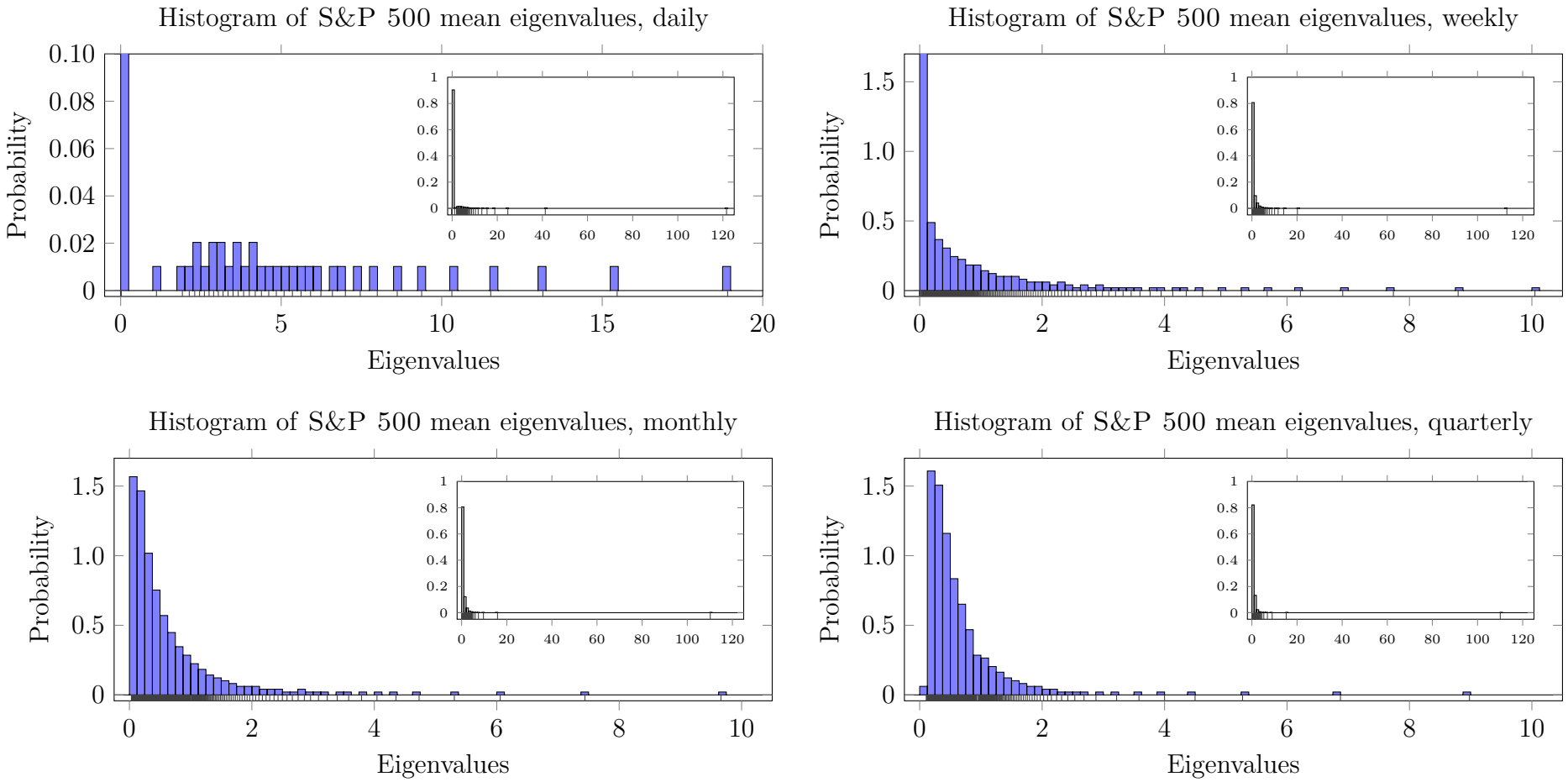

Figure 4.20: Histograms of averaged eigenvalues of the S&P 500.

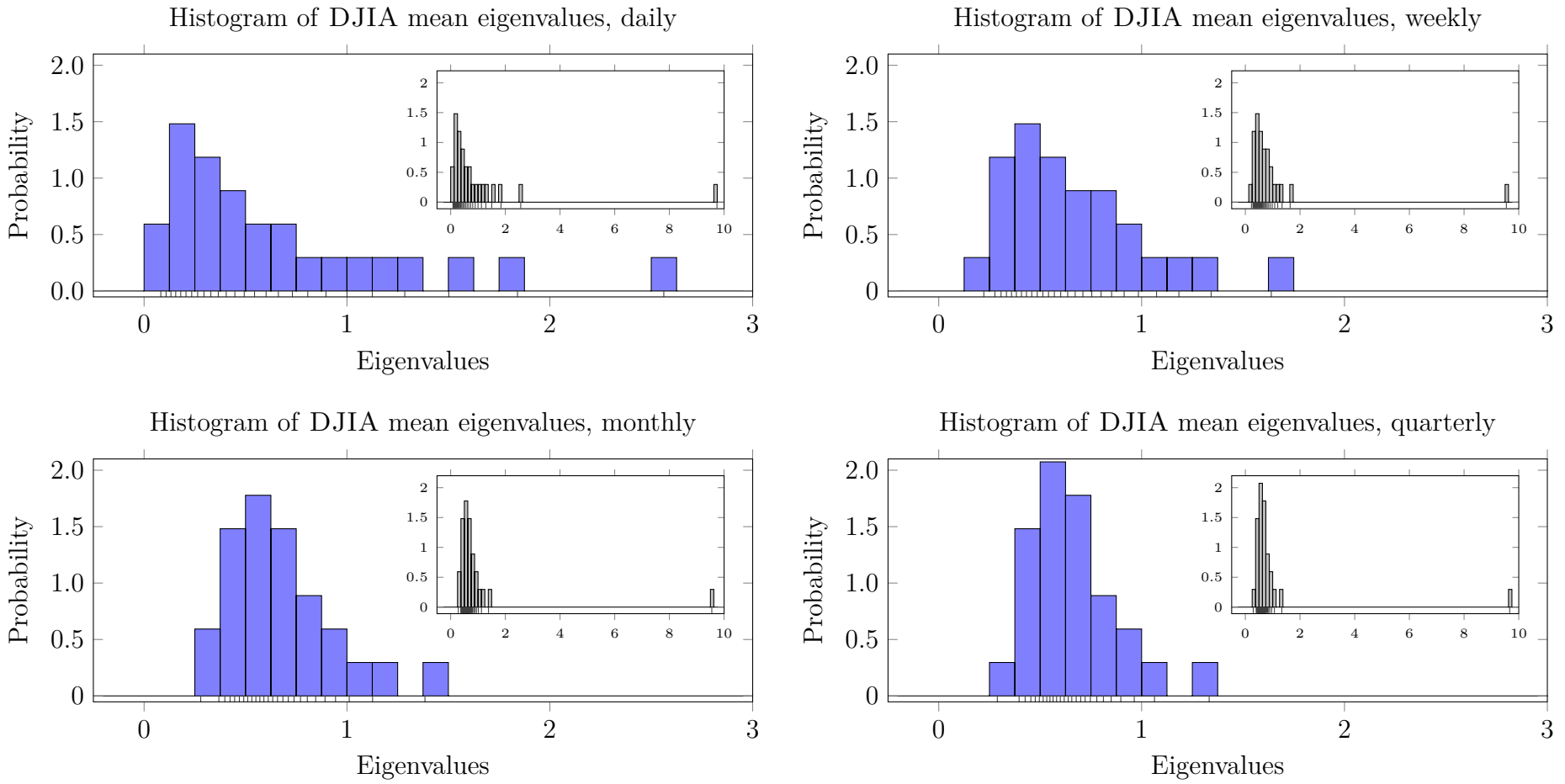

Figure 4.21: Histograms of averaged eigenvalues of the DJIA.

5 Application of Specific Random Matrix Models to Financial Returns

We are interested in a class of preferably parsimonious models that capture and describe the main dynamics of financial stock markets. As explained in the previous chapter, we restrict ourselves to stock returns and the data sets created in section 4.3. In the first part of this work, we layed the foundation of non-commutative probability theory and utilized it to introduce basic random matrix models. These widely-used models are quite general due to the phenomenon of universality. There are many works that utilize GOE matrices and the implied Marchenko–Pastur spectral distribution of their covariance matrix estimators and successfully apply those results to financial data.

A structural caveat of those models is the required and thus assumed independence[1] of all matrix entries, modulo some symmetry constraints. Thus one of the main points of interest, the inter-dependence of stock returns in time and between different stocks, is assumed non-existent. It seems therefore natural to relax the very strict assumptions of independent stock returns and allow for some dependence structure. As we argued in chapter 3, we allow for non-trivial autocovariance matrices and a dependency captured by VARMA(p, q) resp. VARFIMA(p, d, q).

Specifically, we model the evolution of the stock returns as stochastic VARMA(p, q) processes for $\{p, q\}$ small. We call an N-dimensional stochastic VARMA(p, q) process of length T an $(T \times N)$ VARMA(p, q) matrix. In chapter 3, we presented the spectral distribution of covariance matrix estimators of VARMA(p, q) matrices. Analogous, we speak of N-dimensional stochastic VARFIMA(p, d, q) processes and of VARFIMA(p, d, q) matrices.

To assess on how well the theoretical models describe reality, we determine the structural parameters of each data set and estimate the model parameters. The structural and estimated parameters can be plugged back into the model to yield a specific instance of the parametric spectral distribution function. For each data set, the specific shapes implied by estimated parameters can be compared to the empirical spectral distribution of the covariane matrix estimator. We say that a model is well suited, if the theoretical spectral distribution captures the main shape and characteristics of the real data. In addition, one has to perform robustness checks to ensure that the model does not only describe reality well due to overfitting.

[1] In the case of all entries Gaussian uncorrelatedness suffices, because for Gaussians independence and uncorrelatedness are equivalent.

Ideally, one would like that the model results do not depend on certain optimization techniques or only hold for specific data sets. After all, a model should describe a restricted part of reality. The various data sets are invariants insofar as any reasonable model should be applicable with satisfying results to almost all of them.

5.1 Parameter Estimation Methodology

If, by assumption, the vector-valued return process of the stock market under investigation follows a VARMA(p, q) or VARFIMA(p, d, q) processes, these stochastic processes imply a specific empirical spectral distribution of the associated covariance matrix estimator. In section 3.1 and section 3.2 we derived methods on how to obtain said ESD. This involved solving for a specific root of a polynomial with the polynomial coefficients themselves mixed polynomials in the model parameters with high[2] degree.

Based on the observed data used to estimate the covariance matrix, the rectangularity ratio α can be calculated as the ratio of stocks to observations. One even does not need specific data, it suffices to know the matrix size. In contrast, the AR, MA and, if applicable, fractional integration parameters $\{\theta_1, \ldots, \theta_p, \phi_1, \ldots, \phi_q, d\}$ of the assumed VARMA(p, q) or VARFIMA(p, d, q) processes have to be estimated, i.e. for every possible data set from the sample space there has to be a mapping yielding the sample estimate.

We denote the rectangularity ratio α as *structural* parameter, because its value can be derived a priori to observing the stock market dynamics by *meta data* about number of stocks in a market and number of returns in a trading day, week, month or quarter to come.[3] We denote all the other parameters *model parameters*, as they are implied by the specific model chosen.

As presented in section 3.1 and section 3.2, only the most basic models AR(1) and MA(1) have a distribution generating polynomial that is of degree 4 and, in principle, solvable in analytic form. Though, as shown, the closed form solutions are so complicated that a utilization of the analytical solution by, for instance, *method of moments* or *estimation equations* does not seem reasonable, let alone feasible. We employ methods similar to *maximum likelihood type estimators*, which can be subsumed under a class which Huber introduced 1964 and called *M-estimators*[4]. Parameter estimation is done by minimization of a sum of functions g, each depending on the parameter vector $\boldsymbol{\theta}$ and data $x_i \in \mathcal{X}$:

$$\hat{\boldsymbol{\theta}} \coloneqq \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \sum_{i \in \mathcal{I}} g(x_i, \boldsymbol{\theta}), \qquad (\forall i \in \mathcal{I} : x_i \in \mathcal{X}). \tag{5.1}$$

[2] In the context of distribution generating polynomials, every degree greater than 4 is considered high as these polynomials only permits approximate numerical solutions.

[3] Obviously *meta data* is also data in itself and can only be known for certain when finally realized and observed.

[4] See [45, p. 74] and also [44, pp. 43–55, Section 3.2].

M-estimators include a large class of estimators, for instance least-squares estimators. This is a very convenient type of estimators widely used, because minimization of squared residuals is equivalent to norm minimization in a L^2 vector space. Every L^2 vector space can be endowed with a canonical scalar product that makes it into a Hilbert space, possibly after completion. The scalar product induces a norm, that is minimized — or equivalently, the squared norm. Note that L^2 vector spaces are the only ones which can be upgraded to Hilbert spaces.[5] Hilbert space norm minimization also has the nice geometrical interpretation of decomposing a vector into two orthogonal parts, which is roughly the statement of the Hilbert projection theorem.

In this context, data $\mathcal{X}$ is the collection of empirical spectral distributions of the correlation matrix estimators $\hat{\mathbf{C}}$. For each observed return data matrix $\mathbf{X}$ one gets a correlation matrix estimator $\hat{\mathbf{C}}(\mathbf{X})$. The ESD $\tilde{\mu}_{\hat{\mathbf{C}}(\mathbf{X}_k)}$ calculated from $\hat{\mathbf{C}}(\mathbf{X})$ can be denoted by x_k for k indexing a sample.

For a given model and a fixed parameter vector $\boldsymbol{\theta}$, one can obtain a numerical approximation of the density function $f(\boldsymbol{\theta})$. For real density functions from chapter 3, this approximation is done one a grid $(x_i)_i$, with $x_i \in \mathbb{R}$ and i from some index set $\mathcal{I}$. Therefore, the density function $f(\boldsymbol{\theta})$ is approximated by the argument-value pairs $(x_i, f(x_i, \boldsymbol{\theta}))_{i\in\mathcal{I}}$.

Remark: *Note that we do not differentiate between the true density function and its numerical approximation, because we do not know the true density function and its numerical approximation is all we have. In general, for the degree of the distribution generating polynomial greater than* 4*, there do only exist numerical solutions. So one could only differentiate between differing algorithms to approximate the true value, which we will not pursue.*

To be able to compare values of the density function $f(x_i, \boldsymbol{\theta})$ with values of the ESD $\tilde{\mu}_{\hat{\mathbf{C}}(\mathbf{X}_i)}$, the latter has to be aggregated. This is done by choosing a discretization of the domain in bins $[\tilde{x}_i, \tilde{x}_{i+1})$ and counting the number of eigenvalues per bin. For equidistant discretization this gives histograms like Figure 4.17 – Figure 4.21.

We choose a equidistant discretization $\mathbf{x} \coloneqq (x_i)_{i\in\mathcal{I}}$ and calculate bins $[\tilde{x}_i, \tilde{x}_{i+1})$, such that one has

$$x_i = 2^{-1}(\tilde{x}_i + \tilde{x}_{i+1}).^6 \tag{5.2}$$

In addition, we ensure that the discretization covers the whole domain of $\tilde{\mu}_{\hat{\mathbf{C}}(\mathbf{X}_k)}$. For calculations of a sample we fix a discretization suitable for all observations to facilitate intra-model comparability.

Depending on the data set and the maximal range $\max_k \tilde{\mu}_{\hat{\mathbf{C}}(\mathbf{X}_k)}$ of the domain of the

[5] And the other way round, every Hilbert space is isometric to a certain $\ell^2(S)$ space for a set S with appropriate cardinality.

[6] With slight abuse of notation we neglect that the cardinality of index set $\mathcal{I}$ for the bins has to be one bigger than the cardinality of the index set of grid points x_i.

ESD, a discretization of length between 2^9 and 2^{13} and granularity $|x_{i+1} - xi| = 2^{-4}$ is chosen. The relationship of bins and grid points implies that density function values $f(x_i, \boldsymbol{\theta})$ are calculated at the midpoints of the equidistant bins $[\tilde{x}_i, \tilde{x}_{i+1}]$. This coupling of bins and grid points facilitates calculation of residues, because one obtains

$$\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\theta}) \coloneqq f(\mathbf{x}, \boldsymbol{\theta}_t) - \tilde{\mu}_{\hat{\mathbf{C}}(\mathbf{X}_t)}(\mathbf{x}), \quad \text{with } t \in \mathbb{T}, \tag{5.3}$$

for some index set $\mathbb{T}$. The ESD $\tilde{\mu}_{\hat{\mathbf{C}}(\mathbf{X}_t)}(\mathbf{x})$ is the vector of normed histogram counts for the bins given by (5.2).

To illustrate the structure of (5.3), for the NASDAQ-100 data set with 2 minute returns and an aggregation level of one week, the return matrix $\mathbf{X}_t \in \mathbb{R}^{T_t \times N}$ has $N = 71$ stock columns. For 2 minute returns there are $6.5 \cdot 60 \cdot 2^{-1} = 195$ observations per trading day of 6.5 hours length and, depending on the number of trading days in week t, $T_t = 3 \cdot 195 = 585$, $T_t = 4 \cdot 195 = 780$ or $T_t = 5 \cdot 195 = 975$ observations per week. For the whole NASDAQ-100 data set, there are only 3 weeks of length 3 days, 69 weeks of length 4 days and a large majority of 323 weeks with 5 trading days. The elements t of the index set $\mathbb{T}$ enumerate all observed weeks in the data set.

In light of the motivation of Hilbert spaces we minimize the L^2-norm of the residual $\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\theta})$, which for fixed data $\mathbf{X}_t$ and fixed discretization $\mathbf{x}$, is a function of the model parameter vector $\boldsymbol{\theta}$. The structural parameter α is already subsumed into the density function $f(\mathbf{x}, \boldsymbol{\theta}_t)$ and suppressed in notation.

For each $t \in \mathbb{T}$, an estimator is therefore given by

$$\begin{aligned} \hat{\boldsymbol{\theta}}_t &\coloneqq \operatorname*{argmin}_{\boldsymbol{\theta}} \|\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\theta})\|_2^2 = \operatorname*{argmin}_{\boldsymbol{\theta}} \hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\theta})^{\mathsf{T}} \hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\theta}) \\ &= \operatorname*{argmin}_{\boldsymbol{\theta}} \sum_{i \in \mathcal{I}} \Big(f(x_i, \boldsymbol{\theta}_t) - \tilde{\mu}_{\hat{\mathbf{C}}(\mathbf{X}_t)}(x_i) \Big)^2 = \operatorname*{argmin}_{\boldsymbol{\theta}} \sum_{i \in \mathcal{I}} (\tilde{g}(x_i, \boldsymbol{\theta}_t \mid \mathbf{X}_t))^2, \end{aligned} \tag{5.4}$$

with $\tilde{g}(x_i, \boldsymbol{\theta}_t \mid \mathbf{X}_t) \coloneqq f(x_i, \boldsymbol{\theta}_t) - \tilde{\mu}_{\hat{\mathbf{C}}(\mathbf{X}_t)}(x_i)$. If one sets $\tilde{g} \coloneqq \sqrt{g}$, then (5.4) has form (5.1).

Remark: *The* argmin *does not change if one transforms the objective function by a strictly increasing function. Therefore it is more convenient to minimize the squared L^2-norm $\|\cdot\|_2^2$ than the L^2-norm $\|\cdot\|_2$ itself.*

5.2 Numerical Minimization Methods

Minimization of $\|\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\theta})\|_2^2$ is a challenging task because several obstacles, both methodic and numeric, have to be surmounted. The residual vector $\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\theta}) = \hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\theta}, \mathbf{x} \mid \mathbf{X}_t)$ of $\log_2$-length between 9 and 13 has to be calculated efficiently.

[7] See [77].

Algorithm 5.1: Fit VARMA(1, 1) Density to Histogram

Input: $\mathbf{x}$ x values for density $\boldsymbol{\lambda}$ vector of eigenvalues
bw binwidth

Output: parameter estimates $\left(\hat{\phi}_1, \hat{\theta}_1\right)$

Function `DensityVARMA(`$\mathbf{x}, \phi_1, \theta_1, \alpha$`)` **{**
  set $p(\mathcal{G}, z, \phi_1, \theta_1, \alpha) \coloneqq$ to VARMA(1, 1) polynomial (3.34)
  return $x(\boldsymbol{\theta})$
  use Mathematica to find all k roots of $\mathcal{G}$ in $p(\mathcal{G}, z, \phi_1, \theta_1, \alpha)$
  $r(z, k, \phi_1, \theta_1, \alpha) \coloneqq$ root k
  for $i \in \{1, \ldots, k\}$ **do**
    $g_k(z, \phi_1, \theta_1, \alpha) \coloneqq z^{-1}(1 + r(z, k, \phi_1, \theta_1, \alpha))$
    $f_k(z) \coloneqq \left|\frac{1}{\pi} \operatorname{Im} g_k(z, \phi_1, \theta_1, \alpha)\right|^2$
    $h(k) \coloneqq$ `NumIntGaussKronrod(`$f_k(z), 0, 2^{10}$`)` from algorithm 3.1
  end
  $t \coloneqq \operatorname{argmin}_k |h(k)|^2$
  $f(z) \coloneqq f_t(z)$
  foreach *entry* x_i *of* $\mathbf{x}$ **do**
    $y_i \coloneqq f(x_i)$
  end
  return $(\mathbf{y})$
}
Function `ResidualVARMA(`$\mathbf{x}, \boldsymbol{\lambda}, bw$`)` **{**
  $\lambda_{\max} \coloneqq \max\{\boldsymbol{\lambda}\}$ $n \coloneqq \text{Ceiling}(\lambda_{\max})/bw$
  $\mathbf{b}$ sequence from 0 to Ceiling($\lambda_{\max}$) with $n + 1$ equal spaced elements (bins)
  $\mathbf{y}(\mathbf{x}, \phi_1, \theta_1, \alpha) \coloneqq$ `DensityVARMA(`$\mathbf{x}, \phi_1, \theta_1, \alpha$`)`
  foreach *bin* $[b_i, b_{i+1})$ *of* $\mathbf{b}$ **do**
    set $h_\lambda(i)$ to relative count of $\lambda_i \in [b_i, b_{i+1})$
  end
  $\mathbf{g}(\mathbf{x}, \phi_1, \theta_1, \alpha) \coloneqq \mathbf{h}_\lambda - \mathbf{y}(\mathbf{x}, \phi_1, \theta_1, \alpha)$
  $r(\mathbf{x}, \phi_1, \theta_1, \alpha) \coloneqq (\mathbf{g}(\mathbf{x}, \phi_1, \theta_1, \alpha))^\mathsf{T}(\mathbf{g}(\mathbf{x}, \phi_1, \theta_1, \alpha))$
  return $r(\mathbf{x}, \phi_1, \theta_1, \alpha)$
}
Function `EstimateVARMA(`$\mathbf{x}, \boldsymbol{\lambda}, bw$`)` **{**
  $r(\mathbf{x}, \phi_1, \theta_1, \alpha) \coloneqq$ `ResidualVARMA(`$\mathbf{x}, \boldsymbol{\lambda}, bw$`)`
  use Nelder–Mead algorithm[7] to minimize over ϕ_1, θ_1
  $\left(\hat{\phi}_1, \hat{\theta}_1\right) \coloneqq \operatorname{argmin}_{\phi_1, \theta_1} r(\mathbf{x}, \phi_1, \theta_1, \alpha)$
  return $\left(\hat{\phi}_1, \hat{\theta}_1\right)$
}

For the computation of $\tilde{\mu}_{\hat{\mathbf{C}}(\mathbf{X}_t)}(\mathbf{x})$ the eigenvalues of $\hat{\mathbf{C}}(\mathbf{X}_t)$ are fixed, given $\mathbf{X}_t$. As mentioned before, the discretization $\mathbf{x}$ is chosen equal for all $\mathbf{X}_t$, $t \in \mathbb{T}$. It remains to count the eigenvalues in each bin. This computation was optimized and exported for speed to `C`.

Depending on the chosen random matrix model, the way of computation of the spectral density $f(\mathbf{x}, \boldsymbol{\theta}_t)$ and the minimization routine for the squared residual differ.

5.2.1 VMA(1) and VAR(1) processes

Both, VMA(1) and VAR(1) processes, only have 1 model parameter ϕ_1 resp. θ_1 to estimate. In addition, the spectral densities of the covariance matrices for both stochastic processes are given by a distribution generating polynomial of order 4, i.e. all DGPs are solvable in closed form.

As elaborated in subsections 3.1.5 and 3.1.5, Mathematica's `LeafCount` function as a proxy of term length and complexity gives an enormous value of several thousand. This makes the spectral density unfeasible for direct presentation in printed form, but nevertheless it can be computed. As the spectral density $f(\boldsymbol{\theta}_t) \coloneqq f(\mathbf{x}, \boldsymbol{\theta}_t)$ has to be recomputed for every new candidate parameter solution vector $\boldsymbol{\theta}$, fast evaluation is critical. Due to the long length of the expression to be computed we automated the porting from Mathematica to `C`, as writing the code by hand is not feasible anymore.

The residual vector $\hat{\boldsymbol{\varepsilon}}_t$ and thus its norm utilize highly optimized `C` code for fast computation. But the necessary speed improvement comes at the cost of possible numerical instability. In general, the root of a 4^{th} order DGP can be expressed as a function of the polynomials coefficients, involving square and cubic roots. As the DGP coefficients are themselves polynomials in the model parameters, the expressions for DGP roots are long. For some parameter constellations the radicand might be close to zero. Powers, roots and inverses in an unfavorable order might cause some loss in numeric precision. These numerical errors might grow due to the very long expression, involving thousands of operations.

For small parameter values in absolute value the spectral density $f(\boldsymbol{\theta}_t)$ shows evidence of medium numerical instability. We therefore only use parameter values with absolute value not less than $2^{-6} \approx 0.016$.

It remains to find the minimum over the one-dimensional parameter space. The plot of the residual vectors' norm for both, the VMA(1) and VAR(1) random matrix models, are depicted in Figure 5.1.

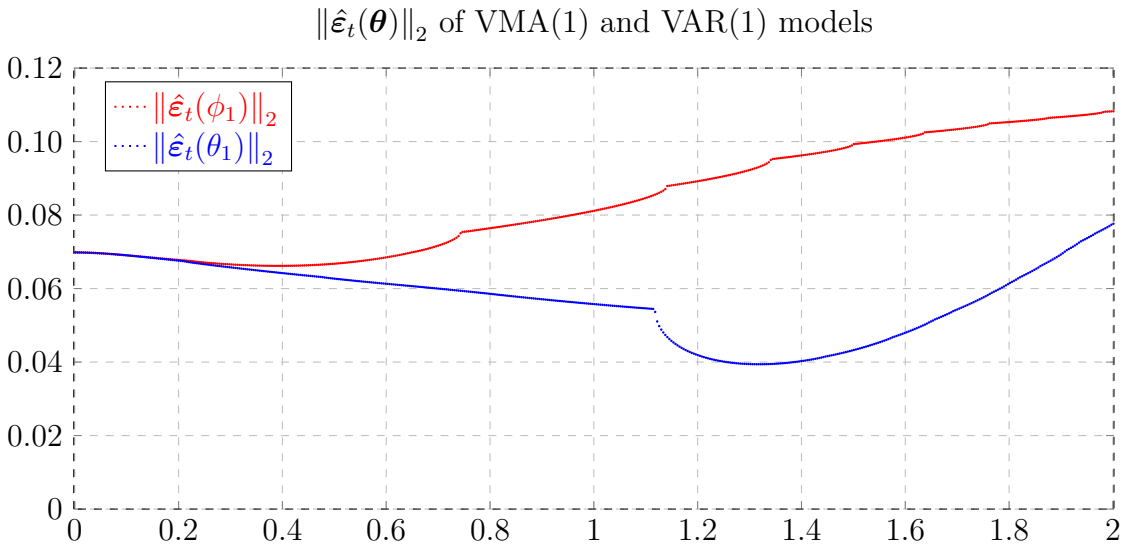

Figure 5.1: $\|\hat{\varepsilon}_t(\boldsymbol{\theta})\|_2$ of VMA(1) (red) and VAR(1) (blue) random matrix models, calculated for parameter values in $(0, 2]$. Depicted is the first day of observation 2007-06-27 of the NASDAQ-100 data set. Algorithm 6.1 was used.

The graphs depicted in Figure 5.1 are roughly exemplary in shape for all the other data sets and sample frequencies. Sometimes, the residual norm as function of the model parameter has distinct multiple minima and narrow valleys, but the shape is never too complicated for standard minimization routines. But as it was the case for small absolute parameter values, there do exists times t with sample $\mathbf{X}_t$ so that the residual norm function $\|\hat{\varepsilon}_t(\boldsymbol{\theta})\|_2$ oscillates increasingly and finally diverges. This oscillating behavior is only visible in very small intervals. Unfortunately, when evaluating the residual norm function one can never be sure on whether its computed value contains any significant numerical error. When under- or overflowing, the presence of numerical errors becomes apparent. However, in a small neighborhood about this pole-like behavior one cannot differentiate oscillating and normal behavior.

Because of this behavior, classical minimization routines that use some form of gradient descent are not easily applicable. If a candidate value happens to be contaminated with numerical error, the gradient gets distorted or, even worse, the algorithm tries to find a minimum in the interval of oscillating function values. Obviously gradient descent routines perform very badly in this conditions.

We opted to minimize a smoothed residual norm function $\widetilde{\|\hat{\varepsilon}_t(\boldsymbol{\theta})\|_2}$ to overcome the problem of numerical outliers. A suitable smoothing procedure has to be insensitive against possible outliers. As numerical errors can accumulate to under- or overflow, outliers are not restricted in magnitude. In addition, for wide applicability cutting of all suspect outliers beyond a threshold is not flexible enough. Outliers slightly beyond the local treshold value can also be influential or distort the minimization routine. Proper weighting schemes do not seem appropriate as the smoothing should be locally confined and with emphasis on only modest many function evaluations

for speed concerns outlier(s) would still get too much weight and thus distort the smoothed function.

Smoothing by local median filtering conforms with the requirements stated above and is very robust. After extensive testing, a median of 7 values yielded a satisfactory compromise between outlier elimination and over-smoothing. The latter flattens away any possible valley with 3 or less consecutive values at the current grid resolution and bigger valleys (and hills) get flattend out, too. In addition, we captured and eliminated all outliers resulting from overflows.

This median smoothing was combined with a robust gradient-free minimization routine. A change in sign of consecutive first differences of the smoothed function values is an indicator of local extrema. One receives curvature information by observing sign changes of the first difference series. This information was utilized and all local minima were investigated by successive parameter grid refinements.

The coarsest grid size is chosen to be of order 2^{-10}, so the previously introduced robust minimization algorithm would only miss vary narrow valleys of absolute wide at most $3 \cdot 2^{-10} \approx 0.003$. Knowing the histogram shape[8] of the ESDs and the shape of the spectral densities of covariance matrix estimators for VMA(1) and VAR(1) processes, such narrow valleys seem unlikely. This is also confirmed by substantial testing, where no such narrow valleys were observed and which yielded the chosen grid size in the first place. The grid is refined in every iterative search step to have only one fourth the grid size compared to the one from the previous step. At 2^{-18} we stop the algorithm and output the parameter with minimal residual norm $\|\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\theta})\|_2$. The so estimated parameter is accurate to $2^{-18} \approx 4 \cdot 10^{-6}$.

With parallel computing and full CPU[9] utilization the run time for the modest sized NASDAQ-100 data set for daily covariance matrix estimates was still about 5.2 hours for VMA(1) model parameter estimation and 1.3 hours for the VAR(1) model. Run time for large data sets like the S&P 500 with 2 minute returns was over 34 hours for the VAR(1) model. This high run times emblematize the need for fast algorithms and efficient implementation. In comparison to the run time cost of residual norm function evaluation the overhead run time cost of the robust minimization algorithm is negligible.

For estimation of the model parameters of VMA(2) or VAR(2) random matrix models, the respectively distribution generating polynomials are of order 9 and have to be solved numerically. In contrast to the 5^{th} order DGP for VARMA$(1,1)$ processes the 4 additional candidate roots aggravate the situation greatly.

The run time for computing the residual norm once is about 8 seconds. The parameter space is 2 dimensional, if a minimization algorithm only needed 15 residual norm evaluations, which is about 1–2 orders of magnitude less than would be expected for reasonable accuracy, estimation time for the smallest data set would be about 5

[8] See Figure 4.17 – Figure 4.21.
[9] CPU used was an Intel i7-2600, which has 4 physical cores and hyper-threading.

days. The methods we employ for the other random matrix models applied to either, VMA(2) or VAR(2) random matrix models for estimation of all data sets would lead to a total run time of nearly one month each. As this rough assessment is a lower bound, estimation of VMA(2) or VAR(2) random matrix models is not feasible at present.

5.2.2 VARMA$(1,1)$ processes

In contrast to the distribution generating polynomial for VMA(1) resp. VAR(1) processes, the DGP for the empirical spectral distribution of the covariance matrix estimator for a VARMA$(1,1)$ process is of order 5. In general, its roots are not solvable in analytical form and therefore have to be approximated numerically. In addition, the VARMA$(1,1)$ random matrix model possesses the two model parameters $\{\theta_1, \phi_1\}$ that have to be estimated simultaneously.

Having to choose one of the five roots that corresponds to the ESD and approximate it numerically is computationally very costly. As an advantage, the quality of the numerical solutions is roughly uniform over the entire parameter space, but for small ϕ_1 values. The numeric challenges one faces when approximating the roots numerically are described in section 3.3.

After comprehensive testing we limited the model parameter space $\Theta := [-1,1]\times[2^{-4},3]$, i.e. we restrict the VAR parameter to $\theta_1 \in [-1,1]$ and the VMA parameter to $\phi_1 \in [2^{-4},3]$. This constraints impose no binding restriction for the parameters, but confinement of the parameter space facilitates numerical minimization. The DGP polynomial is even in ϕ_1, so one can restrict to non-negative parameter values. Because of the numerical difficulty for small absolute ϕ_1 parameter values, we require $\phi_1 \geq 2^{-4}$. This is not really a restriction, as for $\phi_1 = 0$ the VARMA$(1,1)$ model collapses to the VAR(1) model we already estimated.

We employ the robust Nelder–Mead minimization algorithm that was described in the numerical minimization part of section 3.3. The graph of the residual norm is a 2 dimensional manifold and usually there are no distinct local minima. We abort the Nelder–Mead algorithm after only 4 iterations and return the parameter vector with minimal residual norm $\|\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\theta})\|_2$. More iterations for the Nelder–Mead algorithm would be preferable, but it had to be constrained due to excessive run times cost.

Note that the maximal number of only 4 iterations for the Nelder–Mead algorithm does not imply that, after initialization of the 3-simplex that needs 3 objective function evaluation, only 4 more objective function evaluations are performed. In each iteration step the Nelder–Mead algorithm performs a reflextion and expansion sub-step, and if feasible, also a contraction. In contrast to the robust median grid search for VMA(1) and VAR(1) processes, the Nelder–Mead algorithm does not give a accuracy of the argmin. The maximal size of the 3-simplex at termination could be used as a proxy for accuracy. We impose a second termination conditions of the algorithm, so that it stops at an accuracy of 2 decimal places. Usually the precision of

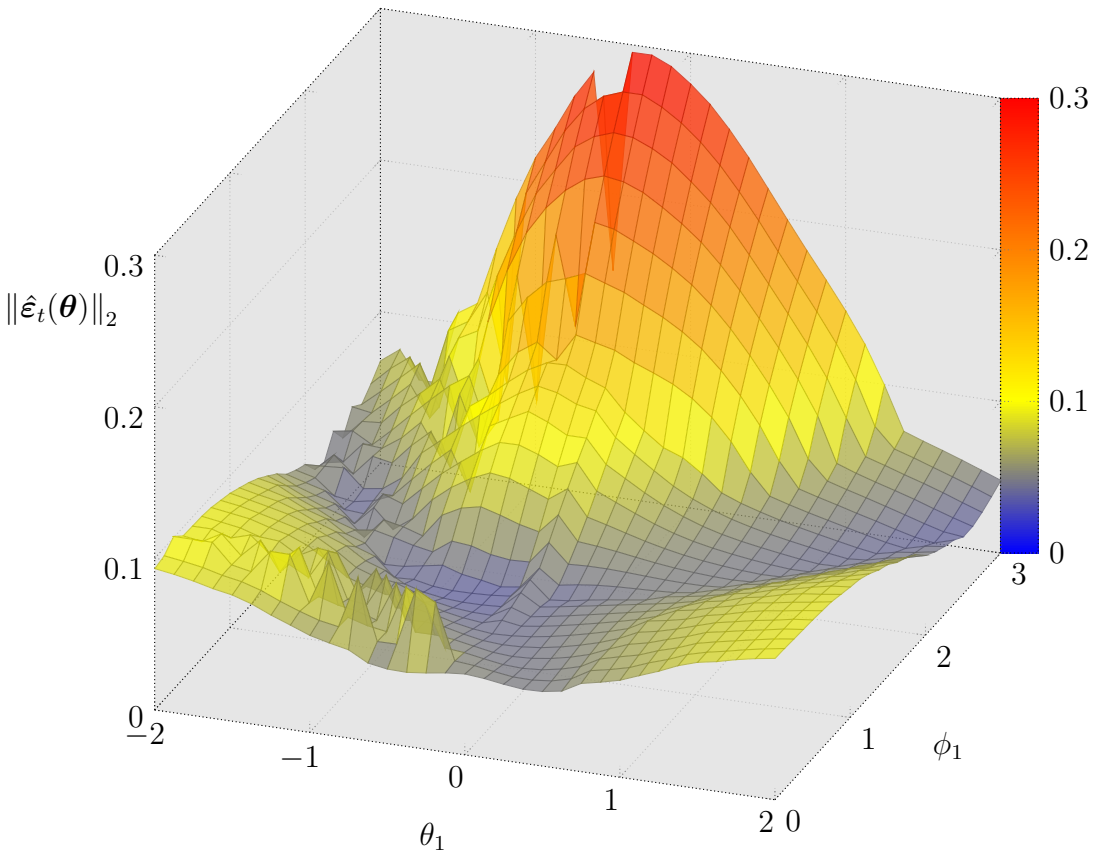

Figure 5.2: $\|\hat{\varepsilon}_t(\boldsymbol{\theta})\|_2$ of VARMA$(1,1)$ random matrix model, calculated for parameter values $(\theta_1, \phi_1) \in [-2, 2] \times (0, 3]$. Depicted is the first day of observation 2007-06-27 of the NASDAQ-100 data set. Algorithm 6.1 was used.

the solution will have said accuracy, but due to the low number of maximal iterations one cannot be sure.

Run time for the NASDAQ-100 data set for daily covariance matrix estimates is about 3 hours, but run time for the much bigger NASDAQ data set is over 30 hours.

5.2.3 VARFIMA(p, d, q) processes

Numerical approximation of the ESD for the covariance estimator of VARFIMA(p, d, q) processes is supremely challenging, as elaborated in subsection 3.2.4 and subsection 3.2.5. Computation of the spectral density was computationally very expensive.

As we remarked at the end of subsection 3.2.4, it is very difficult to automate the generation of VARFIMA$(0, d, 0)$ eigenvalue densities due to the artifacts induced by the numeric algorithm, see Figure 3.6. These artifacts are very problematic for parameter estimation by least-squares minimization, because the histogram data

that should be fitted has actual atoms at the right tail. One does not want to fit those isolated eigenvalues with artifacts. In addition, the artifacts are not stable and change for a different model parameter $\theta = d$.

Calculation of the residual norm once has a run time of 2.5 minutes. The same rationale as for VMA(2) and VAR(2) random matrix models applies. With the algorithms introduced and with present computational power of workstations it is not feasible to estimate the model parameter.

A fortiori, the same holds true for VARFIMA$(1, d, 1)$ eigenvalue densities, see subsection 3.2.5.

5.3 Parameter Estimates and Fit to Historical Data

This section links the theoretical models from chapter 3 in the 1st part with the financial data from chapter 4 in the 2nd part of this thesis. With all the prerequisites and work done, we are finally in a position to estimate the model parameters of the empirical spectral distribution of the covariance estimator for financial return data. Parameter estimation is done as elaborated in the previous paragraph.

As explained in subsection 4.4.2, the ESD of financial stock return data covariance estimates have the characteristic of the biggest eigenvalue being one order of magnitude larger than all the other eigenvalues. For estimation, we use to correlation matrices normed covariance matrix estimators. This normalization ensures that the entries on the main diagonal are all unity and therefore the matrix trace equals the number of stocks in the data set. The market eigenvalue accounts for a huge portion of the matrix trace by itself. This means that the sum of $\lambda_2, \ldots, \lambda_N$ can only achieve a value of $\mathrm{Tr}\left(\hat{\mathbf{C}}\right) - \lambda_1$. This distorts the non-negative eigenvalues $\lambda_2, \ldots, \lambda_N$ to 0. As one wants to employ the ESD of $\hat{\mathbf{C}}$ for parameter estimation, we chose to adjust the ESD for the large value of the market eigenvalue.

One idea to achieve this correction would be to estimate the empirical correlation matrix from the financial stock return data. As the associated eigenvector to the market eigenvalue has estimated components almost all of equal sign and roughly uniform distributed, the L_1-normed eigenvector can be interpreted as a market portfolio and its components as market portfolio weights for each stock. For each return series one could now calculate the markets portfolio return as inner product of the stock returns with the market portfolio vector. A simple linear regression of a constant and the market return series on the original return series projects out the former from the latter. The residuals of this regression are returns with removed influence of the estimated market portfolio, at least approximately and in a linear fashion. Biely and Thurner use this method and also note the close resemblance of the above regression with the CAPM.[10] This also motivates the inclusion of a

[10] See [14, p. 6, eq. (6)].

constant as regressor, which acts by de-meaning the residual returns. These corrected residual returns can be used for further analysis.

A negative aspect with the above described procedure for return cleaning is that the market portfolio is estimated by all data. Thus, the cleaned residual return data will depend on future values. This destroys *adaptiveness* of the return data with any underlying probability model, albeit only slightly so. A more severe practical caveat might be a too positive predictive power for statistics depending solely on in-sample data and considerably worse predictive power for out-of sample statistics. Depending on the purpose of the analysis this method of correction for the market portfolio might be preferable, but we chose an adapted[3] method, i.e. one that does not depend on future data.

We propose to simply exclude the market eigenvalue λ_1 from the ESD, so that it only contains $N-1$ eigenvalues. A corresponding correlation matrix would have a trace of $N-1$. To achieve this constraint all remaining eigenvalues $\lambda_2, \ldots, \lambda_N$ are linearly rescaled to

$$\tilde{\lambda}_k := \frac{N-1}{\mathrm{Tr}\left(\hat{\mathbf{C}}\right) - \lambda_1}\lambda_k = \frac{N-1}{\sum_{k=1}^{N-1}\lambda_k}\lambda_k, \tag{5.5}$$

so that $\sum_{k=1}^{N-1}\tilde{\lambda}_k = N-1$.[11]

We use the by (5.5) rescaled ESD as input for all the parameter estimation algorithms[12] as it removes the ESD distortion of the largest eigenvalue, which the random matrix models do not account for.

5.3.1 Parameter Estimates

The S&P 500 »long« data set for daily estimates and both, the NASDAQ and S&P 500 »big« data sets for daily and weekly estimates, contain more stocks than return observations. By construction, the estimated correlation matrix $\hat{\mathbf{C}}$ has eigenvalues $\lambda_k = 0$ for all $k \in \{N-T+1, \ldots, N\}$. In the ESD, these zero eigenvalues appear as atom at 0 with probability mass of $1-\alpha$, for $\alpha := T/N$.The presence of this atom at zero distorts the empirical spectral distribution twofold. Firstly, the atom at 0 with probability mass $1-\alpha$ and secondly the bulk eigenvalue spectrum only has remaining probability mass α. As the rectangularity ratio α is knowable in advance, α is a structural parameter and one could omit the atom at 0 and rescale the remaining ESD by α^{-1}. Additional rescaling of the remaining eigenvalues is not necessary, as they already sum to $T < N$.

[11] If it is clear from context, we still use the notation λ_k for rescaled eigenvalues $\tilde{\lambda}_k$.
[12] We also use the rescaled ESD for plots.

S&P 500 with 2 minute returns

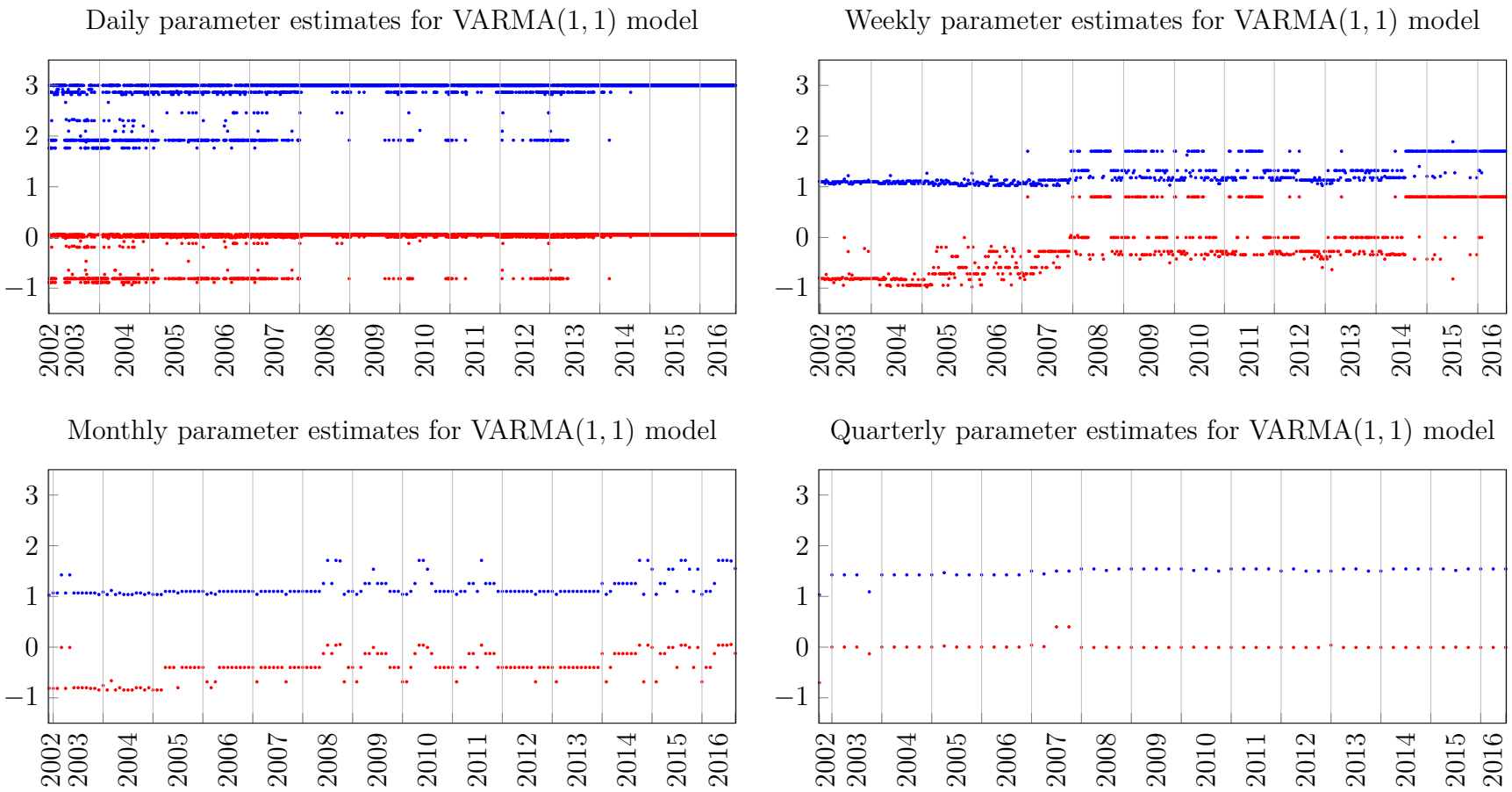

Figure 5.3: Parameter estimates for VARMA$(1,1)$ random matrix model, S&P 500 »long« data set. $\hat{\theta}_1(t) \in (0,3]$ in blue, $\hat{\phi}_1(t) \in [-1,1]$ in red, for all $t \in \mathbb{T}$. Algorithm 6.1 was used.

The removal of the zero eigenvalues would transform the under-identified case to an exactly identified case. This implies that there are no degrees of freedom left and parameter estimation accuracy is poor. In this work, we do not want to compare data sets with structurally different quality of estimated parameters, therefore we refrain from doing so.

Remark: *Note that this line of reasoning yields an active and growing field of research on under-identified statistics. Possible applications include lower subspace approximation by stochastic methods, regularization and lasso methods in regression analysis or* feature extraction *of weak learners in* data mining.

The estimated parameter values are roughly identical between different data sets for the same sampling frequency of either daily, weekly, monthly or quarterly subsamples. Exemplary, we show parameter estimates for the VARMA$(1,1)$ random matrix model for the S&P 500 »long« data set in Figure 5.3.

For daily parameter estimates, the S&P 500 »long« data set contains 224 stocks and a median of 193 observations. This implies for its ESD an atom at 0 with probability mass of about 0.138. As previously elaborated, this atom at 0 distorts the bulk eigenvalues in the ESD and least squares minimization gives parameter estimates whose implied spectral density function of the VARMA$(1,1)$ random matrix model fails to capture the shape of the ESD. This is also visible by the fact that for daily parameter estimates the parameter $\hat{\theta}_1$ admits values near its upper bound of 3. This is a sign that unconstrained optimization would have let to larger parameter estimates. For the three other sampling frequencies the estimated parameters are almost never constrained by the boundaries chosen. But even for larger $\hat{\theta}_1$ parameter values the fit of the implied spectral density function with the ESD would be poor. We confirm and acknowledge that the VARMA$(1,1)$ random matrix model is not applicable in situations with $T < N$, i.e. when there are less observations in time than stocks in the sample used to estimate the correlation matrix $\hat{\mathbf{C}}$.

As expected, parameter variability in time decreases from weekly to monthly and quarterly parameter estimates. Also note that the first quarter only contains the two trading days 2002-12-30 and 2002-12-31, compared with the usual 3 month of observations. For analog reasons, the last week, month or quarter in a data set might also exhibit different characteristics as the rest of the data set, as it might contain notably fewer trading days and thus observations. Therefore, when showing concrete fits of the spectral distribution function with the ESD, we will often chose the second or penultimate subsample.

For quarterly estimates, the MA parameter is essential estimated to be 0, with the exception of the beginning of the financial crisis in the second half of 2007. The monthly estimates show increased variation between 2008–2011 and from 2014 onwards. This pattern is also visible in the weekly estimates from Figure 5.3.

5.3.2 Fit of Spectral Density and ESD

As a first example of the capability of the theoretical spectral densities to approximate the ESD, see Figure 5.4. The NASDAQ-100 data set consists of only 71 stocks, but the spectral density fits the ESD well.

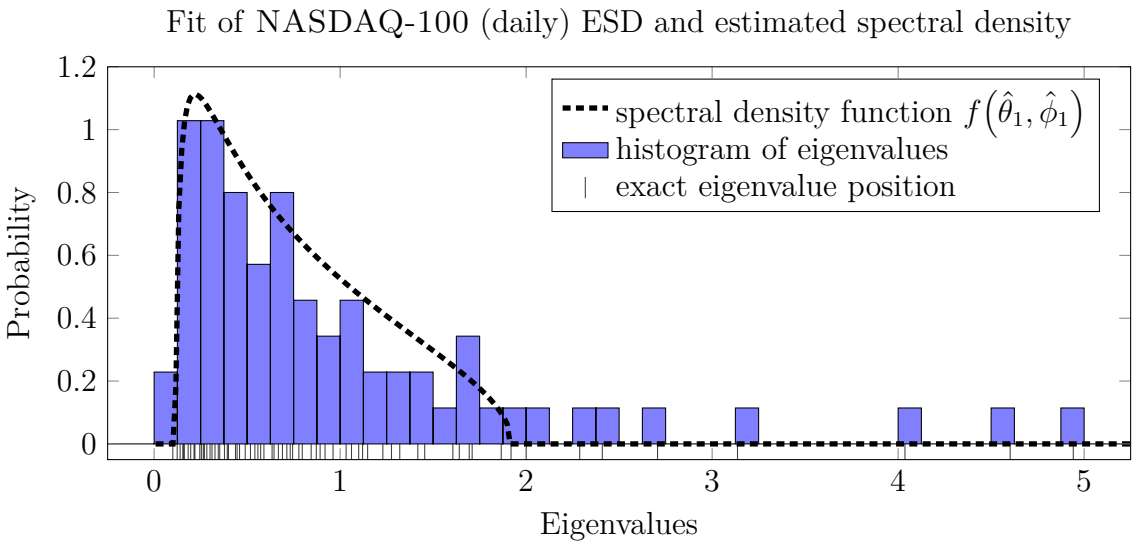

Figure 5.4: ESD of NASDAQ-100 data set (daily) for $t = 2013\text{-}01\text{-}16$ and VARMA$(1,1)$ spectral density function. Algorithm 6.1 was used.

In Figure 5.5, we depict the fit to the S&P 500 »long« data set. Because for daily subsamples one has $T < N$, we only show the three remaining sampling frequencies. For weekly, monthly and quarterly, the spectral density includes all but 6, 5 resp. 7 eigenvalues, also counting the market eigenvalue. As the S&P 500 »long« data set contains 224 stocks, the spectral density approximation contains a share of about 0.97. From a modeling perspective, with only 2 model parameters we were able to approximate the bulk of the ESD of the correlation matrix estimate. The »outlier eigenvalues« can be neglected or have to be modeled in its own right, for instance as a factor model.[13]

For the S&P 500 »big« data set, only for subsamples of frequency month or quarter one has $T \geq N$. The most recent and also complete month is November 2016. Approximation by the 1-parameter VAR(1) model can be seen in Figure 5.6. In contrast to previous VARMA$(1,1)$ matrix models the shape flexibility is less pronounced. In order to be able to capture the peak of very small eigenvalues and approximate the right tail decrease the spectral density function has a long right tail that decreases only very slowly. The right endpoint of its domain is 44.4 and only the two largest eigenvalues exceed this domain endpoint.

[13] See [8].

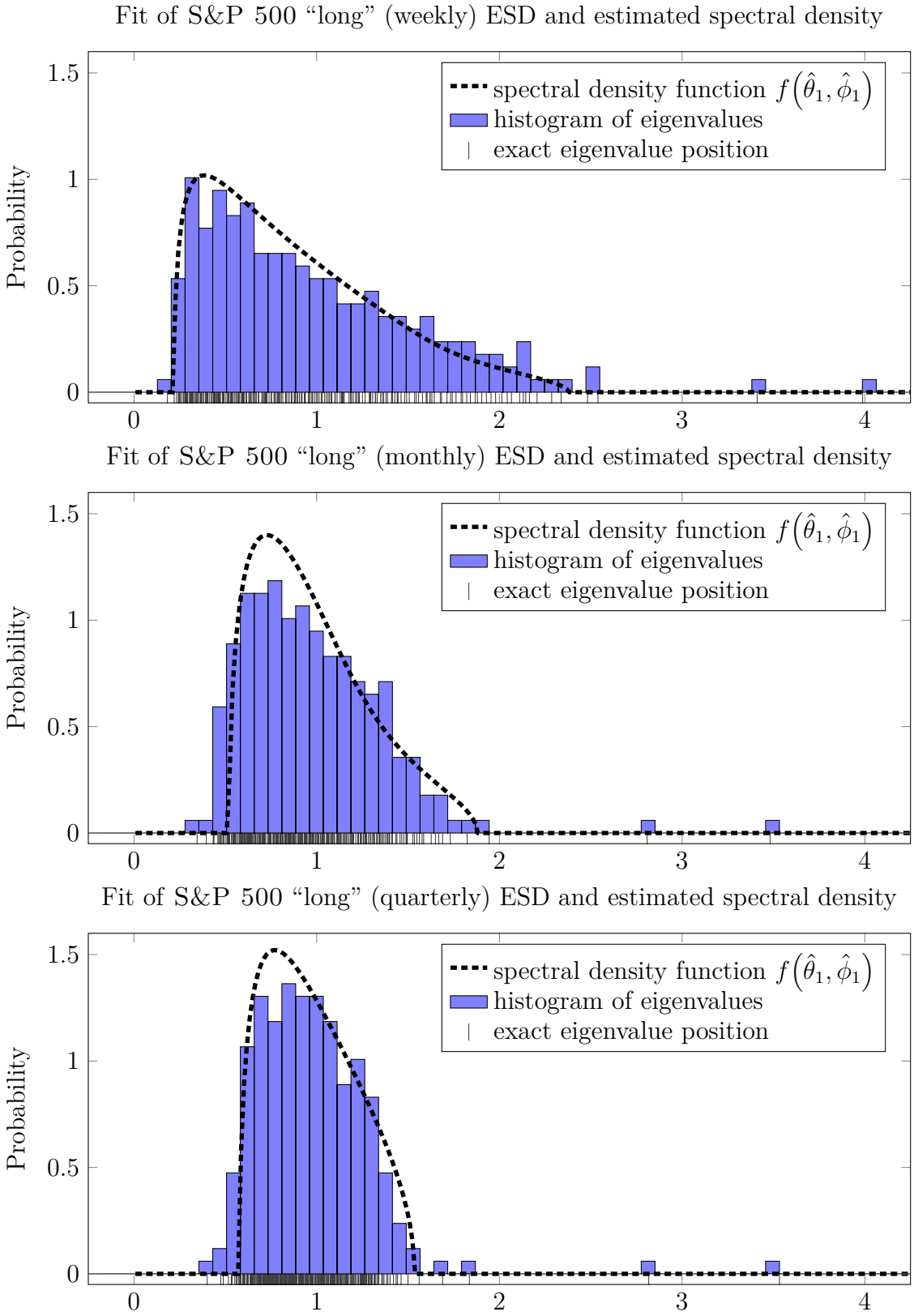

Figure 5.5: Example of model fit for VARMA$(1,1)$ random matrix model, ESD of S&P 500 »long« data set (weekly, monthly and quarterly) for the first complete observation and VARMA$(1,1)$ spectral density function with estimated parameters $\hat{\theta}_1$ and $\hat{\phi}_1$. Algorithm 6.1 was used.

In contrast, the VARMA(1, 1) random matrix model for the same data set, but quarterly parameter estimates has a much narrower domain, see Figure A.3. We show an artificial situation, where we aggregated information of the whole data set by applying the median for all eigenvalues in time. This median ESD is well approximated by the spectral density depicted. Similarly, in Figure A.4 the NASDAQ data set for monthly sampling frequency is aggregated by averaging over the resp. eigenvalues.

The DJIA data set only consists of 26 stocks. But even for this low number satisfactory approximation is possible. In the Figure A.5 we depicted the penultimate observation, 2016-09-15.

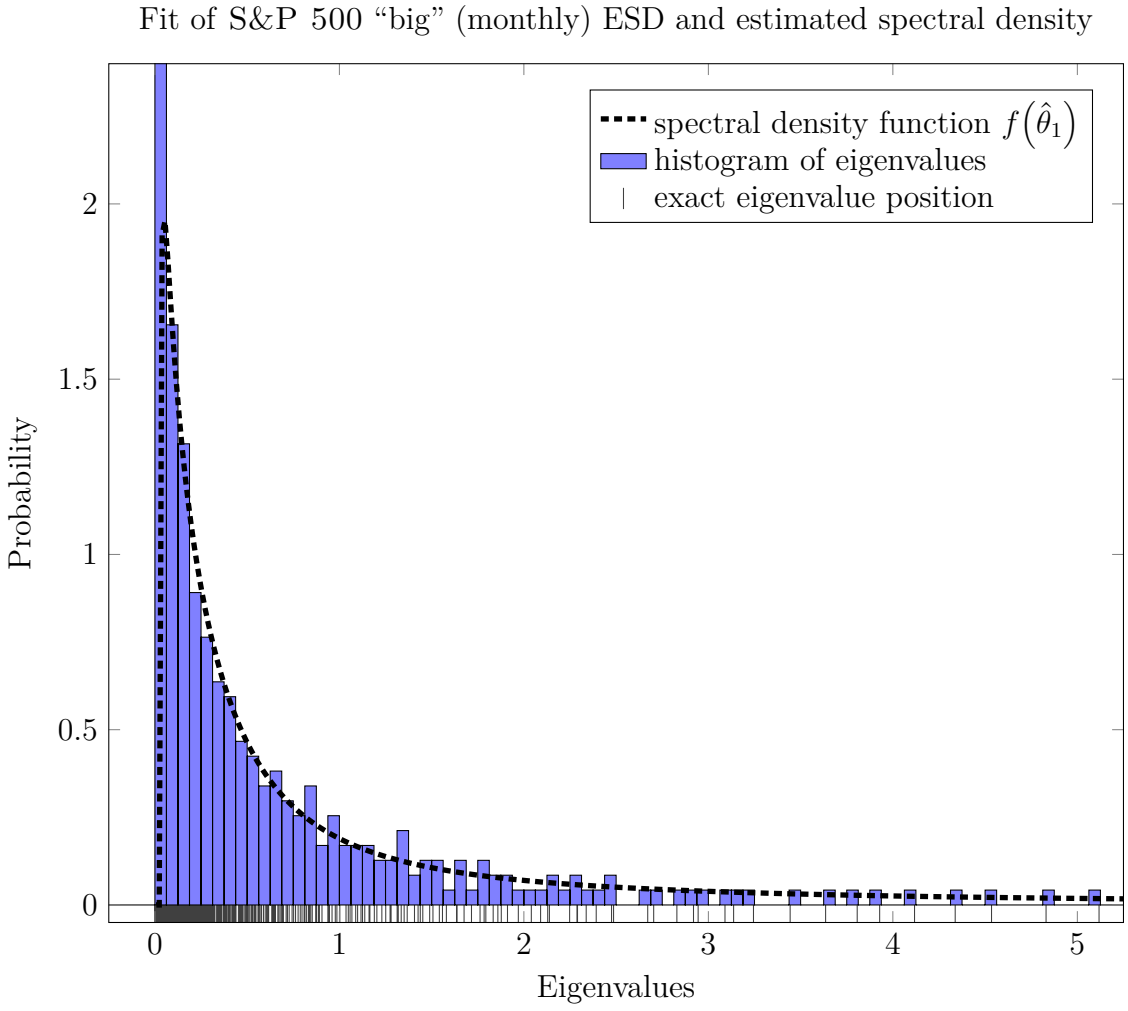

Figure 5.6: Example of model fit for VAR(1, 1) random matrix model, ESD of S&P 500 »big« data set (monthly) for the last complete observation and VAR(1) spectral density function with estimated parameter $\hat{\theta}_1$.

Remark: *To test for robustness, we also estimated some of the data sets by minimization of the L^1-norm and L^∞-norm. The parameter estimates in the L^1-norm are almost identical, which implies a certain degree of robustness. Even in the L^∞-norm the general picture does not change.*

5.4 Reflections on the Results

In subsection 5.3.2 we showed graphically how the fitted VARMA$(1,1)$ densities approximate the observable eigenvalue histogram of the sample covariance matrix.

Reflecting a short moment on what we did and why we did it should help to clarify our approach.

Financial markets are one of the most complex men-made system, as millions of people participate. Financial equity return data is one manifestation one can investigate further. Part of the problem is that we do not have a very long history of observations for the large number of assets available, i.e. in the S&P 500. As the data is high-dimensional in nature, it is quite difficult to assess any cross-dependence but the most simple one, covariances. The sample covariance matrix estimator $\mathbf{C} \coloneqq T^{-1}\mathbf{X}^\intercal\mathbf{X}$ is the simplest formula to decode all second moments of the data without imposing any premature restrictions. But even this simple object decodes very much information and is such very complicated.

The sample covariance matrix estimator is symmetric and at least non-negative definite by construction. Thus, it can be separated by its Eigendecomposition into its spectrum and the eigenbasis. The latter is still very complicated, but the spectrum seems the easier object to research.

It is not our goal to approximate the histogram of sample covariance eigenvalues by any function. With semi- or non-parametric methods like splines one would probably get a fit way better by almost all measures. But the very problem at hand is that the financial market and thus its sample covariance, and even its spectrum, is so complicated, that we try to find some structure. Finding structure means to be able to explain major parts by a simple mechanism.

The sample cross-covariance matrix $\mathbf{C}$ has an interpretation as encoding as its entry $c_{i,j}$ the covariance between stocks i, j for $i \neq j$. But simple linear algebra tells us that the sample auto-covariance $\mathbf{A} \coloneqq T^{-1}\mathbf{X}\mathbf{X}^\intercal$ has the same non-trivial spectrum as the cross-covariance matrix $\mathbf{C}$. It is therefore equally valid to not impose restrictions or assumptions on the cross-covariances, but on the auto-covariances, that is the evolution of stock returns in time.

In this light, Figure 5.5 is quite remarkable. Its statement is, that we can abstract away from the very details of every single asset. Because of how the true spectrum of the population covariance matrix and the spectrum of the sample covariance matrix are related for the model class we investigate, one of the most basic results from Random Matrix Theory (RMT) explains much of the shape of the sample eigenvalue histogram — the Marchenko–Pastur law depicted in Figure 1.1. For this to work, we just had to assume that the financial returns are random variables with finite second moment and otherwise independent and identically distributed (i.i.d.). The moment assumption is no restriction since we want to investigate second moments anyway. The highly unrealistic case of assuming no dependence structure at all already gives

a quite decent fit. One can thus conclude that much of the shapre of the sample eigenvalue histogram is due to algebraic and probabilistic constraints on how such a spectral density can possibly look like.

If we just assume that all assets behave in time like very simple linear processes like VARMA$(1,1)$, then we can improve the fit and capture much of the complexity with just a few additional parameters. We do not claim that financial markets actually behave solely like VARMA$(1,1)$ processes in time, but they give a low-parametric approximation. The structural part of the spectrum that is approximated in such a low-parametric way is therefore not really distinguishable from very basic behavior of assets in time. Of course it is well known that the trajectory of single assets is not well described by a ARMA$(1,1)$ process. But apparently the macro-behavior of many such assets can be well described by this assumption. It is certainly not a much stronger assumption than assuming a block-matrix structure for the cross-covariance, grouping assets by sectors or other characteristics.

As we are in a high-dimensional setting and research the macro structure, the VARMA$(1,1)$ model seems like a good macro explanation. Single stocks do not behave like ARMA$(1,1)$ processes, because this would be easily predictable and lead to opportunities for statistical arbitrage. The evolution of a single asset might very well be described by information newly available to certain groups of investors. But the whole market as an aggregate seems to behave very much like an VARMA$(1,1)$ process. The difference is the coupling to consider it vector-valued so as to capture the macro dynamics of the whole financial market and not of single stocks. It is also quite a good result that financial markets are in big parts indistinguishable in there macro behavior to VARMA$(1,1)$ dynamics.

With this background, the actual fit of the VARMA$(1,1)$ model to the sample covariance spectrum seems only relevant to compare between structurally similar results. It would be interesting to compare with other low-parametric approaches that also explain the shape of the spectrum well.

6 Conclusions and Perspectives

We complete this thesis by a recapitulation of its content and draw conclusions. From this combined, we give perspectives on how this work can be extended.

6.1 Recapitulation and Conclusions

After a short introduction in chapter 1 into the topic of high-dimensional data we showed the relevance of covariance matrices as basic building blocks in many applied areas. We continued to motivate the relevance to describe high-dimensional covariance matrices parsimoniously and the need to model them. We argued that to structurally model high-dimensional matrix models non-commutative probability theory (NCPT) is the right tool.

In the first part of this thesis, we first laid the theoretical foundation necessary to introduce random matrix models capable of containing dependencies in time and across stocks. As a first step, in chapter 2 we introduced the concept of non-commutative probability theory and motivated the necessity of a new concept of stochastic independence suitable for NCPT. The abstract building blocks to advance and enrich NCPT are of interest on its own. They help one understand classical probability theory and might also show new paths on how to extend classical probability theory.

Cumulants are a basic tool to facilitate addition of free NCRVs. This holds true for both, classical and non-commutative probability theory if one selects the associated cumulant type. There is also a branch of research, connecting both types of cumulants to crossing resp. non-crossing partitions and mathematical combinatorics.

Distributions as collection of moments are not analytical tractable, so we introduced various transformations to facilitate additive resp. multiplicative convolution of distributions belonging to free NCRVs. Most notably the $\mathcal{R}$-transform enables addition of free NCRVs and the $\mathcal{S}$-transform enables multiplication.

Rectangular non-commutative random variables are very important, as we employ the theoretic NCRVs to model random matrices. This link was shown in chapter 3, where we also introduced the concept of asymptotic freeness necessary as motivation on why certain non-commutative random variables model specific random matrices. This connection enabled us to turn from abstract NCRVs to concrete random matrix

models. In the same vein the concept of universality was introduced to emphasize the ubiquitous possibilities of application and stability of results. We introduced the Wigner and Wishart random matrix ensembles and worked out some lemmata on convergence of random probability measures. Both are essential if one wants to switch from RVs of empirical spectral distributions to their deterministic limits.

Sections 3.1 and 3.2 unite and use all the theory developed to calculate the spectral density function of correlation matrix estimators for random matrix models which are, broadly, from the VARFIMA(p, d, q) family. Even for the most basic VMA(1) resp. VAR(1) random matrix models the theory is quite complicated. Fortunately, for those two models there exist an explicit solution in closed form. Unfortunately these are too long to express in a meaningful way. Methods from numerical mathematics have to be applied to advance further. All spectral densities for all the other random matrix models introduced are only solvable numerically. As the algorithms for the spectral density are quite complicated an efficient implementation and choice of suitable numerical methods is vital.

In the second part of this thesis, we employed the theoretical findings to exemplary show on how to utilize the first part in practice. In chapter 4 we gave an overview of various data sets. We use high-dimensional financial high frequency data sets to estimate the model parameters of correlation matrix estimates under the dynamics of certain random matrix models.

The high frequency data had to be processed and cleaned. Because of certain requirements on the data in light of minimal liquidity and data completeness the raw data had to be constrained. We gave some basic descriptive statistics of the data and presented the data sets used for subsequent empirical analysis. Finally, the empirical counterpart to the spectral density, the empirical spectral distribution, is estimated.

The final chapter 5 first explains the methodology used for parameter estimation. It continuous with a thorough explanation of the challenges one faces when estimating the model parameters. Due to the complexity of the algorithms needed and constraints on the computational power available only the three most basic models VMA(1), VAR(1) and VARMA$(1, 1)$ could be estimated for the whole data set and for subsampling frequencies. We only depicted a small subset of the estimates, showcasing the ability of the methods used to approximate the bulk of the ESD of correlation matrix estimators.

The appendices collect mathematical definitions and theorems used, some proofs too long for the main part and some additional plots.

6.2 Motivations and Restrictions

In this work we restrict ourselves to symmetric random matrices. Otherwise, the spectrum would not be real-valued but be complex. This is much harder to analyze and another endeavor. We concentrate on the relevant Random Matrix Theory but also wanted to give a self-contained introduction. The Wishart ensemble is important for the Marchenko–Pastur distribution, the first cornerstone of RMT applied to covariance matrices. To understand the concept of *freeness* is essential to understand RMT and also to work with different tools. The family of transformations was used to develop the main building blocks. We shortly touch on rectangular random matrices and how they can be incorporated in the framework presented.

6.3 Perspectives

With more computational power the VMA(2) and VAR(2) random matrix models are estimable, as well as the VARFIMA$(1, d, 1)$ models. The VARFIMA$(0, d, 0)$ random matrix model introduces the idea of fractional integration to random matrix models, but it seems not to be able to replicate the empirical characteristics of ESDs of correlation matrix estimates of financial stock return data. The VARFIMA$(1, d, 1)$ should, in principle, work well as it extends the VARMA$(1, 1)$ model. It is of interest whether fractional integration can, it least in a statistical sense, help explain the shape of the ESD.

A next step after the initial parameter estimation could be the construction of a factor model. We did not analyze the eigenvectors belonging to the largest eigenvalues not the market eigenvalue. It is known that they contain suitable information that can be utilized for inference and prediction. The structure given by our parametrization of the estimated correlation matrix could also be utilized by economic scenario generators to simulate future market behavior. Non-commutative probability theory can be extended to non-commutative stochastic processes. They also can be used to capture high-dimensional marked dynamics in a parsimonious fashion.

Also a new field of research is the explicit modeling of eigenvectors of random matrices. For non-rotational invariant random matrix models the eigenvectors contain much information. As an additional bonus, eigenvectors of normal correlation matrices are quite easy interpretable as portfolio vectors.

Appendices

A Theorems, Proofs and Results

A.1 Definitions

We want to extend convergence of distributions from a sequence of distributions of an implicit single random variable to a sequence of *joint* distributions of a fixed set of random variables.

Definition (Convergence of Joint Distributions)
Let $\mathcal{I}$ be an index set, indexing random variables in a joint distribution. For each $n \in \mathbb{N}$, let $(\mathcal{A}_n, \varphi_n)$ be non-commutative probability spaces and let $\left(\mathbf{a}_i^{(n)}\right)_{i\in\mathcal{I}} \in \mathcal{A}_n$ be a family of non-commutative random variables with joint distribution

$$\mu_n := \mu_{\left(\mathbf{a}_i^{(n)}\right)_{i\in\mathcal{I}}}.$$

We say the sequence of joint distributions $(\mu_n)_{n\in\mathbb{N}}$ converges *for $n \to \infty$, if there exists a joint distribution μ, such that*

$$\lim_{n\to\infty} (\mu_n(P))_{n\in\mathbb{N}} = \mu(P), \qquad (\forall P \in \mathbb{C}\langle X_i \mid i \in \mathcal{I}\rangle).$$

Notation: We say that a sequence of random variables $\left(\mathbf{a}^{(n)}\right)_{n\in\mathbb{N}}$ converges in distribution to some random variable $\mathbf{a} \in \mathcal{A}$ in some non-commutative probability space $(\mathcal{A}, \varphi)$, if the respective sequence of distributions of $\left(\mathbf{a}^{(n)}\right)_{n\in\mathbb{N}}$ converges in distribution to the distribution of $\mathbf{a}$. This naturally extends to sequences of families of random variables converging.

A.2 Theorems

A.2.1 Some Theorems in Complex Analysis

Theorem A.1 (Inverse Cauchy-transform)
Let μ be a real probability measure and let $\mathcal{G}_\mu(z)$ be its Cauchy-transform. Then we have

$$\frac{\mu((a,b)) + \mu([a,b])}{2} = \lim_{\varepsilon\searrow 0} -\frac{1}{\pi}\int_a^b \operatorname{Im}\mathcal{G}_\mu(x+\mathrm{i}\varepsilon)\,\mathrm{d}x, \quad (a < b \in \operatorname{supp}(\mu) \subseteq \mathbb{R}). \tag{A.1}$$

In addition, if a, b are continuity points of μ, i.e. neither of a, b is an atom of μ, we have

$$\mu(\{a\}) = \mu(\{b\}) = 0 \implies \mu((a,b)) = \mu([a,b]),$$

and (A.1) simplifies to

$$\mu((a,b)) = \mu([a,b]) = \lim_{\varepsilon \searrow 0} -\frac{1}{\pi} \int_a^b \operatorname{Im} \mathcal{G}_\mu(x + \mathrm{i}\varepsilon)\, \mathrm{d}x, \quad (a < b \in \operatorname{supp}(\mu) \subseteq \mathbb{R}).$$

Proof (Theorem A.1)
In the situation of Theorem A.1, the imaginary part of the Cauchy-transform $\mathcal{G}_\mu(z)$ is, with $x + \mathrm{i}\varepsilon = z \in \mathbb{C}$

$$\begin{aligned}
\operatorname{Im} \mathcal{G}_\mu(x + \mathrm{i}\varepsilon) &= \operatorname{Im} \int_{\mathbb{R}} \frac{1}{(x + \mathrm{i}\varepsilon) - \lambda}\, \mathrm{d}\mu(\lambda) = \operatorname{Im} \int_{\mathbb{R}} \frac{1}{(x - \lambda) + \mathrm{i}\varepsilon}\, \mathrm{d}\mu(\lambda) \\
&= \operatorname{Im} \int_{\mathbb{R}} \frac{x - \lambda}{(x - \lambda)^2 + \varepsilon^2} - \mathrm{i} \frac{\varepsilon}{(x - \lambda)^2 + \varepsilon^2}\, \mathrm{d}\mu(\lambda) \\
&= -\varepsilon \int_{\mathbb{R}} \frac{1}{(x - \lambda)^2 + \varepsilon^2}\, \mathrm{d}\mu(\lambda). \qquad \text{(A.2)}
\end{aligned}$$

Because of

$$\int \frac{1}{u^2 + 1}\, \mathrm{d}u = \arctan(u) + c, \qquad \text{(A.3)}$$

we want to bring (A.2) in this form. By direct calculation

$$\begin{aligned}
\lim_{\varepsilon \searrow 0} -\frac{1}{\pi} \int_a^b \operatorname{Im} \mathcal{G}_\mu(x + \mathrm{i}\varepsilon)\, \mathrm{d}x &= \lim_{\varepsilon \searrow 0} -\frac{1}{\pi} \int_a^b -\varepsilon \int_{\mathbb{R}} \frac{1}{(x - \lambda)^2 + \varepsilon^2}\, \mathrm{d}\mu(\lambda)\, \mathrm{d}x \\
&\overset{\text{Fubini}}{=} \lim_{\varepsilon \searrow 0} \frac{1}{\pi} \int_{\mathbb{R}} \varepsilon \int_a^b \frac{1}{(x - \lambda)^2 + \varepsilon^2}\, \mathrm{d}x\, \mathrm{d}\mu(\lambda) \\
&= \lim_{\varepsilon \searrow 0} \frac{1}{\pi} \int_{\mathbb{R}} \varepsilon \int_a^b \frac{1}{\varepsilon^2} \frac{1}{\left(\frac{x - \lambda}{\varepsilon}\right)^2 + 1}\, \mathrm{d}x\, \mathrm{d}\mu(\lambda), \qquad \text{(A.4)}
\end{aligned}$$

and the change of variable

$$u := \frac{x - \lambda}{\varepsilon} \qquad \implies \qquad \mathrm{d}x = \varepsilon\, \mathrm{d}u,$$

we get a form where (A.3) is applicable:

$$\int_a^b \frac{1}{\left(\frac{x - \lambda}{\varepsilon}\right)^2 + 1}\, \mathrm{d}x = \int_{\frac{a - \lambda}{\varepsilon}}^{\frac{b - \lambda}{\varepsilon}} \frac{1}{u^2 + 1} \varepsilon\, \mathrm{d}u = \varepsilon \arctan(u) \Big|_{\frac{a - \lambda}{\varepsilon}}^{\frac{b - \lambda}{\varepsilon}}.$$

Choosing a vanishing integration constant $c = 0$ we get for (A.4):

$$\lim_{\varepsilon \searrow 0} \frac{1}{\pi} \int_a^b \operatorname{Im} \mathcal{G}_\mu(x + \mathrm{i}\varepsilon)\, \mathrm{d}x = \lim_{\varepsilon \searrow 0} \frac{1}{\pi} \int_{\mathbb{R}} \arctan\left(\frac{b - \lambda}{\varepsilon}\right) - \arctan\left(\frac{a - \lambda}{\varepsilon}\right) \mathrm{d}\mu(\lambda) \qquad \text{(A.5)}$$

For $\lambda = a$ the argument of arctan and therefore the second summand itself vanishes and we are left with

$$\lim_{\varepsilon \searrow 0} \frac{1}{\pi} \arctan\left(\frac{b-a}{\varepsilon}\right) = \frac{1}{2},$$

because $b - a > 0$ and for $\lim_{\varepsilon \searrow 0}$ the argument goes to infinity. For $\lambda = b$, an analogous argument also yields $\frac{1}{2}$. The range of the usual principal value of arctan is

$$\text{Image}(\arctan) = \left(-\frac{\pi}{2}, \frac{\pi}{2}\right),$$

and arctan is monotonically increasing and we have that $b > a$, therefore we have

$$0 \leq \frac{1}{\pi}\left(\arctan\left(\frac{b-\lambda}{\varepsilon}\right) - \arctan\left(\frac{a-\lambda}{\varepsilon}\right)\right) \leq 1, \tag{A.6}$$

and for $\lim_{\varepsilon \searrow 0}$ (A.6) approximates the characteristic function $\mathbb{1}_{(a,b)}$. Because of (A.6) we can apply the dominated convergence theorem to (A.5) and the claim is proved. ■

Remark: *Theorem A.1 is closely related to the* Sokhotski–Plemelj theorem *from complex analysis.*

To calculate $\operatorname{Im} \mathcal{G}_\mu(x + \mathrm{i}\varepsilon)$, one often needs to know the following:

Lemma A.2 ($\operatorname{Im} \sqrt{x + \mathrm{i}y}$)
For $x, y \in \mathbb{R}$ we have

$$\operatorname{Im} \sqrt{x + \mathrm{i}y} = \sqrt{\frac{\sqrt{x^2+y^2} - x}{2}}, \tag{A.7}$$

where we choose the branch with the same sign as y, if not otherwise stated, so that the square root is well-defined.

Proof (Lemma A.2)
By direct calculation, we have

$$\begin{aligned}
\left(\sqrt{\frac{\sqrt{x^2+y^2}+x}{2}} + \mathrm{i}\sqrt{\frac{\sqrt{x^2+y^2}-x}{2}}\right)^2 &= \frac{\sqrt{x^2+y^2}+x}{2} + \mathrm{i}^2\frac{\sqrt{x^2+y^2}-x}{2} \\
&\quad + \mathrm{i}2\sqrt{\frac{\sqrt{x^2+y^2}+x}{2}}\sqrt{\frac{\sqrt{x^2+y^2}-x}{2}} \\
&= x + \mathrm{i}\sqrt{\left(\sqrt{x^2+y^2}+x\right)\left(\sqrt{x^2+y^2}-x\right)} \\
&= x + \mathrm{i}\sqrt{x^2+y^2-x^2} = x + \mathrm{i}y.
\end{aligned}$$

Because we have $\sqrt{x^2+y^2} \geq x$ for all $x, y \in \mathbb{R}$, the radicand is never negative and thus the RHS of (A.7) is always real. So we have proven the claimed identity. ■

Lemma A.3 (Re $\sqrt{x+\mathrm{i}y}$)
For $x, y \in \mathbb{R}$ we have

$$\operatorname{Re}\sqrt{x+\mathrm{i}y} = \sqrt{\frac{\sqrt{x^2+y^2}+x}{2}}.$$

Proof Analogous to Lemma A.2 ■

Lemma A.4 (Schur's Complement)
Let

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}, \tag{A.8}$$

be an invertible block-matrix, where the entries $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ are matrices with appropriate dimensions and $\mathbf{D}^{-1}$ exists. Then one has

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}^{-1} = \begin{pmatrix} (\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & -(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & \mathbf{D}^{-1}+\mathbf{D}^{-1}\mathbf{C}(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \end{pmatrix}. \tag{A.9}$$

Proof Multiplication of (A.8) with (A.9) gives the identity matrix $\mathbf{I}$. ■

A.3 Results

A.3.1 Polynomial for VMA(2)

$$M^9 z^5 (\phi_2-1)^2 \left(\phi_1^2-4\phi_2\right) \left(\phi_1^2-(\phi_2+1)^2\right)^4$$

$$-M^8 \Big(z^5 \Big((-4\alpha-5) \left(\phi_1^2-4\phi_2\right) (\phi_2-1)^2 \left(\phi_1^2-(\phi_2+1)^2\right)^4\Big)$$
$$+4z^4\alpha \left(\phi_1^2-(\phi_2+1)^2\right)^2 \left(\left((\phi_2-3)\phi_2+1\right)\phi_1^4+\left(\phi_2\left(\phi_2\left((\phi_2-2)\phi_2+10\right)-2\right)+1\right)\phi_1^2-\phi_2(\phi_2+1)^2\left(\phi_2(5\phi_2-6)+5\right)\right)\Big)$$

$$-M^7 \Big(z^5 \Big(2(-\alpha(3\alpha+10)-5) \left(\phi_1^2-4\phi_2\right) (\phi_2-1)^2 \left(\phi_1^2-(\phi_2+1)^2\right)^4\Big)$$
$$+z^4 \Big(16\alpha(\alpha+1) \left(\phi_1^2-(\phi_2+1)^2\right)^2 \left(\left((\phi_2-3)\phi_2+1\right)\phi_1^4+\left(\phi_2\left(\phi_2\left((\phi_2-2)\phi_2+10\right)-2\right)+1\right)\phi_1^2-\phi_2(\phi_2+1)^2\left(\phi_2(5\phi_2-6)+5\right)\right)\Big)$$
$$-2z^3\alpha^2 \Big(\left(\phi_2(3\phi_2-14)+3\right)\phi_1^6+2\left(\phi_2\left(\phi_2\left((\phi_2-2)\phi_2+18\right)-2\right)+1\right)\phi_1^4$$
$$+(\phi_2+1)^2\left(3\phi_2^4+26\phi_2^2+3\right)\phi_1^2-4\phi_2(\phi_2+1)^4\left(\phi_2(5\phi_2-2)+5\right)\Big)\Big)$$

$$-M^6 \Big(z^5 \Big(2(-\alpha(\alpha(2\alpha+15)+20)-5) \left(\phi_1^2-4\phi_2\right) (\phi_2-1)^2 \left(\phi_1^2-(\phi_2+1)^2\right)^4\Big)$$
$$+z^4 \Big(8\alpha(\alpha(3\alpha+8)+3) \left(\phi_1^2-(\phi_2+1)^2\right)^2 \left(\left((\phi_2-3)\phi_2+1\right)\phi_1^4+\left(\phi_2\left(\phi_2\left((\phi_2-2)\phi_2+10\right)-2\right)+1\right)\phi_1^2-\phi_2(\phi_2+1)^2\left(\phi_2(5\phi_2-6)+5\right)\right)\Big)$$
$$+z^3 \Big(2(-4\alpha-3)\alpha^2 \Big(\left(\phi_2(3\phi_2-14)+3\right)\phi_1^6+2\left(\phi_2\left(\phi_2\left((\phi_2-2)\phi_2+18\right)-2\right)+1\right)\phi_1^4+(\phi_2+1)^2$$
$$\cdot\left(3\phi_2^4+26\phi_2^2+3\right)\phi_1^2-4\phi_2(\phi_2+1)^4\left(\phi_2(5\phi_2-2)+5\right)\Big)\Big)$$
$$+4z^2\alpha^3 \Big(\left((\phi_2-8)\phi_2+1\right)\phi_1^4+\left(\phi_2\left(\phi_2\left((\phi_2-8)\phi_2-2\right)-8\right)+1\right)\phi_1^2-2\phi_2(\phi_2+1)^2\left(\phi_2(5\phi_2+2)+5\right)\Big)\Big)$$

$$
\begin{aligned}
&-M^5\left(z^5\left((-\alpha(\alpha(\alpha(\alpha+20)+60)+40)-5)\left(\phi_1^2-4\phi_2\right)\left(\phi_2-1\right)^2\left(\phi_1^2-\left(\phi_2+1\right)^2\right)^4\right)\right.\\
&+z^4\left(16\alpha(\alpha+1)(\alpha(\alpha+5)+1)\left(\phi_1^2-\left(\phi_2+1\right)^2\right)^2\left(\left(\left(\phi_2-3\right)\phi_2+1\right)\phi_1^4+\left(\phi_2\left(\phi_2\left(\left(\phi_2-2\right)\phi_2+10\right)-2\right)+1\right)\phi_1^2-\phi_2\left(\phi_2+1\right)^2\left(\phi_2\left(5\phi_2-6\right)+5\right)\right)\right)\\
&+z^3\left(2\alpha^2\left(-2\left(\alpha(7\alpha+12)+\phi_2\left(\phi_2\left((\alpha(7\alpha+12)+3)\phi_2^2-3(\alpha(3\alpha+8)+2)\phi_2+8\alpha(14\alpha+27)+54\right)-3(\alpha(3\alpha+8)+2)\right)+3\right)\phi_1^4\right.\right.\\
&\quad-\left(\phi_2+1\right)^2\left(\alpha(17\alpha+36)+\phi_2\left(\phi_2\left(-6\phi_2\alpha^2+2(73\alpha+156)\alpha+(\alpha(17\alpha+36)+9)\phi_2^2+78\right)-6\alpha^2\right)+9\right)\phi_1^2-\phi_1^6\left((\alpha(17\alpha+36)+9)\phi_2^2\right.\\
&\quad\left.\left.-42(2\alpha(\alpha+2)+1)\phi_2+\alpha(17\alpha+36)+9\right)+2\phi_2\left(\phi_2+1\right)^4\left((\alpha(59\alpha+120)+30)\phi_2^2-2(\alpha(13\alpha+24)+6)\phi_2+\alpha(59\alpha+120)+30\right)\right))\\
&+z^2\left(8\alpha^3(2\alpha+1)\left(\left(\left(\phi_2-8\right)\phi_2+1\right)\phi_1^4+\left(\phi_2\left(\phi_2\left(\left(\phi_2-8\right)\phi_2-2\right)-8\right)+1\right)\phi_1^2-2\phi_2\left(\phi_2+1\right)^2\left(\phi_2\left(5\phi_2+2\right)+5\right)\right)\right)\\
&\left.+z\left(\alpha^4\left(4\phi_2\left(\phi_2\left(5\phi_2+6\right)+5\right)-\phi_1^2\left(\left(\phi_2-18\right)\phi_2+1\right)\right)\right)\right)
\end{aligned}
$$

$$
\begin{aligned}
&-M^4\left(z^5\left(-(5\alpha(\alpha(\alpha+2)(\alpha+6)+4)+1)\left(\phi_1^2-4\phi_2\right)\left(\phi_2-1\right)^2\left(\phi_1^2-\left(\phi_2+1\right)^2\right)^4\right)+z^4\left(4\alpha(\alpha(\alpha+2)(\alpha(\alpha+14)+8)+1)\left(\phi_1^2-\left(\phi_2+1\right)^2\right)^2\right.\\
&\quad\left.\cdot\left(\left(\left(\phi_2-3\right)\phi_2+1\right)\phi_1^4+\left(\phi_2\left(\phi_2\left(\left(\phi_2-2\right)\phi_2+10\right)-2\right)+1\right)\phi_1^2-\phi_2\left(\phi_2+1\right)^2\left(\phi_2\left(5\phi_2-6\right)+5\right)\right)\right)\\
&+z^3\left(2\alpha^2\left(\left(-(\alpha(\alpha(10\alpha+51)+36)+3)\phi_2^2+14(2\alpha(\alpha(2\alpha+9)+6)+1)\phi_2-\alpha(\alpha(10\alpha+51)+36)-3\right)\phi_1^6\right.\right.\\
&\quad+2\left(-3\alpha(\alpha(2\alpha+7)+4)+\phi_2\left(\alpha(\alpha(2\alpha+27)+24)+\phi_2\left(\phi_2\left(\alpha(\alpha(2\alpha+27)+24)-(3\alpha(\alpha(2\alpha+7)+4)+1)\phi_2+2\right)\right.\right.\right.\\
&\quad\left.\left.\left.-2(4\alpha(2\alpha(5\alpha+21)+27)+9))+2\right)-1\right)\phi_1^4-\left(\phi_2+1\right)^2\left(\alpha(\alpha(10\alpha+51)+36)+\phi_2\left(\phi_2\left(-6(2\alpha+3)\phi_2\alpha^2+6(\alpha(14\alpha+73)+52)\alpha\right.\right.\right.\\
&\quad\left.\left.\left.+(\alpha(\alpha(10\alpha+51)+36)+3)\phi_2^2+26\right)-6\alpha^2(2\alpha+3)\right)+3\right)\phi_1^2+2\phi_2\left(\phi_2+1\right)^4\left((\alpha(\alpha(38\alpha+177)+120)+10)\phi_2^2-2(\alpha(\alpha(10\alpha+39)+24)+2)\phi_2\right.\\
&\quad\left.\left.+\alpha(\alpha(38\alpha+177)+120)+10\right)\right))+z\left(\alpha^4(4\alpha+1)\left(4\phi_2\left(\phi_2\left(5\phi_2+6\right)+5\right)-\phi_1^2\left(\left(\phi_2-18\right)\phi_2+1\right)\right)\right)\\
&+z^2\left(4\alpha^3\left((\alpha(5\alpha+8)+\phi_2\left(-\alpha(47\alpha+64)+(\alpha(5\alpha+8)+1)\phi_2-8\right)+1\right)\phi_1^4\right.\\
&\quad+\left((\alpha(5\alpha+8)+1)\phi_2^4-2(\alpha(25\alpha+32)+4)\phi_2^3-2(\alpha+1)(7\alpha+1)\phi_2^2-2(\alpha(25\alpha+32)+4)\phi_2\right.\\
&\quad\left.\left.+\alpha(5\alpha+8)+1\right)\phi_1^2-\phi_2\left(\phi_2+1\right)^2\left(\alpha(57\alpha+80)+\phi_2\left(2\alpha(9\alpha+16)+(\alpha(57\alpha+80)+10)\phi_2+4\right)+10\right)\right)\\
&\left.-4\phi_2\alpha^5\right)
\end{aligned}
$$

$$
\begin{aligned}
&-M^3\left(z^5\left(2\alpha(-5\alpha(\alpha+1)(\alpha+3)-2)\left(\phi_1^2-4\phi_2\right)\left(\phi_2-1\right)^2\left(\phi_1^2-\left(\phi_2+1\right)^2\right)^4\right)\right.\\
&+z^4\left(16\alpha^2(\alpha+1)(\alpha(\alpha+5)+1)\left(\phi_1^2-\left(\phi_2+1\right)^2\right)^2\left(\left(\left(\phi_2-3\right)\phi_2+1\right)\phi_1^4+\left(\phi_2\left(\phi_2\left(\left(\phi_2-2\right)\phi_2+10\right)-2\right)+1\right)\phi_1^2\right.\right.\\
&\quad\left.\left.-\phi_2\left(\phi_2+1\right)^2\left(\phi_2\left(5\phi_2-6\right)+5\right)\right)\right)\\
&+z^3\left(2\alpha^3\left(-2\left(\alpha(2\alpha(\alpha+9)+21)+\phi_2\left((\alpha-9)\alpha(\alpha+3)+\phi_2\left(2\alpha(\alpha(11\alpha+120)+168)+\phi_2\left((\alpha-9)\alpha(\alpha+3)+(\alpha(2\alpha(\alpha+9)+21)+4)\phi_2\right.\right.\right.\right.\right.\right.\\
&\quad\left.\left.\left.\left.\left.\left.-8\right)+72\right)-8\right)+4\right)\phi_1^4-\left(\phi_2+1\right)^2\left(\left(\alpha(2\alpha(\alpha+15)+51)+12\right)\phi_2^4-6\alpha(\alpha(\alpha+6)+3)\phi_2^3+2(\alpha(2\alpha(4\alpha+63)+219)+52)\phi_2^2-6\alpha(\alpha(\alpha+6)+3)\phi_2\right.\right.\\
&\quad\left.\left.+\alpha(2\alpha(\alpha+15)+51)+12\right)\phi_1^2-\phi_1^6\left(\left(\alpha(2\alpha(\alpha+15)+51)+12\right)\phi_2^2-14(\alpha(\alpha(\alpha+12)+18)+4)\phi_2+\alpha(2\alpha(\alpha+15)+51)+12\right)+2\phi_2\left(\phi_2+1\right)^4\right.\\
&\quad\left.\cdot\left(\left(3\alpha(\alpha(3\alpha+38)+59)+40\right)\phi_2^2-2(3\alpha(\alpha(\alpha+10)+13)+8)\phi_2+3\alpha(\alpha(3\alpha+38)+59)+40\right)\right)\\
&+z^2\left(8\alpha^4\left(\left(\alpha(\alpha+5)+\phi_2\left(-\alpha(15\alpha+47)+(\alpha(\alpha+5)+2)\phi_2-16\right)+2\right)\phi_1^4+\left(\left(\alpha(\alpha+5)+2\right)\phi_2^4-2(\alpha(9\alpha+25)+8)\phi_2^3-2(\alpha+2)(3\alpha+1)\phi_2^2\right.\right.\\
&\quad\left.\left.-2(\alpha(9\alpha+25)+8)\phi_2+\alpha(\alpha+5)+2\right)\phi_1^2-\phi_2\left(\phi_2+1\right)^2\left(\left(\alpha(17\alpha+57)+20\right)\phi_2^2+2(\alpha(\alpha+9)+4)\phi_2+\alpha(17\alpha+57)+20\right)\right)\\
&+z\left(4\alpha^5\left(\phi_2\left(\left(27\alpha+20\right)\phi_2^2+6(5\alpha+4)\phi_2+27\alpha+20\right)-\phi_1^2\left(\alpha+\phi_2\left(-25\alpha+(\alpha+1)\phi_2-18\right)+1\right)\right)\right.\\
&\left.\left.-16\phi_2\alpha^6\right)\right)
\end{aligned}
$$

$$
\begin{aligned}
&-M^2\left(z^5\left(2\alpha^2(-5\alpha(\alpha+2)-3)\left(\phi_1^2-4\phi_2\right)\left(\phi_2-1\right)^2\left(\phi_1^2-\left(\phi_2+1\right)^2\right)^4\right)\right.\\
&+z^4\left(8\alpha^3(\alpha(3\alpha+8)+3)\left(\phi_1^2-\left(\phi_2+1\right)^2\right)^2\left(\left(\left(\phi_2-3\right)\phi_2+1\right)\phi_1^4+\left(\phi_2\left(\phi_2\left(\left(\phi_2-2\right)\phi_2+10\right)-2\right)+1\right)\phi_1^2-\phi_2\left(\phi_2+1\right)^2\left(\phi_2\left(5\phi_2-6\right)+5\right)\right)\right)\\
&+z^3\left(2\alpha^4\left(-2\left(6\alpha(\alpha+3)+\phi_2\left(3(\alpha-3)(\alpha+1)+\phi_2\left((6\alpha(\alpha+3)+7)\phi_2^2+3(\alpha-3)(\alpha+1)\phi_2+6\alpha(11\alpha+40)+112\right)\right)+7\right)\phi_1^4\right.\\
&\quad-\left(\phi_2+1\right)^2\left(6\alpha(\alpha+5)+\phi_2\left(\phi_2\left((6\alpha(\alpha+5)+17)\phi_2^2-6(3\alpha(\alpha+2)+1)\phi_2+12\alpha(4\alpha+21)+146\right)-6(3\alpha(\alpha+2)+1)\right)+17\right)\phi_1^2\\
&\quad-\phi_1^6\left((6\alpha(\alpha+5)+17)\phi_2^2-42(\alpha(\alpha+4)+2)\phi_2+6\alpha(\alpha+5)+17\right)+2\phi_2\left(\phi_2+1\right)^4\\
&\quad\left.\cdot\left((3\alpha(9\alpha+38)+59)\phi_2^2-2\left(9\alpha^2+30\alpha+13\right)\phi_2+3\alpha(9\alpha+38)+59\right)\right)\\
&+z^2\left(4\alpha^5\left(\left(4\alpha+\phi_2\left(-\alpha(7\alpha+60)+(4\alpha+5)\phi_2-47\right)+5\right)\phi_1^4+\left((4\alpha+5)\phi_2^4-2(\alpha(5\alpha+36)+25)\phi_2^3-2(2\alpha(\alpha+6)+7)\phi_2^2\right.\right.\\
&\quad\left.\left.-2(\alpha(5\alpha+36)+25)\phi_2+4\alpha+5\right)\phi_1^2-\phi_2\left(\phi_2+1\right)^2\left(\alpha(7\alpha+68)+\phi_2\left(-2(\alpha-4)\alpha+(\alpha(7\alpha+68)+57)\phi_2+18\right)+57\right)\right)\\
&+z\left(4\alpha^6\left(\phi_2\left(\left(14\alpha+27\right)\phi_2^2+6(2\alpha+5)\phi_2+14\alpha+27\right)-\phi_1^2\left(\phi_2\left(-14\alpha+\phi_2-25\right)+1\right)\right)\right.\\
&\left.\left.-20\phi_2\alpha^7\right)\right)
\end{aligned}
$$

$$
\begin{aligned}
&-M^1\left(z^5\left((-5\alpha-4)\alpha^3\left(\phi_1^2-4\phi_2\right)\left(\phi_2-1\right)^2\left(\phi_1^2-\left(\phi_2+1\right)^2\right)^4\right)+z\left(8\alpha^7\phi_2\left((\alpha+7)\phi_1^2+\alpha+\phi_2\left((\alpha+7)\phi_2+6\right)+7\right)\right)\right.\\
&+z^4\left(16\alpha^4(\alpha+1)\left(\phi_1^2-\left(\phi_2+1\right)^2\right)^2\left(\left(\left(\phi_2-3\right)\phi_2+1\right)\phi_1^4+\left(\phi_2\left(\phi_2\left(\left(\phi_2-2\right)\phi_2+10\right)-2\right)+1\right)\phi_1^2-\phi_2\left(\phi_2+1\right)^2\left(\phi_2\left(5\phi_2-6\right)+5\right)\right)\right)\\
&+z^3\left(4\alpha^5\left(-\left(6(\alpha+1)+\phi_2\left(3\alpha+\phi_2\left(66\alpha+\phi_2\left(3\alpha+6(\alpha+1)\phi_2-2\right)+80\right)-2\right)\right)\phi_1^4\right.\right.\\
&\quad-\left(\phi_2+1\right)^2\left(3\alpha+\phi_2\left(-9\alpha+\phi_2\left(6(4\alpha+7)+\phi_2\left(-9\alpha+(3\alpha+5)\phi_2-6\right)\right)-6\right)+5\right)\phi_1^2\\
&\quad\left.\left.-\phi_1^6\left((3\alpha+5)\phi_2^2-7(3\alpha+4)\phi_2+3\alpha+5\right)+\phi_2\left(\phi_2+1\right)^4\left((27\alpha+38)\phi_2^2-2(9\alpha+10)\phi_2+27\alpha+38\right)\right)\right)\\
&+z^2\left(8\alpha^6\left(\left(\phi_2\left(-7\alpha+\phi_2-15\right)+1\right)\phi_1^4+\left(\phi_2\left(\phi_2\left(\phi_2^2-2(5\alpha+9)\phi_2-4\alpha-6\right)-2(5\alpha+9)\right)+1\right)\phi_1^2\right.\right.\\
&\quad\left.\left.-\phi_2\left(\phi_2+1\right)^2\left(7\alpha+\phi_2\left(-2\alpha+(7\alpha+17)\phi_2+2\right)+17\right)\right)\right.\\
&\left.-8\phi_2\alpha^8\right)
\end{aligned}
$$

$$
\begin{aligned}
&+\alpha^4 z\left(-8\phi_2\left(\phi_1^2+\phi_2^2+1\right)\alpha^4+4z\phi_2\left(7\phi_1^4+2\left(\phi_2\left(5\phi_2+2\right)+5\right)\phi_1^2+\left(\phi_2+1\right)^2\left(\phi_2\left(7\phi_2-2\right)+7\right)\right)\alpha^3+4z^2\left(\left(\left(\phi_2-7\right)\phi_2+1\right)\phi_1^6\right.\right.\\
&+\left(2\phi_2^4+\phi_2^3+22\phi_2^2+\phi_2+2\right)\phi_1^4+\left(\phi_2+1\right)^2\left(\phi_2\left(\phi_2\left(\left(\phi_2-3\right)\phi_2+8\right)-3\right)+1\right)\phi_1^2-3\phi_2\left(\phi_2+1\right)^4\left(\phi_2\left(3\phi_2-2\right)+3\right)\alpha^2\\
&-4z^3\left(\phi_1^2-\left(\phi_2+1\right)^2\right)^2\left(\left(\left(\phi_2-3\right)\phi_2+1\right)\phi_1^4+\left(\phi_2\left(\phi_2\left(\left(\phi_2-2\right)\phi_2+10\right)-2\right)+1\right)\phi_1^2-\phi_2\left(\phi_2+1\right)^2\left(\phi_2\left(5\phi_2-6\right)+5\right)\right)\alpha\\
&\left.+z^4\left(\phi_2-1\right)^2\left(\phi_1^2-4\phi_2\right)\left(\phi_1^2-\left(\phi_2+1\right)^2\right)^4\right)
\end{aligned}
$$

A.3.2 Polynomial for VAR(2)

$$4M^9z^5\theta_2$$

$$+M^8\left(20z^5\theta_2-\alpha z^4\left(\left(\theta_2\left(\theta_2+18\right)+1\right)\theta_1^2+4\theta_2\left(\theta_2\left(5\theta_2-6\right)+5\right)\right)\right)$$

$$
\begin{aligned}
&+M^7\left(-4\left(\alpha^2-10\right)\theta_2 z^5-4\alpha\left(\left(\theta_2\left(\theta_2+18\right)+1\right)\theta_1^2+4\theta_2\left(\theta_2\left(5\theta_2-6\right)+5\right)\right)z^4\right.\\
&\left.+4\alpha^2\left(\left(\theta_2\left(\theta_2+8\right)+1\right)\theta_1^4+\left(\theta_2\left(\theta_2\left(\theta_2\left(\theta_2+8\right)-2\right)+8\right)+1\right)\theta_1^2+2\left(\theta_2-1\right)^2\theta_2\left(\theta_2\left(5\theta_2-2\right)+5\right)\right)z^3\right)\\
&+M^6\left(-20z^5\left(\alpha^2-2\right)\theta_2+2z^4\alpha\left(\left(\alpha^2+\theta_2\left(4\alpha^2+\left(\alpha^2-3\right)\theta_2-54\right)-3\right)\theta_1^2+6\theta_2\left(\alpha^2+\left(\alpha^2-10\right)\theta_2^2-2\left(\alpha^2-6\right)\theta_2-10\right)\right)\right.\\
&+12z^3\alpha^2\left(\left(\theta_2\left(\theta_2+8\right)+1\right)\theta_1^4+\left(\theta_2\left(\theta_2\left(\theta_2\left(\theta_2+8\right)-2\right)+8\right)+1\right)\theta_1^2+2\left(\theta_2-1\right)^2\theta_2\left(\theta_2\left(5\theta_2-2\right)+5\right)\right)\\
&\left.-2z^2\alpha^3\left(\left(\theta_2\left(3\theta_2+14\right)+3\right)\theta_1^6+2\left(\theta_2\left(\theta_2\left(\theta_2\left(\theta_2+2\right)+18\right)+2\right)+1\right)\theta_1^4+\left(\theta_2-1\right)^2\left(3\theta_2^4+26\theta_2^2+3\right)\theta_1^2+4\left(\theta_2-1\right)^4\theta_2\left(\theta_2\left(5\theta_2+2\right)+5\right)\right)\right)\\
&+M^5\left(20z^5\left(1-2\alpha^2\right)\theta_2+4z^4\alpha\left(\left(2\alpha^2+\theta_2\left(8\alpha^2+\left(2\alpha^2-1\right)\theta_2-18\right)-1\right)\theta_1^2+4\theta_2\left(3\alpha^2+\theta_2\left(-6\alpha^2+\left(3\alpha^2-5\right)\theta_2+6\right)-5\right)\right)\right.\\
&+z^3\left(12\alpha^2\left(\left(\theta_2\left(\theta_2+8\right)+1\right)\theta_1^4+\left(\theta_2\left(\theta_2\left(\theta_2\left(\theta_2+8\right)-2\right)+8\right)+1\right)\theta_1^2+2\left(\theta_2-1\right)^2\theta_2\left(\theta_2\left(5\theta_2-2\right)+5\right)\right)\right.\\
&\left.-4\alpha^4\left(\left(\theta_2^2+\theta_2+1\right)\theta_1^4+\left(\theta_2-1\right)^2\left(\theta_2^2+1\right)\theta_1^2+3\left(\theta_2-1\right)^4\theta_2\right)\right)\\
&-4z^2\alpha^3\left(\left(\theta_2\left(3\theta_2+14\right)+3\right)\theta_1^6+2\left(\theta_2\left(\theta_2\left(\theta_2\left(\theta_2+2\right)+18\right)+2\right)+1\right)\theta_1^4+\left(\theta_2-1\right)^2\left(3\theta_2^4+26\theta_2^2+3\right)\theta_1^2+4\left(\theta_2-1\right)^4\theta_2\left(\theta_2\left(5\theta_2+2\right)+5\right)\right)\\
&\left.+4z\alpha^4\left(\theta_1^2-\left(\theta_2-1\right)^2\right)^2\left(\left(\theta_2\left(\theta_2+3\right)+1\right)\theta_1^4+\left(\theta_2\left(\theta_2\left(\theta_2\left(\theta_2+2\right)+10\right)+2\right)+1\right)\theta_1^2+\left(\theta_2-1\right)^2\theta_2\left(\theta_2\left(5\theta_2+6\right)+5\right)\right)\right)\\
&+M^4\left(4z^5\left(1-10\alpha^2\right)\theta_2\right.\\
&+z^4\left(4\alpha\theta_2\left(18\alpha^2+\theta_2\left(-36\alpha^2+\left(18\alpha^2-5\right)\theta_2+6\right)-5\right)-\alpha\theta_1^2\left(\alpha^4-12\alpha^2+\left(\alpha^4-12\alpha^2+1\right)\theta_2^2-2\left(\alpha^4+24\alpha^2-9\right)\theta_2+1\right)\right)\\
&+4z^3\left(\left(\theta_2\left(\theta_2+8\right)+1\right)\theta_1^4+\left(\theta_2\left(\theta_2\left(\theta_2\left(\theta_2+8\right)-2\right)+8\right)+1\right)\theta_1^2+2\left(\theta_2-1\right)^2\theta_2\left(\theta_2\left(5\theta_2-2\right)+5\right)\right.\\
&\left.-3\alpha^2\left(\left(\theta_2^2+\theta_2+1\right)\theta_1^4+\left(\theta_2-1\right)^2\left(\theta_2^2+1\right)\theta_1^2+3\left(\theta_2-1\right)^4\theta_2\right)\right)\alpha^2+z^2\left(2\alpha^5\left(\theta_1^2-\left(\theta_2-1\right)^2\right)^2\left(\left(\theta_2^2+1\right)\theta_1^2+2\left(\theta_2-1\right)^2\theta_2\right)\right.\\
&\left.-2\alpha^3\left(\left(\theta_2\left(3\theta_2+14\right)+3\right)\theta_1^6+2\left(\theta_2\left(\theta_2\left(\theta_2\left(\theta_2+2\right)+18\right)+2\right)+1\right)\theta_1^4+\left(\theta_2-1\right)^2\left(3\theta_2^4+26\theta_2^2+3\right)\theta_1^2+4\left(\theta_2-1\right)^4\theta_2\left(\theta_2\left(5\theta_2+2\right)+5\right)\right)\right)\\
&+4z\left(\theta_1^2-\left(\theta_2-1\right)^2\right)^2\left(\left(\theta_2\left(\theta_2+3\right)+1\right)\theta_1^4+\left(\theta_2\left(\theta_2\left(\theta_2\left(\theta_2+2\right)+10\right)+2\right)+1\right)\theta_1^2+\left(\theta_2-1\right)^2\theta_2\left(\theta_2\left(5\theta_2+6\right)+5\right)\right)\alpha^4\\
&\left.-\left(\theta_1^2-\left(\theta_2-1\right)^2\right)^4\left(\theta_2+1\right)^2\left(\theta_1^2+4\theta_2\right)\alpha^5\right)
\end{aligned}
$$

$$+M^{3}\left(-20z^{5}\theta_{2}\alpha^{2}+4z^{4}\left(12\left(\theta_{2}-1\right)^{2}\theta_{2}-\theta_{1}^{2}\left(\alpha^{2}+\left(\alpha^{2}-2\right)\theta_{2}^{2}-2\left(\alpha^{2}+4\right)\theta_{2}-2\right)\right)\alpha^{3}\right.$$
$$\left.-12z^{3}\left(\left(\theta_{2}^{2}+\theta_{2}+1\right)\theta_{1}^{4}+\left(\theta_{2}-1\right)^{2}\left(\theta_{2}^{2}+1\right)\theta_{1}^{2}+3\left(\theta_{2}-1\right)^{4}\theta_{2}\right)\alpha^{4}4z^{2}\left(\theta_{1}^{2}-\left(\theta_{2}-1\right)^{2}\right)^{2}\left(\left(\theta_{2}^{2}+1\right)\theta_{1}^{2}+2\left(\theta_{2}-1\right)^{2}\theta_{2}\right)\alpha^{5}\right)$$

$$+M^{2}\left(-4z^{5}\theta_{2}\alpha^{2}+2z^{4}\left(\left(-3\alpha^{2}+\theta_{2}\left(6\alpha^{2}+\left(1-3\alpha^{2}\right)\theta_{2}+4\right)+1\right)\theta_{1}^{2}+6\left(\theta_{2}-1\right)^{2}\theta_{2}\right)\alpha^{3}\right.$$
$$\left.-4z^{3}\left(\left(\theta_{2}^{2}+\theta_{2}+1\right)\theta_{1}^{4}+\left(\theta_{2}-1\right)^{2}\left(\theta_{2}^{2}+1\right)\theta_{1}^{2}+3\left(\theta_{2}-1\right)^{4}\theta_{2}\right)\alpha^{4}2z^{2}\left(\theta_{1}^{2}-\left(\theta_{2}-1\right)^{2}\right)^{2}\left(\left(\theta_{2}^{2}+1\right)\theta_{1}^{2}+2\left(\theta_{2}-1\right)^{2}\theta_{2}\right)\alpha^{5}\right)$$

$$-4M^{1}z^{4}\alpha^{5}\theta_{1}^{2}\left(\theta_{2}-1\right)^{2}$$

$$-\alpha^{5}z^{4}\theta_{1}^{2}\left(\theta_{2}-1\right)^{2}$$

A.3.3 Argument of Roots for a VAR(2) Distribution Generating Polynomial

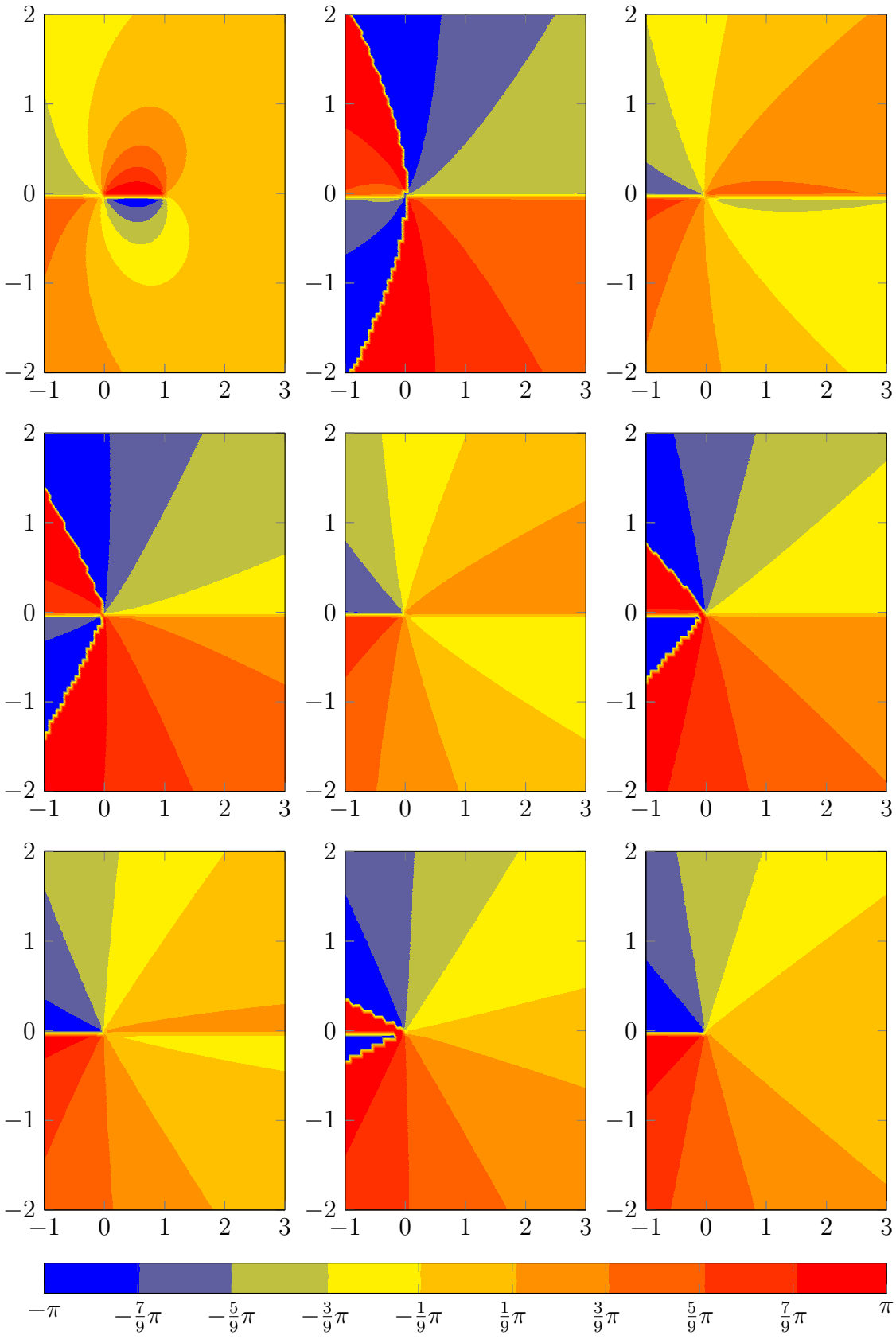

Figure A.1: arg of complex-valued root of polynomial $M^9 - (x + \mathrm{i}y)$.

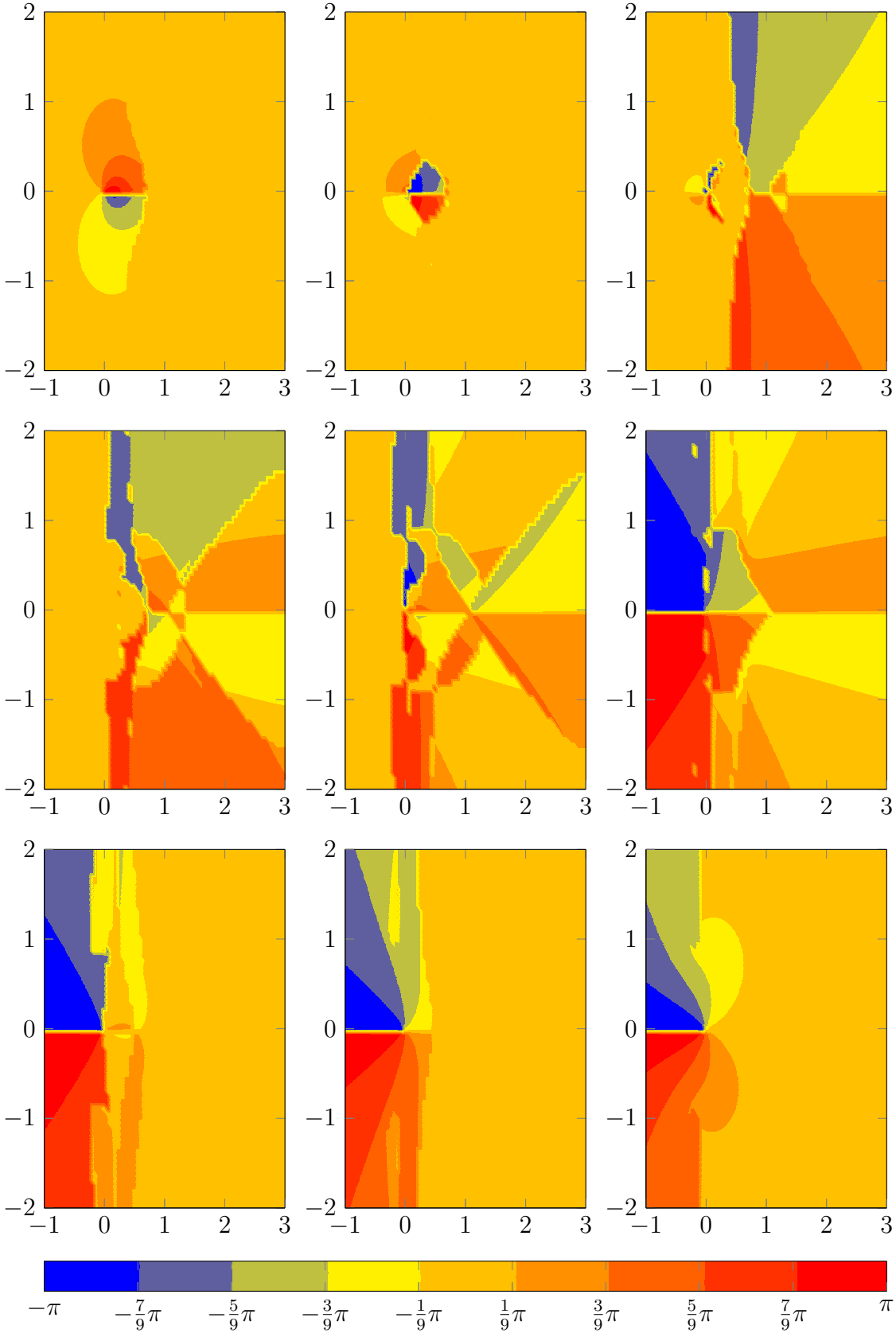

Figure A.2: arg of complex-valued root of VAR(2) polynomial for parameters $\theta_1 = 0.5$, $\theta_2 = 0.25$ and $\alpha = 4$.

A.3.4 Fit of Spectral Density and ESD

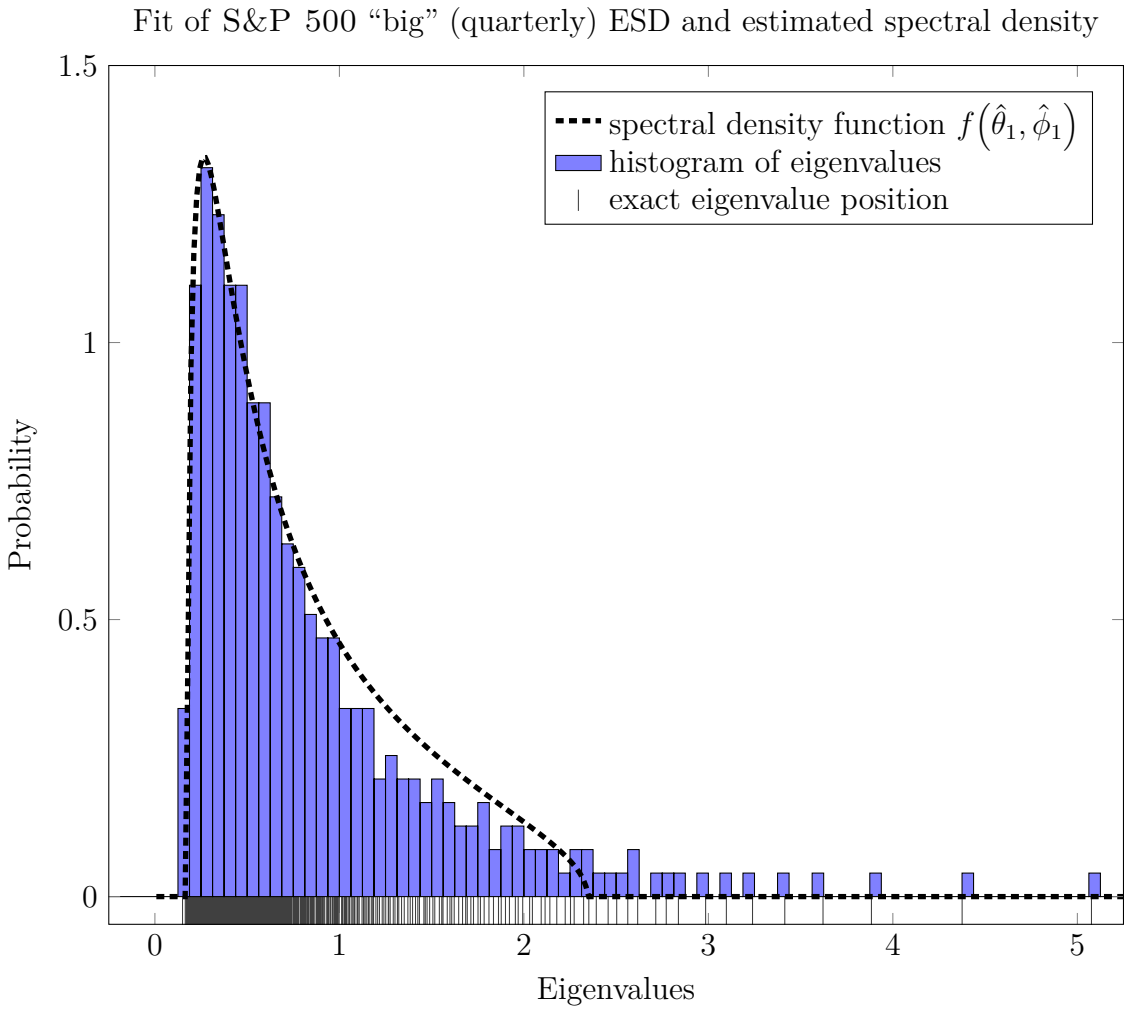

Figure A.3: Example of model fit for VARMA$(1,1)$ random matrix model, ESD of S&P 500 »big« data set (quarterly) for the median of all eigenvalues and VARMA$(1,1)$ spectral density function with estimated parameters $\hat{\theta}_1$ and $\hat{\phi}_1$.

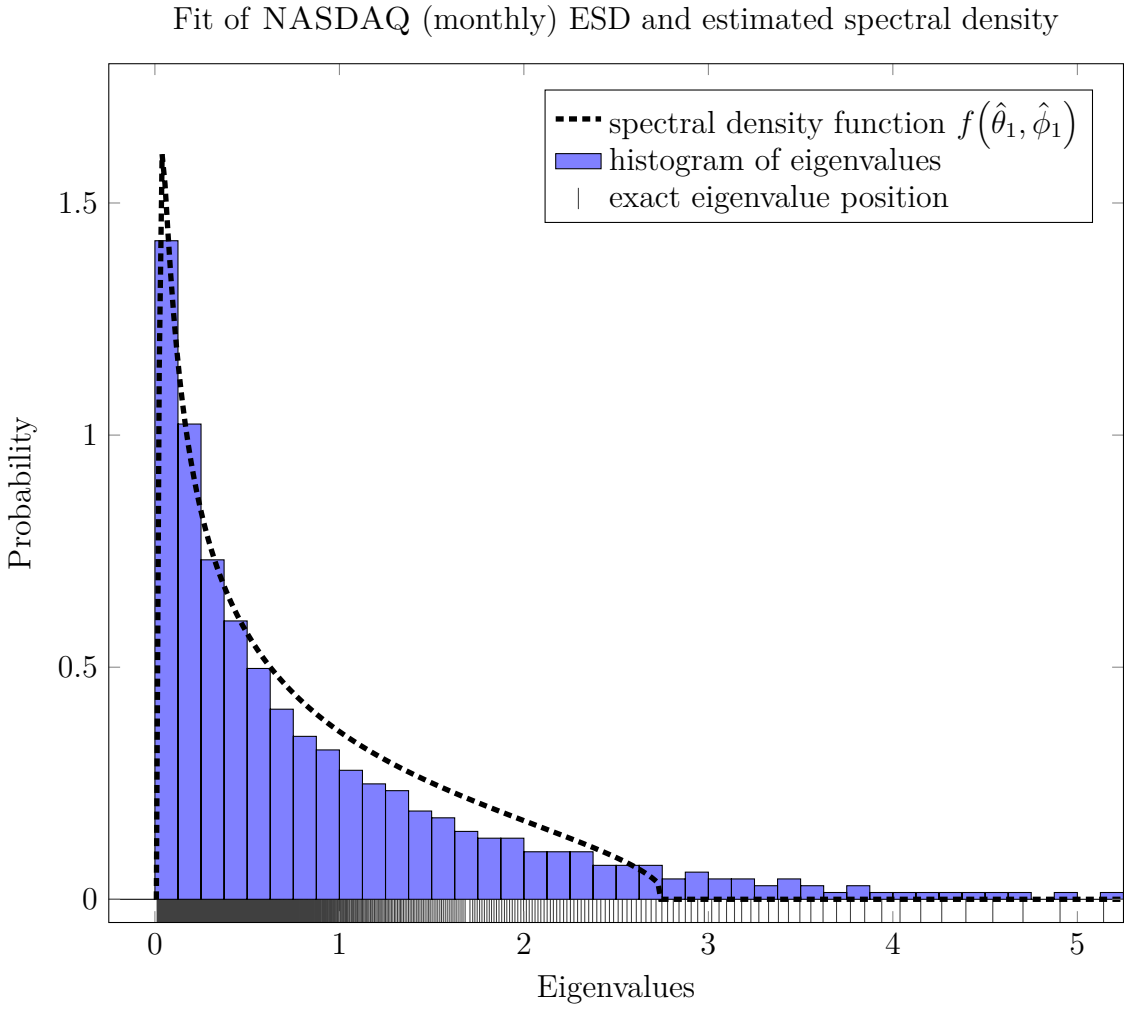

Figure A.4: Example of model fit for VARMA$(1,1)$ random matrix model, ESD of NASDAQ data set (monthly) for the mean of all eigenvalues and VARMA$(1,1)$ spectral density function with estimated parameters $\hat{\theta}_1$ and $\hat{\phi}_1$.

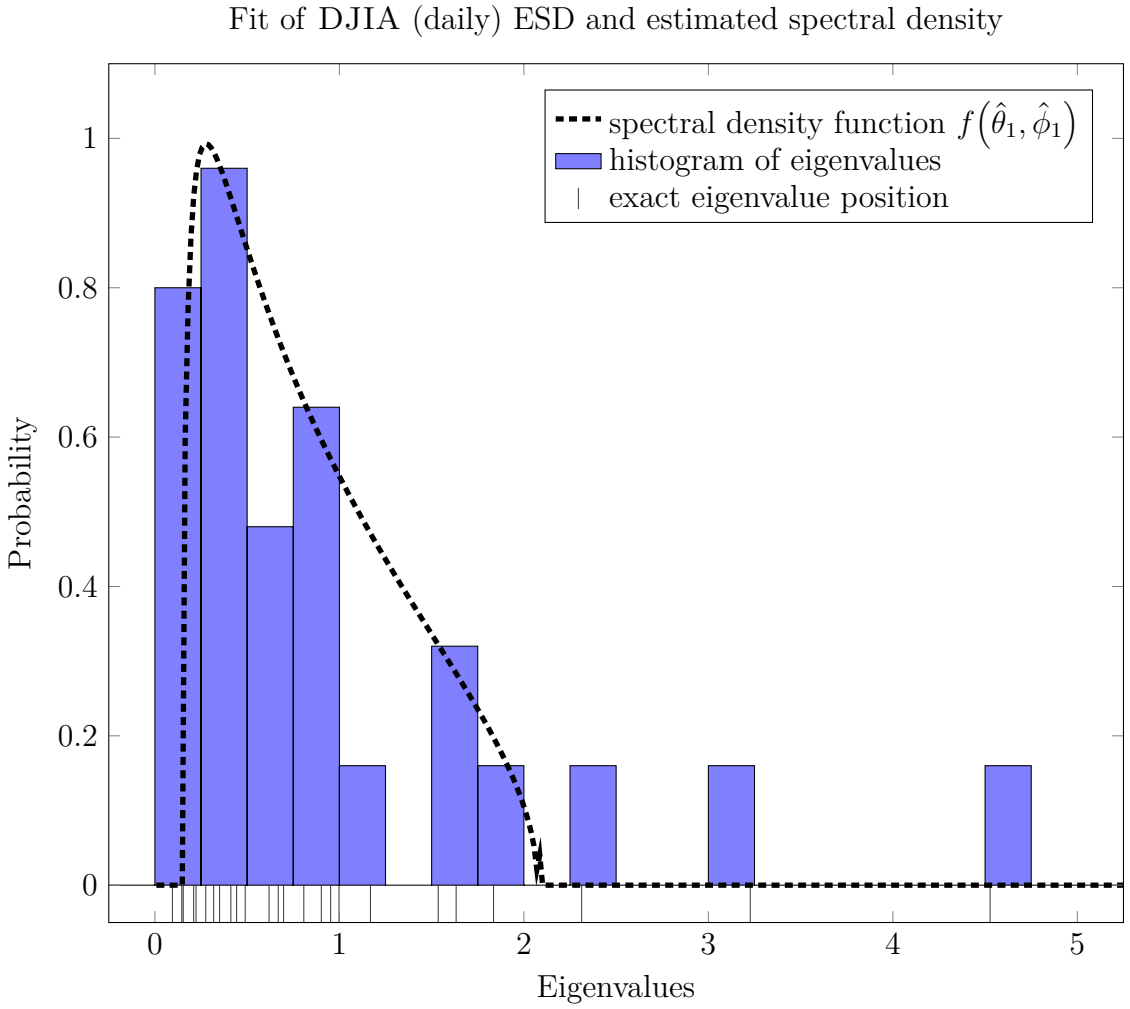

Figure A.5: Example of model fit for VARMA$(1,1)$ random matrix model, ESD of DJIA data set (daily) for the last complete observation and VARMA$(1,1)$ spectral density function with estimated parameters $\hat{\theta}_1$ and $\hat{\phi}_1$.

B Program Code

B.1 Mathematica Code

B.1.1 Code for Generating DGP

In[1]:=

```
ClearAll[FactorPoly,PolySquareFree,PolyReduce,NormalizePoly,SimplifyRoot,PowerContract,showPoly];

FactorPoly[poly_,var_:M]:=Collect[FactorList[Numerator[Together[Numerator[Rationalize[ExpandAll[poly]]]
  //.{e1_+e2_ Power[e3_,r_Rational]:→With[{d=Denominator[r]},-(-e1)^d+(e2 e3^r)^d]}]][[-1,1]],var];

NormalizePoly[poly_,polyarg_:M]:=Collect[Numerator[Together[poly]],polyarg];

PolySquareFree[poly_,polyarg_:M]:=Module[{
  num=(PowerContract[Expand[Numerator[Together[-(-(Replace[poly,{a_Plus:→Most@a}]))^2
    +(Replace[poly,{a_Plus:→Last@a}])^2]]]//.powerdistr]//.powersimplify),
  denom=(PowerContract[Expand[Denominator[Together[-(-(Replace[poly,{a_Plus:→Most@a}]))^2
    +(Replace[poly,{a_Plus:→Last@a}])^2]]]//.powerdistr]//.powersimplify)},
  NormalizePoly[Simplify[Cancel[Together[(num/denom)/.sumpowred]]],polyarg]];

PolyReduce[poly_,var_:M]:=Module[{termlist,explist,reslist,vars,exp},
  termlist=With[{Mlist=CoefficientList[poly,var]},Table[Part[Mlist,i],{i,Length[Mlist]}]];
  explist=Range[Length[termlist]]-1;
  reslist=Parallelize[Table[Power[var,Part[explist,k]]*(With[{termpoly=Part[termlist,k]},
    vars=Variables[termpoly];
    exp=Exponent[termpoly,vars];
    vars=Take[vars,Flatten[Position[exp,Min[exp]]]];
    reslist=Map[Apply[Plus,With[{mlist=MonomialList[termpoly,#]},
      Table[FullSimplify[ExpandAll[Part[mlist,i]]],{i,Length[mlist]}]]]&,vars];
    Part[reslist,First@Flatten[{Ordering[Map[LeafCount,reslist]]}]]
  ]),{k,Length[termlist]}]];
  Apply[Plus,reslist]
];
SimplifyRoot=PowerExpand[Factor//@#]//.{1/Sqrt[x_]:→Sqrt[1/x],Sqrt[x_] Sqrt[y_]:→Sqrt[x y]}&;
```

```
  PowerContract[expr_]:=expr//.{m_^q_ n_^q_:→(m n)^q/;!IntegerQ[m]&&!IntegerQ[n],m_^q_
n_^p_:→(m/n)^q/;q≥0&&p==-q&&!IntegerQ[m]&&!IntegerQ[n]};
showPoly=PolynomialForm[#,TraditionalOrder→True]&;
```

In[2]:=
```
ClearAll[fracdistr,powerdistr,powersimplify,distr,square,sumpowred,oneSide,PowerContract,CollectPoly,showPoly];

fracdistr={a_/b_+c_/d_:→FullSimplify@Numerator[a] FullSimplify@Denominator[b]
  /(FullSimplify@Denominator[a]FullSimplify@Numerator[b])
  +FullSimplify@Numerator[c] FullSimplify@Denominator[d]
  /(FullSimplify@Denominator[c] FullSimplify@Numerator[d])};

powerdistr={a_ Power[b_,exp_Rational]+c_ Power[b_,exp_Rational]:→(a+c)Power[b,exp]};
powersimplify={Power[a_,exp_Rational]:→Power[Simplify[ExpandAll[a]],exp]};
distr={a_ b_+c_ b_:→(a+c)b,b_ a_+b_ c_:→b(a+c),(2+2 M) a_+(-2-2 M) c_:→(2+2 M)(a-c)};
square={e1_+e2_ e3_^r_Rational:→With[{d=Denominator[r]},-(-e1)^d+(e2 e3^r)^d]};
sumpowred={(a_+b_ Power[c_,exp_Rational]) Power[d_,-1]:→-(-(a/d))^(1/exp)
  +((b Power[c,exp])/d)^(1/exp)};
oneSide=(Head[#][Subtract@@#,0]&);
PowerContract[expr_]:=expr//.{m_^q_ n_^q_:→(m n)^q/;!IntegerQ[m]&&!IntegerQ[n],m_^q_
n_^p_:→(m/n)^q/;q≥0&&p==-q&&!IntegerQ[m]&&!IntegerQ[n]};
CollectPoly[poly_,polyarg_]:=Nest[Simplify[PolynomialQuotient[Numerator[#1],Denominator[#1],#2]]
  +Simplify[PolynomialRemainder[Numerator[#1],Denominator[#1],#2]/Denominator[#1]]&[#,polyarg]&,
  FullSimplify[poly,ComplexityFunction→
  LeafCount[#]+2^10*StringCount[ToString[#,InputForm],ToString[poly]]&],3];
showPoly=PolynomialForm[#,TraditionalOrder→True]&;
```

Bibliography

[1] Y. Aït-Sahalia, P. Mykland, and L. Zhang. »How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise«. In: *Review of Financial Studies* 18.2 (2005), pp. 351–416. DOI: 10.1093/rfs/hhi016.

[2] R. Allez and J.-P. Bouchaud. »Eigenvector dynamics: theory and some applications«. In: *ArXiv e-prints* (Aug. 2011).

[3] R. Allez and J.-P. Bouchaud. »Eigenvector dynamics: General theory and some applications«. In: *Phys. Rev. E* 86 (4 Oct. 2012), p. 046202. DOI: 10.1103/PhysRevE.86.046202.

[4] R. Allez and J.-P. Bouchaud. »Eigenvector dynamics under free addition«. In: *Random Matrices: Theory and Applications* 03.03 (2014), pp. 1450010-1–1450010-17. DOI: 10.1142/S2010326314500105.

[5] R. Allez, J. Bun, and J.-P. Bouchaud. »The eigenvectors of Gaussian matrices with an external source«. In: *ArXiv e-prints* (Dec. 2014).

[6] T. Andersen and T. Bollerslev. »Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-Run in High Frequency Returns«. In: *The Journal of Finance* 52.3 (1997), pp. 975–1005. ISSN: 1540-6261. DOI: 10.1111/j.1540-6261.1997.tb02722.x.

[7] E. Andreou and E. Ghysels. »Rolling-Sample Volatility Estimators: Some New Theoretical, Simulation, and Empirical Results«. In: *Journal of Business & Economic Statistics* 20.3 (2002), pp. 363–376. ISSN: 07350015.

[8] J. Bai. »Inferential Theory for Factor Models of Large Dimensions«. In: *Econometrica* 71.1 (2003), pp. 135–171. ISSN: 00129682, 14680262.

[9] Z. Bai and J. Silverstein. *Spectral Analysis of Large Dimensional Random Matrices*. 2nd. Springer Series in Statistics. Springer, Dec. 2009, p. 552. ISBN: 9781441906601.

[10] R. Baillie, C.-F. Chung, and M. Tieslau. »Analysing inflation by the fractionally integrated ARFIMA–GARCH model«. In: *Journal of Applied Econometrics* 11.1 (1996), pp. 23–40. ISSN: 1099-1255. DOI: 10.1002/(SICI)1099-1255(199601)11:1<23::AID-JAE374>3.0.CO;2-M.

[11] F. Bandiy and J. Russell. »Microstructure Noise, Realized Variance, and Optimal Sampling«. In: *The Review of Economic Studies* 75.2 (2008), pp. 339–369. DOI: 10.1111/j.1467-937X.2008.00474.x.

[12] F. Benaych-Georges. »Rectangular random matrices, entropy, and Fisher's information«. In: *J. Operator Theory* 62.2 (2009), pp. 371–419. ISSN: 0379-4024.

[13] F. Benaych-Georges. »Rectangular random matrices, related convolution«. In: *Probability Theory and Related Fields* 144.3–4 (2009), pp. 471–515. ISSN: 0178-8051. DOI: 10.1007/s00440-008-0152-z.

[14] C. Biely and S. Thurner. »Random matrix ensembles of time-lagged correlation matrices: derivation of eigenvalue spectra and analysis of financial time-series«. In: *Quantitative Finance* 8.7 (2008), pp. 705–722. DOI: 10.1080/14697680701691477.

[15] F. Black. »Noise«. In: *The Journal of Finance* 41.3 (1986), pp. 529–543. ISSN: 00221082.

[16] C. Bordenave and A. Guionnet. »Localization and delocalization of eigenvectors for heavy-tailed random matrices«. In: *Probability Theory and Related Fields* 157.3 (2013), pp. 885–953. ISSN: 1432-2064. DOI: 10.1007/s00440-012-0473-9.

[17] J.-P. Bouchaud and M. Potters. *Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management*. Cambridge University Press, 2003. ISBN: 9780521819169.

[18] N. Bourbaki. *Elements of Mathematics: Integration*. Vol. Chapters 1–6. Actualités scientifiques et industrielles. Hermann, 1956.

[19] G. Box and. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day series in Time series analysis. Holden-Day, 1971.

[20] P. Brockwell and R. Davis. *Time Series: Theory and Methods*. 2nd ed. Springer, 2009. ISBN: 9781441904003.

[21] J. Bun, J.-P. Bouchaud, and M. Potters. »On the overlaps between eigenvectors of correlated random matrices«. In: *ArXiv e-prints* (Mar. 2016).

[22] Z. Burda, A. Görlich, et al. »Signal and noise in correlation matrix«. In: *Physica A: Statistical Mechanics and its Applications* 343 (2004), pp. 295–310. ISSN: 0378-4371. DOI: 10.1016/j.physa.2004.05.048.

[23] Z. Burda, A. Jarosz, et al. »A random matrix approach to VARMA processes«. In: *New Journal of Physics* 12.7 (2010), p. 075036. DOI: 10.1088/1367-2630/12/7/075036.

[24] Z. Burda, J. Jurkiewicz, and B. Wacław. »Spectral moments of correlated Wishart matrices«. In: *Phys. Rev. E* 71 (2 Feb. 2005), p. 026111. DOI: 10.1103/PhysRevE.71.026111.

[25] Z. Burda, J. Grela, et al. »Unveiling the significance of eigenvectors in diffusing non-Hermitian matrices by identifying the underlying Burgers dynamics«. In: *Nuclear Physics B* 897 (2015), pp. 421–447. ISSN: 0550-3213. DOI: http://dx.doi.org/10.1016/j.nuclphysb.2015.06.002.

[26] H. Cartan. *Théorie élémentaire des fonctions analytiques d'une ou plusieurs variables complexes*. Collection Enseignement des sciences. Hermann, 1961. ISBN: 9782705652159.

[27] J. Chalker and B. Mehlig. »Eigenvector Statistics in Non-Hermitian Random Matrix Ensembles«. In: *Phys. Rev. Lett.* 81 (16 Oct. 1998), pp. 3367–3370. DOI: 10.1103/PhysRevLett.81.3367.

[28] M. Dacorogna et al. *An Introduction to High-Frequency Finance*. Elsevier Science, 2001. ISBN: 9780080499048.

[29] R. F. Engle and J. R. Russell. »Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data«. In: *Econometrica* 66.5 (1998), pp. 1127–1162. ISSN: 00129682.

[30] T. W. Epps. »Comovements in Stock Prices in the Very Short Run«. In: *Journal of the American Statistical Association* 74.366 (1979), pp. 291–298. ISSN: 01621459.

[31] L. Euler. *Introductio in analysin infinitorum*. Introductio in analysin infinitorum v. 1. apud Marcum-Michaelem Bousquet & socios, 1748.

[32] Y. Fang. »Volatility Modeling and Estimation of High-Frequency Data With Gaussian Noise«. PhD thesis. Massachusetts Institute of Technology, Sloan School of Management, 1996.

[33] C. Granger. »Long memory relationships and the aggregation of dynamic models«. In: *Journal of Econometrics* 14.2 (1980), pp. 227–238. ISSN: 0304-4076. DOI: 10.1016/0304-4076(80)90092-5.

[34] C. Granger and Z. Ding. »Varieties of long memory models«. In: *Journal of Econometrics* 73.1 (1996), pp. 61–77. ISSN: 0304-4076. DOI: 10.1016/0304-4076(95)01733-X.

[35] C. Granger and R. Joyeux. »An Introduction to Long-Memory Time Series Models and Fractional Differencing«. In: *Journal of Time Series Analysis* 1.1 (1980), pp. 15–29. ISSN: 1467-9892. DOI: 10.1111/j.1467-9892.1980.tb00297.x.

[36] T. Graves et al. »A brief history of long memory: Hurst, Mandelbrot and the road to ARFIMA«. In: *ArXiv e-prints* (June 2014).

[37] A. Haar. »Der Massbegriff in der Theorie der Kontinuierlichen Gruppen«. In: *Annals of Mathematics* 34.1 (1933), pp. 147–169. ISSN: 0003486X.

[38] J. Hamilton. *Time Series Analysis*. Princeton University Press, 1994. ISBN: 9780691042893.

[39] P. Hansen and A. Lunde. »Realized Variance and Market Microstructure Noise«. In: *Journal of Business & Economic Statistics* 24.2 (2006), pp. 127–161. DOI: 10.1198/073500106000000071.

[40] A. Hasegawa, N. Sakuma, and H. Yoshida. »Random matrices by MA models and compound free Poisson laws«. In: *Probability and Mathematical Statistics* 33.2 (2013), pp. 243–254.

[41] M. Henry and P. Zaffaroni. »Theory and Applications of Long-Range Dependence«. In: ed. by P. Doukhan, G. Oppenheim, and M. Taqqu. Birkhäuser Boston, 2003. Chap. The Long-Range Dependence Paradigm for Macroeconomics and Finance, pp. 417–438. ISBN: 9780817641689.

[42] C. Heyde and S. Liu. »Empirical realities for a minimum description risky asset model. The need for fractal Features«. In: *Journal of the Korean Mathematical Society* 38.5 (2001), pp. 1047–1069.

[43] J. Hosking. »Fractional differencing«. In: *Biometrika* 68.1 (1981), pp. 165–176. DOI: 10.1093/biomet/68.1.165.

[44] P. Huber. *Robust Statistics*. Wiley Series in Probability and Statistics. Wiley, 1981. ISBN: 9780471418054.

[45] P. J. Huber. »Robust Estimation of a Location Parameter«. In: *Ann. Math. Statist.* 35.1 (Mar. 1964), pp. 73–101. DOI: 10.1214/aoms/1177703732.

[46] H. Hurst. »Long-Term Storage Capacity of Reservoirs«. In: *Transactions of the American Society of Civil Engineers* 116.1 (1951), pp. 770–799.

[47] H. Hurst. »The Problem of Long-Term Storage in Reservoirs«. In: *International Association of Scientific Hydrology. Bulletin* 1.3 (1956), pp. 13–27. DOI: 10.1080/02626665609493644.

[48] H. Hurst. »Methods of Using Long-term storage in reservoirs.« In: *Proceedings of the Institution of Civil Engineers* 5.5 (Oct. 1956), pp. 519–543. DOI: 10.1680/iicep.1956.11503.

[49] H. Hurst. »A Suggested Statistical Model of some Time Series which occur in Nature«. In: *Nature* 180.4584 (Sept. 1957), pp. 494–494.

[50] IEEE Computer Society. »754-2008 IEEE Standard for Floating-Point Arithmetic«. In: *IEEE Std 754-2008* (Aug. 2008), pp. 1–70. DOI: 10.1109/IEEESTD.2008.4610935.

[51] M. Iri, S. Moriguti, and Y. Takasawa. »On a certain quadrature formula«. Japanese. In: *Kokyuroku of the Research Institute for Mathematical Sciences* 91 (1970). English translation in [52], pp. 82–118.

[52] M. Iri, S. Moriguti, and Y. Takasawa. »On a certain quadrature formula«. In: *Journal of Computational and Applied Mathematics* 17.1 (1987), pp. 3–20. ISSN: 0377-0427. DOI: 10.1016/0377-0427(87)90034-3.

[53] A. T. James. »Distributions of Matrix Variates and Latent Roots Derived from Normal Samples«. In: *The Annals of Mathematical Statistics* 35.2 (1964), pp. 475–501. ISSN: 00034851.

[54] R. Janik et al. »Correlations of eigenvectors for non-Hermitian random-matrix models«. In: *Phys. Rev. E* 60 (3 Sept. 1999), pp. 2699–2705. DOI: 10.1103/PhysRevE.60.2699.

[55] A. Knowles and J. Yin. »Eigenvector distribution of Wigner matrices«. In: *Probability Theory and Related Fields* 155.3 (2013), pp. 543–582. ISSN: 1432-2064. DOI: 10.1007/s00440-011-0407-y.

[56] A. Kronrod. *Nodes and weights of quadrature formulas: sixteen-place tables*. Consultants Bureau, 1965.

[57] D. P. Laurie. »Calculation of Gauss–Kronrod quadrature rules«. In: *Math. Comp.* 66.219 (1997), pp. 1133–1145. ISSN: 0025-5718. DOI: 10.1090/S0025-5718-97-00861-2.

[58] F. Lillo and D. Farmer. »The Long Memory of the Efficient Market«. In: *Studies in Nonlinear Dynamics & Econometrics* 8.3 (Sept. 2004). DOI: 10.2202/1558-3708.1226.

[59] J. Lintner. »The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets«. In: *The Review of Economics and Statistics* 47.1 (1965), pp. 13–37. ISSN: 00346535, 15309142.

[60] H. L. Lütkepohl and M. Krätzig. *Applied Time Series Econometrics*. Themes in Modern Econometrics. Cambridge University Press, 2004. ISBN: 9780521547871.

[61] B. Mandelbrot. »Une classe processus stochastiques homothétiques à soi; application à la loi climatologique H. E. Hurst.« French. In: *C. R. Acad. Sci. Paris* 260 (1965), pp. 3274–3277. ISSN: 0001-4036.

[62] B. Mandelbrot. »The Variation of Some Other Speculative Prices«. In: *The Journal of Business* 40.4 (1967), pp. 393–413. ISSN: 00219398, 15375374.

[63] V. A. Marchenko and L. A. Pastur. »Distribution of eigenvalues for some sets of random matrices«. In: *Mathematics of the USSR-Sbornik* 1.4 (1967), pp. 457–483. DOI: 10.1070/SM1967v001n04ABEH001994.

[64] H. Markowitz. »Portfolio Selection«. In: *The Journal of Finance* 7.1 (1952), pp. 77–91. ISSN: 00221082.

[65] B. Mehlig and J. Chalker. »Eigenvector correlations in non-Hermitian random matrix ensembles«. In: *Annalen der Physik* 7.5-6 (1998), pp. 427–436. ISSN: 1521-3889. DOI: 10.1002/(SICI)1521-3889(199811)7:5/6<427::AID-ANDP427>3.0.CO;2-1.

[66] B. Mehlig and J. Chalker. »Statistical properties of eigenvectors in non-Hermitian Gaussian random matrix ensembles«. In: *Journal of Mathematical Physics* 41.5 (2000), pp. 3233–3256. DOI: http://dx.doi.org/10.1063/1.533302.

[67] R. C. Merton. »On estimating the expected return on the market: An exploratory investigation«. In: *Journal of Financial Economics* 8.4 (1980), pp. 323–361. ISSN: 0304-405X. DOI: 10.1016/0304-405X(80)90007-0.

[68] Microsoft. *Visual Studio Community 2015*. Version 14.0.25213.00 Update 2. June 2014.

[69] T. Mikosch and C. Stărică. »Extremes and Integrated Risk Management«. In: ed. by P. Embrechts. Risk Books, 2000. Chap. Is it Really Long Memory we See in Financial Returns?, pp. 149–168. ISBN: 9781899332748.

[70] T. Mikosch and C. Stărică. »Nonstationarities in Financial Time Series, the Long-Range Dependence, and the IGARCH Effects«. In: *Review of Economics and Statistics* 86.1 (Feb. 2004), pp. 378–390. ISSN: 0034-6535. DOI: 10.1162/003465304323023886.

[71] J. Mossin. »Equilibrium in a Capital Asset Market«. In: *Econometrica* 34.4 (1966), pp. 768–783. ISSN: 00129682, 14680262.

[72] R. J. Muirhead. *Aspects of Multivariate Statistical Theory*. Wiley, 1982.

[73] M. C. Münnix, R. Schäfer, and T. Guhr. »Impact of the tick-size on financial returns and correlations«. In: *Physica A: Statistical Mechanics and its Applications* 389.21 (2010), pp. 4828–4843. ISSN: 0378-4371. DOI: 10.1016/j.physa.2010.06.037.

[74] NASDAQ. *Company List (NASDAQ, NYSE, & AMEX)*. Last visited 19/03/2015. 2015.

[75] NASDAQ OMX Group. *Annual Changes to the NASDAQ-100 Index*. Last visited 18/03/2015. Dec. 2014.

[76] NASDAQ OMX Group. *Market Calendar & Trading Hours*. Last visited 18/03/2015. 2015.

[77] N. Nelder and R. Mead. »A Simplex Method for Function Minimization«. In: *The Computer Journal* 7.4 (1965), pp. 308–313. DOI: 10.1093/comjnl/7.4.308.

[78] A. Nica and R. Speicher. *Lectures on the Combinatorics of Free Probability*. Lectures on the combinatorics of free probability Bd. 13. Cambridge University Press, 2006. ISBN: 9780521858526.

[79] S. O'Rourke, V. Vu, and K. Wang. »Eigenvectors of random matrices: A survey«. In: *ArXiv e-prints* (Jan. 2016).

[80] R. C. A. Oomen. *Modelling realized variance when returns are serially correlated*. Discussion Papers, Research Unit: Market Processes and Governance SP II 2004-11. WZB Berlin Social Science Center, 2004.

[81] O. Pfaffel and E. Schlemm. »Eigenvalue distribution of large sample covariance matrices of linear processes«. In: *Probability and Mathematical Statistics* 31.2 (2011), pp. 313–329.

[82] Pi Trading Inc. *Historical Intraday Stock Data*. http://pitrading.com/historical-stock-data.html. Last visited 27/03/2015. Mar. 2015.

[83] R. Piessens and M. Branders. »A Note on the Optimal Addition of Abscissas to Quadrature Formulas of Gauss and Lobatto Type«. In: *Mathematics of Computation* 28.125 (1974), pp. 135–139. ISSN: 00255718, 10886842.

[84] W. H. Press et al. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. 3rd ed. New York, NY, USA: Cambridge University Press, 2007. ISBN: 0521880688.

[85] M. Rudelson and R. Vershynin. »Delocalization of eigenvectors of random matrices with independent entries«. In: *Duke Math. J.* 164.13 (Oct. 2015), pp. 2507–2538. DOI: 10.1215/00127094-3129809.

[86] S&P Dow Jones Indices. *Celgene, Kinder Morgan and Actavis Set to Join the S&P 100; Several Constituent Changes Announced for S&P 500, S&P MidCap 400 and S&P SmallCap 600*. Press Release. published 2015-03-13. Mar. 2015.

[87] G. Samorodnitsky. *Long Range Dependence*. Vol. 1. Foundations and Trends® in Stochastic Systems 3. Now Publishers, 2007. ISBN: 9781601980908.

[88] G. A. Seber. *Multivariate Observations*. Wiley Series in Probability and Statistics. Wiley, 2009. ISBN: 9780470317310.

[89] M. Sewell. »Characterization of financial time series«. In: *Research Note* 11.01 (2011), p. 01.

[90] W. F. Sharpe. »A Simplified Model for Portfolio Analysis«. In: *Management Science* 9.2 (1963), pp. 277–293. DOI: 10.1287/mnsc.9.2.277.

[91] W. F. Sharpe. »Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk«. In: *The Journal of Finance* 19.3 (1964), pp. 425–442. ISSN: 00221082, 15406261.

[92] A. Shiryaev and J. Jacod. *Limit Theorems for Stochastic Processes*. Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen. Springer-Verlag, 1987. ISBN: 9783540178828.

[93] J. W. Silverstein. »Strong Convergence of the Empirical Distribution of Eigenvalues of Large Dimensional Random Matrices«. In: *Journal of Multivariate Analysis* 55.2 (Nov. 1995), pp. 331–339. ISSN: 0047-259X. DOI: 10.1006/jmva.1995.1083.

[94] J. W. Silverstein and Z. Bai. »On the Empirical Distribution of Eigenvalues of a Class of Large Dimensional Random Matrices«. In: *Journal of Multivariate Analysis* 54.2 (1995), pp. 175–192. ISSN: 0047-259X. DOI: 10.1006/jmva.1995.1051.

[95] F. Sowell. »Maximum likelihood estimation of stationary univariate fractionally integrated time series models«. In: *Journal of Econometrics* 53.1 (1992), pp. 165–188. ISSN: 0304-4076. DOI: 10.1016/0304-4076(92)90084-5.

[96] R. Speicher. *Free Probability Theory – Lectures at the »Summer School in Operator Algebras«, Ottawa*. June 2005.

[97] R. Speicher, C. Vargas, and T. Mai. »Free Deterministic Equivalents, Rectangular Random Matrix Models, and Operator-Valued Free Probability Theory«. In: *CoRR* abs/1110.1237 (2011).

[98] T. J. Stieltjes. »Recherches sur les fractions continues«. French. In: *Annales de la faculté des sciences de Toulouse* 8.4 (1894), pp. 1–122.

[99] P. Suárez-García and D. Gómez-Ullate. »Multifractality and long memory of a financial index«. In: *Physica A: Statistical Mechanics and its Applications* 394 (2014), pp. 226–234. ISSN: 0378-4371. DOI: 10.1016/j.physa.2013.09.038.

[100] T. Tao and V. Vu. »Random Matrices: Universal Properties of Eigenvectors«. In: *Random Matrices: Theory and Applications* 01.01 (2012), pp. 1150001-1–1150001-27. DOI: 10.1142/S2010326311500018.

[101] M. Taqqu. »Theory and Applications of Long-Range Dependence«. In: ed. by P. Doukhan, G. Oppenheim, and M. Taqqu. Birkhäuser Boston, 2003. Chap. Fractional Brownian Motion and Long Range Dependence, pp. 5–38. ISBN: 9780817641689.

[102] E. C. Titchmarsh. *Introduction to the Theory of Fourier Integrals*. Clarendon Press, 1937.

[103] J. L. Treynor. »Market Value, Time, and Risk«. Unpublished manuscript. 1961.

[104] J. L. Treynor. »Toward a Theory of Market Value of Risky Assets«. Unpublished manuscript. A final version was published 1999, in [105, pp. 15–22]. 1962.

[105] J. L. Treynor. »Toward a Theory of Market Value of Risky Assets«. In: *Asset Pricing and Portfolio Performance: Models, Strategy, and Performance Metrics*. Ed. by R. A. Korajczyk. Risk Books, 1999, pp. 15–22. ISBN: 9781899332366.

[106] A. M. Tulino and S. Verdú. »Random Matrix Theory and Wireless Communications«. In: *Foundations and Trends® in Communications and Information Theory* 1.1 (2004), pp. 1–182. ISSN: 1567-2190. DOI: 10.1561/0100000001.

[107] D. Voiculescu. »Addition of certain non-commuting random variables«. In: *Journal of Functional Analysis* 66.3 (1986), pp. 323–346. ISSN: 0022-1236. DOI: 10.1016/0022-1236(86)90062-5.

[108] D. Voiculescu. »Limit laws for Random matrices and free products«. In: *Inventiones mathematicae* 104.1 (1991), pp. 201–220. ISSN: 0020-9910. DOI: 10.1007/BF01245072.

[109] D. Voiculescu. *Free Probability Theory*. Vol. 12. American Mathematical Society, 1997. ISBN: 9780821871201.

[110] D. Voiculescu. »A strengthened asymptotic freeness result for random matrices with applications to free entropy«. In: *International Mathematics Research Notices* 1998.1 (1998), pp. 41–63. DOI: 10.1155/S107379289800004X.

[111] D. Voiculescu, K. Dykema, and A. Nica. *Free Random Variables*. CRM Monograph Series. American Mathematical Society, 1992. ISBN: 9780821811405.

[112] D. V. Widder. »The Stieltjes transform«. In: *Trans. Amer. Math. Soc.* 43.1 (1938), pp. 7–60. DOI: 10.1090/S0002-9947-1938-1501933-2.

[113] D. V. Widder. *The Laplace Transform*. Dover books on mathematics. Dover Publications, 2010. ISBN: 9780486477558.

[114] E. P. Wigner. »On the Distribution of the Roots of Certain Symmetric Matrices«. In: *Annals of Mathematics*. Second Series 67.2 (Mar. 1958), pp. 325–327. ISSN: 0003486X.

[115] W. Willinger, M. Taqqu, and V. Teverovsky. »Stock market prices and long-range dependence«. In: *Finance and Stochastics* 3.1 (1999), pp. 1–13. ISSN: 0949-2984. DOI: 10.1007/s007800050049.

[116] A. Wintner. *Spektraltheorie Der Unendlichen Matrizen: Einführung in Den Analytischen Apparat Der Quantenmechanik*. S. Hirzel, 1929.

[117] J. Wishart. »The Generalised Product Moment Distribution in Samples from a Normal Multivariate Population«. In: *Biometrika* 20A.1/2 (1928), pp. 32–52. ISSN: 00063444.

[118] Wolfram Research Inc. *Mathematica 11*. Version 11.2.0.0 x64. Dec. 2014.

[119] N. Xia, Y. Qin, and Z. Bai. »Convergence rates of eigenvector empirical spectral distribution of large dimensional sample covariance matrix«. In: *Ann. Statist.* 41.5 (Oct. 2013), pp. 2572–2607. DOI: 10.1214/13-AOS1154.

Index

Symbols

⋆
-Probability Space *see* Probability Space

A

ARFIMA 54
ARMA
Multivariate 30
ARMA 30
Average Eigenvalue Distribution . . . *see* Eigenvalue Distribution

B

Backshift Operator . . *see* Operators

C

Centering 16
Convegrence
Distribution
Joint 123
Convergence
Absolute 11

D

Data 69
DJIA Intraday 73
NASDAQ Intraday 72
NASDAQ-100 Intraday 71
S&P500 Intraday
Big 84
Long 84
S&P 500 Intraday 73
Dependence
Long-Range 52
Short-Range 52
Distribution
Empirical Spectral . . *see* Empirical Spectral
Wishart 12
Distribution Generating Polynomial 40

E

Eigenvalue Distribution
Average 9
Expected 11
Empirical Spectral
Cumulative Distribution Function 12
Distribution 11
Ensemble
Wishart 12
Euler–Maclaurin approximation . 65

F

Fractional Difference Operator . . . . *see* Operators
Freeness 17
Free Independence . *see* Freeness
of Subsets 18

G

GARCH 53
Gaussian quadrature rule . . . . . *see* Quadrature rule

I

Im $\sqrt{x + \mathrm{i}y}$ . . . . . *see* Lemma
Independence
Classical 16
Free 17
Tensor 16
Inverse Cauchy-transform 123

K

Kronrod–Points 64

L

Lemma
  Im $\sqrt{x+\mathrm{i}y}$ 125
  Re $\sqrt{x+\mathrm{i}y}$ 126
Lindeberg-type condition 49
Liquidity 75
  Average 77
  Single Stock 77

N

Newton–Cotes formula 64
Non-Commutative
  Probability Space . *see* Probability Space
  Random Variable . . . . *see* Random Variable
Numerical Integration
  Global Adaptive 64
  Singularity Handling 65
Numerical Optimization
  Grid Search 67
  Monte Carlo 67
  Nelder–Mead 67
  Quasi Monte Carlo 67

O

Operators
  Backshift 29
  Fractional Backshift 53
Order Book 71

P

Positivity 8
Probability Space
  $\star$- 8
  Classical 8
  Non-Commutative 8

Q

Quadrature rule 64
  Equidistant Points 64
  Gauss–Kronrod 64
  Gaussian 64

R

Random Matrices 9
Random Variable
  Classical 8
  Non-Commutative 8
    Rectangular 24
Re $\sqrt{x+\mathrm{i}y}$ . . . . . *see* Lemma
Return 74
Runge's phenomenon 64

S

Schur's Complement 126
Singularity Handling . . . *see* Numerical Integration
Spectral Density 38
Stieltjes
  Polynomials 64

T

Tensor
  Independence *see* Independence
Trading Hours 71

V

VAR 31
VARMA 31
VMA 31

W

White Noise 29
Wishart Distribution . . . . . *see* Distribution
Wishart Ensemble . . . . *see* Ensemble

Eidesstattliche Versicherung

Hiermit erkläre ich, Marco Breitig, an Eides statt, dass ich die Dissertation mit dem Titel »Non-Commutative Probability Theory and Applications in Finance« selbständig — und bei einer Zusammenarbeit mit anderen Wissenschaftlern gemäß der beigefügten Darlegung — nach §6 Abs. 3 der Promotionsordnung der Fakultät für Wirtschafts- und Sozialwissenschaften vom 24. August 2010 verfasst und keine anderen als die von mir angegebenen Hilfsmittel benutzt habe. Die den herangezogenen Werken wörtlich oder sinngemäß entnommenen Stellen sind als solche gekennzeichnet.

Ich versichere, dass ich keine kommerzielle Promotionsberatung in Anspruch genommen habe und die Arbeit nicht schon in einem früheren Promotionsverfahren im In- oder Ausland angenommen oder als ungenügend beurteilt worden ist.

.....	.....
Ort, Datum	Marco Breitig