Continuous Time Heterogeneous Agent Models

Universität Hamburg
Fakultät für Wirtschafts- und Sozialwissenschaften

Dissertation

Zur Erlangung der Würde der Doktorin/des Doktors der Wirtschafts- und Sozialwissenschaften

(gemäß der PromO vom 18. Januar 2017)

vorgelegt von
René Marco Glawion
aus Bremen

Hamburg 2025

Vorsitzende: Prof. Dr. Ole Wilms
Erstgutachter: Prof. Dr. Olaf Posch
Zweitgutachter: Prof. Dr. Alexander Szimayer
Datum der Disputation: 19.02.2026

Acknowledgements / Dankessagung

Diese Dissertation wäre ohne die Unterstützung und Hilfe zahlreicher Personen nicht möglich gewesen, denen ich an dieser Stelle herzlich danken möchte.

Mein tief empfundener Dank gilt zunächst meinem Doktorvater, Olaf Posch. Seit meinem Bachelorstudium hat er mich kontinuierlich begleitet, inspiriert und in meiner wissenschaftlichen Entwicklung maßgeblich unterstützt. Er begeisterte mich bereits im Bachelorstudium für das Lösen makroökonomischer Modelle in stetiger Zeit und vermittelte mir dafür umfassende mathematische und numerische Methoden. Von meiner Zeit als studentische Hilfskraft bis hin zu meiner Tätigkeit als wissenschaftlicher Mitarbeiter konnte ich von zahlreichen produktiven Diskussionen, wertvollem Feedback und seiner engagierten Führung profitieren. Auch die persönlichen Momente, etwa unsere jährlichen Bowlingwettbewerbe, bleiben unvergessen.

Ein besonderer Dank geht ebenso an meinen Zweitbetreuer, Alexander Szimayer, der mich ebenfalls seit dem Bachelorstudium begleitet hat und mir stets die Möglichkeit bot, aktiv in seiner Forschungsgruppe mitzuwirken. Seine Unterstützung hat mir nicht nur die Möglichkeit eröffnet, meinen Master an der Universität Hamburg zu absolvieren, sondern er hat mich während meines gesamten Studiums fortlaufend gefördert und begleitet. Darüber hinaus danke ich den Mitarbeitern seines Lehrstuhls, insbesondere Johannes Heul, Mick Schäfer, Aidin Miri Lavasani, Sophie Döpp, und Christian Hilpert sowie allen ehemaligen Mitarbeitern, die durch fachliche Gespräche und gemeinsame Eispausen bei Luicella's meine Zeit in Hamburg zu etwas Besonderem machten.

Herzlich danken möchte ich zudem meinen engen Freunden, Koautoren und Mitgründern Frédéric Enrico Haller und Marc Puche Niubó. Unsere gemeinsamen Projekte, von Algotrading-Ideen über Forschungsarbeiten, von denen eines Eingang in diese Dissertation fand, bis hin zur Teilnahme am Hamburg Triathlon, zeigten stets, dass ich mich uneingeschränkt auf euch verlassen kann. Besonderer Dank gilt euch zusätzlich für die Bereitstellung der LaTeX-Vorlage dieser Dissertation.

Ein herzliches Dankeschön gilt Josie Oetjen, die mich über die gesamte Promotionszeit begleitet und unterstützt hat. Ihre Motivation und ihr Ansporn waren entscheidend, diese Dissertation zum Abschluss zu bringen. Ihre detaillierten Korrekturvorschläge haben wesentlich zur Qualität dieser Arbeit beigetragen, und ich hoffe, in Zukunft gemeinsam mit ihr weiter forschen zu dürfen.

Adrien Auclert danke ich für die spontane Einladung zu ihm, um gemeinsame Forschungsprojekte zu besprechen. Durch ihn habe ich die Methodik der Sequence-Space Jacobians kennengelernt, und dank seiner intensiven Hilfe und detailliertem Feedback konnte Kapitel 2 dieser Dissertation entstehen.

Schließlich möchte ich meinen Eltern, Anja und Marco Glawion, danken. Ihre kontinuierliche Unterstützung und ihr uneingeschränktes Vertrauen haben nicht nur mein Studium, sondern insbesondere diese Dissertation ermöglicht.

Ihnen allen gilt mein aufrichtiger Dank.

René Marco Glawion

Contents

List of Figures **i**

List of Tables **iii**

Synopsis **v**

S.1 Motivation and overview v
S.2 A General Equilibrium Model of Earnings, Income, and Wealth v
S.3 Sequence-Space Jacobians in continuous time vi
S.4 Predicting the COVID-19 pandemic in Germany vii
Bibliography Synopsis viii

1 A General Equilibrium Model of Earnings, Income, and Wealth **1**

1.1 Introduction 1
1.2 The model 4
1.2.1 Framework 4
1.2.2 Stationary equilibrium 7
1.3 Numerical analysis 9
1.3.1 Algorithm 9
1.3.2 Calibration and results 11
1.4 The effects of a minimum wage on wealth inequality 13
1.5 Channels of inequality 17
1.6 Conclusion 18
A.1 Derivation of the HJB equations 19
A.2 Derivation of the FP equation 23
A.3 Income distribution 25
A.4 Stationary distribution as an eigenvector 26
A.5 Numeric scheme for the equilibrium model 29
Bibliography Chapter 1 35

2 Sequence-Space Jacobians in Continuous Time **41**
2.1 Introduction . . . 41
2.2 Model environment . . . 43
2.2.1 A simple HANK model in continuous time . . . 44
2.2.2 Definition of equilibrium . . . 44
2.2.3 Certainty equivalence . . . 46
2.2.4 Incorporating Poisson shocks . . . 46
2.3 Computing Jacobians in continuous time . . . 47
2.3.1 General model representation . . . 47
2.3.2 Structure of the Jacobian $\mathcal{J}$ . . . 48
2.3.3 Efficiently computing expectation vectors . . . 50
2.4 The model as a DAG . . . 51
2.4.1 Fiscal policy DAG for the one-asset HANK model . . . 51
2.4.2 Taking derivatives along the DAG . . . 52
2.5 Computing the transition path after a shock to government expenditures . . . 53
2.5.1 Calibration procedure . . . 53
2.5.2 Resulting transition paths . . . 54
2.6 Conclusion . . . 55
B.1 Krusell-Smith model in continuous time . . . 56
B.1.1 Model environment . . . 56
B.1.2 Equilibrium . . . 57
B.1.3 Linearized dynamics using Jacobians . . . 57
B.1.4 Estimation . . . 58
B.1.5 Numerical results . . . 59
B.1.6 Existence and uniqueness of infinitesimal generator . . . 61
B.2 Discrete-time one-asset HANK model . . . 61
B.3 Comparison with the method used in Achdou et al. (2022) . . . 62
Bibliography Chapter 2 . . . 63

3 Projecting the Spread of COVID-19 for Germany **65**
3.1 Introduction . . . 66
3.2 Data . . . 69
3.3 The model . . . 71
3.3.1 The structure . . . 71
3.3.2 Transition rates . . . 72
3.3.3 Probabilities and means . . . 75
3.3.4 The number of individuals who get and ever got sick and deaths . . . 76

3.3.5 Model properties . . . 77
3.4 Calibration . . . 78
3.4.1 Data . . . 78
3.4.2 Targets and parameters . . . 78
3.4.3 The goodness of fit of the calibration . . . 80
3.5 Findings . . . 81
3.5.1 The number of sick individuals and health care . . . 81
3.5.2 Individual risks . . . 82
3.5.3 Robustness check . . . 83
3.6 The effect of policy measures . . . 85
3.6.1 The experiments . . . 85
3.6.2 The quantitative effects of policy measures . . . 86
3.7 Conclusion . . . 89
C.1 Data sources . . . 91
C.2 Intermediate steps for (3.15) . . . 91
C.3 Goodness of fit . . . 92
Bibliography Chapter 3 . . . 92

Appendices **97**

D General Appendix **ix**
D.1 Abstract . . . ix
D.2 Zusammenfassung . . . xii

List of Figures

1.1 Gini coefficient of earnings for different levels of unemployment benefits 14
1.2 Gini coefficient of income for different levels of unemployment benefits 15
1.3 Gini coefficient of wealth for different levels of unemployment benefits 16

2.1 DAG for the one-asset HANK model 52
2.2 Comparison of the discrete-time and continuous-time transition path 55
2.3 Transition paths of balanced budget and deficit-financed government spending shocks 56
2.4 Impulse responses in the Krusell-Smith model 60
2.5 Log-Likelihood estimation in the Krusell-Smith model 61
2.6 Comparison of SSJ to Achdou et al. (2022) 63

3.1 Total cases in Hubei seem to converge to a constant 70
3.2 Total cases in South Korea exhibit a similar pattern to the Hubei province 71
3.3 Transitions between the state of health (initial state), sickness, death and health despite infection or after recovery 73
3.4 New incidences (reported) in the model ($N_2^{\text{new}}(t)$, curve) and in the data (dots) (left figure) and the number $N_2^{\text{ever}}(t)$ of sick in model and data (right figure) for the epidemic without public health measures 81
3.5 The number of sick individuals ($s = 2$) by day starting on 24 February 2020 in the absence of public health measures 82
3.6 The probability per day (dotted) and per week (solid) to get sick in the absence of public health measures. 83
3.7 The number $N_2(t)$ of sick individuals as a function of assumptions on the probability r to get sick after an infection 84
3.8 The effect of a shut down 87
3.9 The effect of changes in the long-run infection rate $\bar{\rho}$ 89
3.10 Enlarged version of figure 3.4 with data and model fit 92

List of Tables

1.1 Calibration of the θ-w-model. 12
1.2 Results of the calibrated model 13
1.3 Channels of inequality 18

2.1 Calibration of the one-asset HANK model 53

3.1 Current infection rates and risks in a sample of European countries 72
3.2 Parameters for the epidemic without public health measures 79
3.3 Varying the long-run sickness rate r 84

Synopsis

S.1 Motivation and overview

Heterogeneous-agent models have revolutionized macroeconomic analysis by incorporating individual-level shocks and decisions into aggregate outcomes. Traditional models often relied on representative agents, overlooking the rich diversity of behaviors and experiences across different economic actors. The continuous-time approach to heterogeneous-agent models further enhances this framework by exploiting advanced mathematical techniques and offering computational advantages over discrete-time models.

The motivation for this dissertation stems from the need to better understand the distributional aspects of the economy and to develop efficient computational methods when analyzing dynamic responses to shocks. By focusing on continuous-time models, we can capture the nuances of agent heterogeneity more accurately and handle complex state spaces more effectively.

The three chapters of this dissertation are interconnected in their goal of advancing continuous-time heterogeneous-agent modeling. The first chapter lays the theoretical groundwork by extending the Bewley-Huggett-Aiyagari framework to jointly model earnings, income, and the distribution of wealth. The second chapter builds upon this approach by addressing the computational challenges of solving transition dynamics, introducing a novel method that accelerates computations significantly. The third chapter applies continuous-time methods to an epidemiological context, modeling the COVID-19 pandemic using techniques from search and matching theory.

S.2 A General Equilibrium Model of Earnings, Income, and Wealth

In the first chapter, we extend the standard Bewley-Huggett-Aiyagari framework to simultaneously describe the steady-state distributions of earnings, income, and the distribution of wealth in a general equilibrium setting. By casting the model in continuous time, we leverage advanced mathematical techniques and newly developed numerical tools to handle multiple continuous state variables. Specifically, we model labor income and interest rates as continuous stochastic processes, allowing for a

more accurate replication of the empirical distributional patterns observed in U.S. microdata.

Our work builds upon the continuous-time model developed by Achdou et al. (2022), who formulated a version of the Aiyagari (1994) model and detailed numerical solutions using finite-difference methods. We extend their numerical algorithm to accommodate multiple continuous state variables, specifically stochastic labor income and stochastic interest rates. This combination is crucial for explaining the distinct distributions of earnings, income, and wealth without resorting to implausible assumptions such as "superstar" income states, as criticized by Benhabib and Bisin (2018).

Previous studies have often modeled wealth distribution effectively but failed to jointly model the distribution of both earnings and income. For instance, Krusell and Smith (1998) achieved a good fit for wealth distribution using stochastic discount factors but focused solely on wealth. Castaneda et al. (2003) attempted to match both income and wealth distributions but had to introduce an unrealistically high income state to generate a sufficiently thick right tail in terms of wealth distribution. By integrating stochastic labor income and heterogeneous interest rates, our model captures the essential features of earnings, income, and wealth distributions. We calibrate the model to U.S. data and demonstrate that it can replicate the empirical observations without extreme assumptions. Furthermore, we conduct a policy experiment to examine the effects of increasing unemployment benefits. Our findings indicate that higher unemployment benefits lead to increased wealth inequality in the steady state. This counterintuitive result arises because higher benefits reduce the need for precautionary savings among lower-income agents, thereby increasing the concentration of wealth among those who continue to save and invest.

Our contribution lies not only in developing new methodologies but also in providing an empirical framework that jointly models the three dimensions of inequality highlighted by Quadrini and Rıos-Rull (1997): earnings, income, and wealth. By addressing the gap in the literature, we offer a more comprehensive understanding of how different sources of heterogeneity interact within the economy.

S.3 Sequence-Space Jacobians in continuous time

The second chapter addresses one of the significant challenges in heterogeneous-agent modeling: the computation of transition dynamics and impulse responses following aggregate shocks. Traditional methods for solving these dynamics are computationally intensive, especially in models with large state spaces and rich agent heterogeneity.

We introduce a novel approach by extending the sequence-space Jacobian method proposed by Auclert et al. (2021) for discrete-time models to the continuous-time setting. This extension utilizes the infinitesimal generator of the Hamilton-Jacobi-Bellman equation to compute expectation vectors, allowing for the efficient computation of transition dynamics in continuous-time heterogeneous-agent models.

The sequence-space Jacobian method offers substantial computational advantages by solving the

model in the sequence space rather than the state space. This approach decouples the computation from the size of the state space, which is particularly beneficial for models with high-dimensional state variables. By working in continuous time, we further exploit the properties of stochastic calculus to avoid discretization errors associated with numerical expectations in discrete-time models.

We apply our method to two models: a neoclassical heterogeneous-agent model in the spirit of Krusell and Smith (1998) and a one-asset New Keynesian model similar to McKay et al. (2016). In both cases, the idiosyncratic shocks to labor productivity follow either a Poisson process or a diffusion process. Our results demonstrate a substantial computational acceleration at least 30 times faster than conventional methods, such as those used by Achdou et al. (2022) and Kaplan et al. (2018). This acceleration makes it feasible to analyze models that were previously too computationally demanding.

Our work contributes to the literature by showing how continuous-time approaches can benefit from recent developments in computational methods for heterogeneous-agent models. Compared to alternative methods that involve model reduction or automatic differentiation (e.g., Ahn et al., 2018), our approach avoids calculating large Jacobians and allows for higher-order perturbations without a significant increase in computational complexity.

By providing an efficient way to compute transition dynamics in continuous-time models, we facilitate the study of the rich interactions between agent heterogeneity and aggregate outcomes more effectively. This advancement is crucial for analyzing policy interventions and understanding the dynamic responses of the economy to shocks.

S.4 Predicting the COVID-19 pandemic in Germany

The third chapter showcases the versatility of continuous-time heterogeneous-agent models by applying the methods to an epidemiological context. Motivated by the urgent need for reliable predictions during the COVID-19 pandemic, we model the spread of the virus in Germany using an epidemiological Markov model formulated in continuous time.

Our model is rooted in the SEIR (Susceptible-Exposed-Infectious-Recovered) framework but incorporates continuous-time methods from search and matching theory, as seen in Diamond (1982), Mortensen (1982), and Pissarides (1985). By utilizing continuous-time Markov chains, we can capture the stochastic nature of transitions between health states and predict the probabilities of individuals being in various states over time.

We begin by analyzing quantitative data from Hubei province in China and South Korea to understand the dynamics of the pandemic in regions where it had progressed significantly. This analysis informs our model calibration for Germany. We match the time series data of infected individuals provided by the Robert Koch Institute (Robert Koch Institut (RKI), 2020) and use the model to predict future developments.

Our model allows us to estimate key quantities of interest, such as the risk of infection for individuals,

the expected number of individuals who will become infected, and the number of active infections at any point in time. We conduct counterfactual analyses to assess the effectiveness of various policy interventions, such as social distancing measures and lockdowns.

One of the critical contributions of this chapter is the emphasis on the importance of modeling assumptions in evaluating policy measures. We discuss how different assumptions about parameters like transmission rates and recovery rates can significantly impact the predictions and policy implications. By making these assumptions explicit, we provide a transparent framework that can be updated as new data become available.

Our approach offers advantages over traditional epidemiological models by leveraging the computational efficiencies of continuous-time methods. This methodological choice allows for more accurate and timely predictions, which are essential for informing public health policies during a rapidly evolving pandemic.

Bibliography Synopsis

Achdou, Y., Han, J., Lasry, J.-M., Lions, P.-L., and Moll, B. (2022). Income and wealth distribution in macroeconomics: A continuous-time approach. *The Review of Economic Studies*, 89(1):45–86.

Ahn, S., Kaplan, G., Moll, B., Winberry, T., and Wolf, C. (2018). When inequality matters for macro and macro matters for inequality. *NBER Macroeconomics Annual*, 32(1):1–75.

Aiyagari, S. R. (1994). Uninsured idiosyncratic risk and aggregate saving. *The Quarterly Journal of Economics*, 109(3):659–684.

Auclert, A., Bardóczy, B., Rognlie, M., and Straub, L. (2021). Using the sequence-space jacobian to solve and estimate heterogeneous-agent models. *Econometrica*, 89(5):2375–2408.

Benhabib, J. and Bisin, A. (2018). Skewed wealth distributions: Theory and empirics. *Journal of Economic Literature*, 56(4):1261–91.

Castaneda, A., Diaz-Gimenez, J., and Rios-Rull, J.-V. (2003). Accounting for the us earnings and wealth inequality. *Journal of political economy*, 111(4):818–857.

Diamond, P. A. (1982). Aggregate demand management in search equilibrium. *Journal of Political Economy*, 90:881–94.

Kaplan, G., Moll, B., and Violante, G. L. (2018). Monetary policy according to hank. *American Economic Review*, 108(3):697–743.

Krusell, P. and Smith, Jr, A. A. (1998). Income and wealth heterogeneity in the macroeconomy. *Journal of Political Economy*, 106(5):867–896.

McKay, A., Nakamura, E., and Steinsson, J. (2016). The power of forward guidance revisited. *American Economic Review*, 106(10):3133–58.

Mortensen, D. T. (1982). Property rights and efficiency in mating, racing, and related games. *American Economic Review*, 72:968–79.

Pissarides, C. A. (1985). Short-run equilibrium dynamics of unemployment vacancies, and real wages. *American Economic Review*, 75:676–90.

Quadrini, V. and Rıos-Rull, J.-V. (1997). Dimensions of inequality: Facts on the us distribution of earnings, income and wealth. *Federal Reserve Bank of Minneapolis Quarterly Review*, 21(2):3–21.

Robert Koch Institut (RKI) (2020). Covid-19: Fallzahlen in deutschland und weltweit. *https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Fallzahlen.html*.

Chapter 1

A General Equilibrium Model of Earnings, Income, and Wealth

Frédéric Enrico Haller and Marc Puche Niubó contributed to the work presented in this chapter.

In what follows, we extend the standard Bewley-Hugget-Aiyagari framework to describe the steady-state distributions of earnings, income, and wealth simultaneously. We cast our model in continuous time to exploit more powerful mathematical techniques as well as newly developed numerical tools. We provide detailed proofs to explain our numerical methods. Afterwards, we calibrate the model to U.S. microdata and demonstrate its ability to explain the different distributional outcomes of earnings, income, and wealth by assuming a continuous state space of income and heterogeneous interest rates. Finally, we conduct a policy experiment and demonstrate that, ceteris paribus, an increase in unemployment benefits leads to a rise in steady-state wealth inequality.

1.1 Introduction

This work combines several results from the theory of heterogeneous agent models that have been established in recent years. First, the core of our model, as well as of our numerical results, builds on the work of Achdou et al. (2022). These authors develop a continuous-time version of the Aiyagari (1994) model and describe in great detail how to solve it numerically using finite-difference methods, namely the upwind scheme. We extend their numerical algorithm to handle multiple continuous state variables. More precisely, we model labor income and interest rates continuously.
The idea to combine these two variables is not new. In a seminal work, Benhabib et al. (2011) show that modeling a risk-free interest rate together with idiosyncratic stochastic interest rates leads to a Pareto tail of the wealth distribution. Khieu and Wälde (2018) adapt this idea and develop a model with both stochastic labor and stochastic interest rates. They achieve a good fit of the

wealth distribution. However, the authors assume that interest rates and labor income are two-state processes that switch via Poisson jumps. Further, they only work in partial equilibrium, whereas we work in general equilibrium throughout the whole analysis.

In a comprehensive study, Castaneda et al. (2003) model the earnings process and match both the income and the wealth distribution almost perfectly. Yet, to generate sufficiently thick right tails in the wealth distribution, they have to introduce a labor income state where the income level is an order of magnitude larger than what would be plausible empirically. This situation is what Benhabib and Bisin (2018) refer to as an "awesome state" or a "superstar state."[1]

It therefore seems reasonable to search for other determinants of the wealth distribution. Quadrini and Rıos-Rull (1997) analyze empirical facts of inequality based on the 1992 Survey of Consumer Finances and 1984-85 and 1989-90 Panel Study of Income Dynamics. They find three dimensions of inequality. First, inequality in *earnings*, which are related to labor input. Second, *income* inequality, where income is defined as revenue from all sources before taxes but after transfers. In our model, income is composed of labor earnings plus income generated by wealth, i.e., interest rate payments. Third, inequality in *wealth*, which is defined as the net worth of households. The authors find that wealth is by far the most concentrated of the three variables, earnings rank second, and income is the most dispersed of the three. The authors contend that "a good theory of inequality should be able to account for the differences among the distributions of earnings, income and wealth [...]." Yet, economists did not come up with a joint theory of these three inequality measures. Our study aims to close this gap.

On the methodological front, since Achdou et al. (2022), efforts have been made to further enhance computational efficiency and the capacity to tackle increasingly complex models. For instance, Garcke and Ruttscheidt (2025) and Du (2025) investigate sparse grid techniques. While these studies offer promising avenues for solving a broader class of continuous-time models, our contribution is distinct in its specific application and extension of the established finite-difference upwind scheme to a multi-dimensional continuous state space tailored to jointly analyze the distribution of earnings, income, and wealth. The robustness and interpretability of the finite-difference method, especially for problems with clear boundary conditions like ours, provide a solid foundation for our quantitative analysis, whereas the newer sparse grid, while powerful, is still being broadly adopted and tested for the specific type of multi-distributional analysis we undertake.

In terms of economic mechanisms, the recent continuous-time literature has also pushed the boundaries. Benhabib et al. (2024), for example, develop a tractable continuous-time model that incorporates capital income jumps via the affine jump-diffusion framework of Duffie et al. (2000). Their focus on jumps in capital income to match wealth and earnings distributions offers an alternative to our approach of using continuous stochastic processes for both labor income and interest rates. While jumps

[1] For models with superstars see, for example, Rosen (1981), Garicano and Rossi-Hansberg (2006), Gabaix and Landier (2008), Tervio (2008), and Geerolf (2016).

can be crucial for capturing extreme, rare events, our framework with continuous processes allows for a smoother, more persistent form of heterogeneity in both labor market experiences and investment returns, which we argue is essential for simultaneously matching the broader features of all three distributions (earnings, income, and wealth) without an over-reliance on tail events. Similarly, Daminato and Pistaferri (2024) and Smith et al. (2021) emphasize the importance of return heterogeneity (both type and scale dependence) for understanding wealth and consumption inequality.

Our model shares the premise of heterogeneous returns but specifically integrates it with continuous labor income risk to address the tripartite distributional question posed by Quadrini and Rıos-Rull (1997). Other works, such as Krueger et al. (2024), explore continuous-time models with idiosyncratic risk in the context of limited commitment and transition dynamics. While valuable, their focus differs from our detailed steady-state characterization of the joint distributions of earnings, income, and wealth. Our contribution lies in demonstrating that, within a general equilibrium setting, the combination of continuous labor income risk and continuous interest rate risk is sufficient to replicate key empirical regularities across these three dimensions, a specific challenge not always at the forefront of these other important contributions.

While our paper leverages the strengths of continuous-time modeling, significant advancements also continue within the discrete-time paradigm. For instance, Reiter (2023) makes notable progress in enhancing the computational toolkit for discrete-time models by developing methods for second-order perturbations, which can offer higher accuracy when analyzing responses to larger shocks. Similarly, the work of Kaplan and Violante (2022) continues to refine the understanding of consumption behavior within discrete-time Bewley-Huggett-Aiyagari frameworks by exploring aspects such as recursive utility. Methodological contributions like those of Ocampo and Robinson (2024) on histogram iteration methods improve the efficiency of computing moments in discrete-time settings. Furthermore, the application of machine learning and deep learning techniques (e.g., Han et al. (2021); Huang (2023)) is a rapidly growing area aimed at tackling the curse of dimensionality in complex discrete-time models. These discrete-time advancements are crucial for the field; however, our choice of a continuous-time framework is deliberate, motivated by its analytical tractability for handling multiple continuous state variables and its direct application of powerful tools from stochastic calculus. We argue that these features provide a more natural and transparent way to model the specific continuous income and return processes central to our analysis of the subtle interplay between earnings, income, and wealth distributions.

Within our framework, we conduct a policy analysis. We find that an increase in unemployment benefits, i.e., the lower bound of earnings, leads to a rise in wealth inequality. At first sight, this result seems odd. However, it is inherited from the Bewley precautionary savings core of our heterogeneous-agent model. The unemployment benefits serve as insurance for agents. Hence, if policymakers raise the lower bound of income, agents have fewer incentives to save, and thus the left tail of the wealth distribution grows more prominent.

Lastly, we analyze which channels in our model drive the results. We find that we need stochastic labor earnings to generate a left tail and to achieve convergence of the numerical algorithm. Some fraction of agents have to hit the lower bound of earnings to generate sufficient skewness in the wealth distribution. Additionally, we need stochastic interest rates for the right tail. This mechanism moves a small fraction of agents into a "Goldilocks zone," where they already own a large amount of wealth but also generate high returns on investments. Thus, these agents still have a high propensity to save. This result is in line with the findings of Benhabib et al. (2011) and Krusell and Smith (1998). The rest of the paper proceeds as follows. Section 1.2 describes our model and the stationary equilibrium. Section 1.3 provides the numerical algorithms to solve the model, as well as the calibration and resulting distributions it generates. In section 1.4, we use the model to analyze how an increase in unemployment benefits deforms the distributions of earnings, income, and wealth. We then show which mechanisms of the model drive the different results in section 1.5. Lastly, section 1.6 concludes.

1.2 The model

For our study, we consider a heterogeneous agent model à la Bewley-Hugget-Aiyagari set up in continuous time following Achdou et al. (2022). The standard Aiyagari model features stochastic labor endowments that generate some skewness in the distribution of wealth, especially if the labor endowment process is itself skewed and persistent. A substantial body of literature indeed studies models in which households face an uninsurable idiosyncratic labor income risk (typically referred to as Bewley models). Yet, the standard Bewley models of Aiyagari (1994) and Huggett (1993) produce low Gini coefficients and cannot generate heavy tails. More specifically, at high wealth levels, the incentives for precautionary savings taper off, and the right tails of the wealth distribution persist.[2] To model the thick right tail, we introduce stochastic returns. This decision is motivated by the seminal paper of Benhabib et al. (2015), which shows that in Bewley economies with idiosyncratic capital income risk, the wealth distribution is a unique ergodic distribution with a fat (Pareto) tail. Hence, in our model, capital income risk induces the necessary mobility across wealth levels to generate a fat-tailed wealth distribution. Another difference from the standard model is our continuous earnings process, which overcomes the aforementioned "superstar" problem.

1.2.1 Framework

There is a continuum of unit mass of infinitely lived households that are heterogeneous in their labor earnings z, realized interest rate r, and wealth a. Each household consists of one individual, so we use the terms households and individuals interchangeably. The state of the economy is the joint distribution of interest rates, wages, and wealth. Individuals have standard preferences over utility

[2] See Carroll (1997) and Quadrini (2000) for a discussion of these issues.

flows from future consumption c_t discounted at rate $\rho \geq 0$:

$$U_0 \equiv E_0 \int_0^\infty e^{-\rho t} u(c_t)\, \mathrm{d}t\,. \tag{1.1}$$

We assume $u_c' > 0$ and $u_c'' < 0$, i.e., the utility function is concave. Throughout the analysis, the instantaneous utility function is given by:

$$u(c_t) = \begin{cases} \frac{c_t^{1-\gamma}}{1-\gamma} & \text{for } \gamma \neq 1 \\ \log c_t & \text{for } \gamma = 1, \end{cases}$$

where $\gamma > 0$ is the coefficient of relative risk aversion.

Each individual earns a wage z_t, which is governed by a mean-reverting process:

$$\mathrm{d}z_t \;= \varphi(w_t - z_t)\, \mathrm{d}t \;+ \nu z_t\, \mathrm{d}W_t\,, \tag{1.2}$$

where $\mathrm{d}W_t$ is a Brownian motion. The parameter φ governs the speed of mean reversion, and w_t corresponds to the mean level of z_t. We vary the parameter w_t, so that it can be matched with market-clearing conditions in equilibrium. For the variance, we use the specification of Courtadon (1982). Fischer (2018) argues that this process is well-suited to model labor income dynamics.[3] Hence, ν is a constant that is scaled by the current wage level. Since w_t has continuous trajectories, it cannot take negative values. The wage of the individual is paid as an endowment of the economy's final good.

The standard Bewley-Aiyagari model generates too few agents that hold low levels of wealth, and the concentration of wealth among the richest agents is far too small. In the U.S. data, 12 percent of the population has below-zero wealth, whereas the wealthiest 5 percent of the population hold roughly 53 percent of the wealth. One of the main purposes of our line of research is to extend the standard macroeconomic framework to allow heterogeneity among agents. Thus, it is important that the heterogeneity in the new framework is quantitatively adequate to match the observed income and wealth distribution in the United States. To generate a fatter right tail, it is necessary to either endow rich agents with a higher propensity to save or to give them higher returns on savings (or both). Krusell and Smith (1998) give them a higher propensity to save by assuming a stochastic discount factor.

In contrast, Benhabib et al. (2011) introduce a Merton's portfolio selection problem to enable agents to achieve a higher return on capital, and thus give them a higher savings rate. We follow their approach and introduce two investment options. First, we include a risk-free bond b_t that is in zero

[3] The Courtadon process has an inverse Gamma asymptotic distribution. Fischer (2018) shows that its Gini coefficient can be computed directly via

$$Gini(z) = \frac{\Gamma(\alpha - 0.5)}{\Gamma(\alpha)\sqrt{\pi}},$$

where $\alpha = 2\zeta/\nu^2 + 1$. Thus, for parameter combinations that result in an $\alpha < 2$, we obtain Gini coefficients greater than 0.4.

net supply and whose return r^f is constant over time and across agents. Second, we allow agents to invest their wealth in productive capital k_t, whose return follows a mean-reverting process:

$$\mathrm{d}r_t = \kappa(\theta_t - r_t)\,\mathrm{d}t + \eta\sqrt{r_t}\,\mathrm{d}W_t\,, \tag{1.3}$$

where again $\mathrm{d}W_t$ is a Brownian motion, κ governs the speed of mean reversion, and θ_t corresponds to the mean level of r_t. We let the parameter θ_t vary, so that we can match it with market-clearing conditions in equilibrium later. For the variance η, we use the specification of Cox et al. (1985).[4] This model has the advantage that the process r_t can take negative values. Hence, agents are able to realize negative returns on capital. To enforce no-arbitrage, we further assume the capital stock is predetermined, i.e., instantaneous adjustment costs are 100%.

Further, agents' realized returns on those financial assets vary stochastically. The realized return evolves according to

$$\mathrm{d}R = r_t\,\mathrm{d}t + \sigma\,\mathrm{d}B_t\,, \tag{1.4}$$

where the Brownian motion $\mathrm{d}B_t$ is independent of the Brownian motion $\mathrm{d}W_t$. Hence, an investment k leads to the return $k \cdot \mathrm{d}R = k(r_t\,\mathrm{d}t + \sigma\,\mathrm{d}B_t) = kr_t\,\mathrm{d}t + k\sigma\,\mathrm{d}B_t$. This decomposition has a nice interpretation. As in Longstaff and Schwartz (1992), we have a two-factor model, where the first term $X_1 = r_t\,\mathrm{d}t$ reflects the systematic risk, whereas the second term $X_2 = \sigma\,\mathrm{d}B_t$ describes the idiosyncratic risk. Putting everything together, the wealth of an individual can be expressed as follows[5]:

$$\mathrm{d}k_t + \mathrm{d}b_t = (b_t r^f + k_t r_t + z_t - c_t)\,\mathrm{d}t + \sigma k_t\,\mathrm{d}W_t\,. \tag{1.5}$$

The agents now face an optimal portfolio allocation problem, as in Merton (1969). We simplify the problem by writing the budget constraint in terms of wealth or net worth, $a_t = b_t + k_t$:

$$\mathrm{d}a_t = (a_t r^f + k_t(r_t - r^f) + z_t - c_t)\,\mathrm{d}t + \sigma k_t\,\mathrm{d}W_t\,. \tag{1.6}$$

Finally, agents also face a borrowing limit:

$$a_t \geq -a_{min}, \tag{1.7}$$

where $a_{min} \in \mathbb{R}_+ \coloneqq [0, \infty)$.

Capital satisfies $k_t \geq 0$; hence, the borrowing constraint $b_t \geq -a_{min}$ can be written as

$$k_t \leq a_t + a_{min}. \tag{1.8}$$

[4] Other specifications for the short-rate process are possible as well. For example, one could also use the Vasicek (1977) specification. In this case, the variance is η and thus can take negative values. We checked this specification, and all algebra and numerical implementations are carried out exactly the same. For other short-rate specifications, see Aït-Sahalia (1996) for an excellent overview.

[5] Note that we could also account for growth in our setting. Let g be a constant growth rate. By detrending all variables by e^{gt}, we obtain our setting.

Individuals maximize (1.1) subject to (1.3), (1.5), and (1.7), taking as given the evolution of θ_t and w_t for $t \geq 0$. There are no private insurance markets for the household-specific endowment shocks. However, the wealth accumulation process in Equation (1.5) creates a mechanism used by agents to insure themselves against shocks and allows for consumption smoothing, as demonstrated in Bewley (1977).

We close the model by following Aiyagari (1994); hence, wealth takes the form of productive capital that is used by a representative firm, which also hires labor. The total amount of capital supplied in the economy equals the total amount of capital k. In a stationary equilibrium, it is given by:

$$K = \int_{\Omega} k(a, r, z) p(a, r, z) \, \mathrm{d}(a, r, z) \, ,$$

where $\Omega \coloneqq [-a_{min}, \infty) \times [0, \infty) \times [0, \infty)$. We further assume capital depreciates at the rate δ and there exists a representative firm with a Cobb-Douglas production function $Y = F(K, L) = K^{\alpha} L^{1-\alpha}$. Since factor markets are competitive, the wage and the risk-free interest rate are given by:

$$\begin{aligned} \theta &= \partial_K F(K, 1) - \delta = \alpha K^{\alpha - 1} - \delta, \\ w &= \partial_L F(K, 1) = (1 - \alpha) K^{\alpha}. \end{aligned} \tag{1.9}$$

The risky asset is now a real asset in the sense that k_t units produce $r_t k_t$ units of physical output, and only positive asset positions are possible. One interpretation of the risky asset is that r_t is the return from owning and running a private firm.[6] Nevertheless, other interpretations are possible.

1.2.2 Stationary equilibrium

An individual's consumption-saving decision and the evolution of the joint distribution of their income, interest rates, and wealth can be described by a coupled system of partial differential equations. More precisely, a *Hamilton-Jacobi-Bellman* (HJB) equation and a *Fokker-Planck* (FP) equation, also known as a *Kolmogorov Forward* equation, i.e., a framework that Lasry and Lions (2007) named a *Mean Field Game*.

Theorem 1.2.1 (Stationary equilibrium of the stochastic $\theta - w$-model). *Let the following drifts be given:*

$$\begin{aligned} s &\coloneqq z + a r^f + k(r - r^f) - c, \\ \psi &\coloneqq \kappa(\theta - r), \\ \mu &\coloneqq \varphi(w - z), \end{aligned}$$

[6] For example, assume that private firms produce profit using capital and labor based on a constant returns to scale production function $Z_t f(k_t, l_t)$ as in Angeletos (2007), and define

$$r_t k_t = \max_{l_t} \{ Z_t f(k_t, l_t) - w_t l_t - \delta k_t \}.$$

Then, the process for r_t inherits the properties of the process for Z_t. See Quadrini (2009) and Cagetti and De Nardi (2006) for related models of private firms.

.

The stationary equilibrium of the stochastic $\theta - w$-model takes the form:

$$\rho v = \max_{\substack{0<c \\ 0\leq k\leq a+a_{min}}} \left\{ u(c) + \psi v_r + s v_a + \mu v_z + \frac{\eta^2 r}{2} v_{rr} + \frac{\sigma^2 k^2}{2} v_{aa} + \frac{\nu^2 z^2}{2} v_{zz} \right\}, \tag{1.10a}$$

$$0 = \partial_a \left(\partial_a \frac{\sigma^2 k^2}{2} p - sp \right) + \partial_r \left(\partial_r \frac{\eta^2 r}{2} p - \psi p \right) + \partial_z \left(\partial_z \frac{\nu^2 z^2}{2} p - \mu p \right), \tag{1.10b}$$

with

$$\int_\Omega p(a,r,z)\,\mathrm{d}(a,r,z) \,=1, \tag{1.10c}$$

and

$$\begin{aligned} K &= \int_\Omega k(a,r,z) p(a,r,z)\,\mathrm{d}(a,r,z)\,, \\ \theta &= \alpha K^{\alpha-1} - \delta, \\ w &= (1-\alpha) K^\alpha. \end{aligned} \tag{1.10d}$$

We call s the savings policy function*.*

Proof. See Remark A.1.3 and Remark A.2.3. Further, since firms maximize profit, define the profit function of the firm $\Pi \coloneqq F(K,L) - wL - \theta K - \delta K$, and the maximization of Π with respect to K, L yields (1.10d). □

From the stationary equilibrium described in (1.10) we obtain the stationary distributions for earnings z, net wealth a, and returns on investment r. As economists, we are also interested in the distribution of total income. The standard procedure to obtain it would be to find the stochastic process of total income and derive a new stationary equilibrium model (1.10a)-(1.10d) with new variables (a, r, ι), where ι stands for total income. However, this approach is cumbersome and unnecessary, since the information regarding total income is hidden in the former result. We elucidate it in the following corollary.

Corollary 1.2.2. *Let $\iota \coloneqq z + ar^f + k(r - r^f)$ denote the* income policy function*. Then, the stationary income distribution is given by*

$$q(\iota) = \int_{-a_{min}}^{\infty} \int_0^\infty \frac{p(a,r,z(a,r,\iota))}{1 + (r-r^f) k_z(a,r,z(a,r,\iota))}\,\mathrm{d}r\;\mathrm{d}a\,, \tag{1.11}$$

where

$$z(a,r,\iota) = (I + ar^f + (r-r^f)k(a,r,\cdot))^{-1}(\iota) \tag{1.12}$$

and

$$(I + ar^f + (r-r^f)k(a,r,\cdot))^{-1}$$

denotes the inverse function of $z + ar^f + (r-r^f)k(a,r,z)$ with respect to z.

Proof. See Appendix A.3. □

Theorem 1.2.1 and Corollary 1.2.2 provide the stationary behavior of the economy described by our model. Attempting to find a closed-form solution for (1.10a)-(1.10d) can become rather involved if not impossible, so we use numerical methods to solve (1.10). Nevertheless, this is not an impediment, as we will show in Section 1.3. With that, we have acquired the necessary tools—consumption and investment policy function, distributions—in order to calibrate our model to show that it can match empirical observations from microeconomic data.

It is important to note that the optimal problem (1.10a)-(1.10d) is nonstandard, in the sense that the optimality condition of capital k alters the "usual structure"—see Lasry and Lions (2007, Section 1.2). Moreover, there is a nontrivial fact in Corollary 1.2.2: k may depend on z, since k_z does not need to vanish identically.

1.3 Numerical analysis

Now that we have established the mathematical setup for our model and provided the necessary results to study the economy described by it in Theorem 1.2.1 and Corollary 1.2.2, we move on to the numerical results. First, we give a compact version of the algorithms we use to obtain the numerical solution of 1.10 and 1.11. We lay out the complete description of the algorithms in Appendix A.5 and provide the codes with this paper. Second, we calibrate the model to match the observed distributions of earnings, income, and wealth. At the end of the section, we compare the results with empirical data.

1.3.1 Algorithm

We base our implementation on the methods described in Achdou et al. (2022), but we differ in some steps. Therefore, we describe our whole algorithm here in detail. First, we need an auxiliary result for the algorithm to compute the stationary distribution of our model, which we present here. Even though it seems to be well-known, we are unaware of any reference that it can be traced back to; for this reason, we give a proof of it in Appendix A.4. Nevertheless, the proof in the appendix not only proves Theorem 1.3.1, but it pushes the boundary a bit further. It is tailor-made for the class of models we discuss here. First, by defining the linear differential operator associated to (1.10a), we obtain its formal adjoint to be the linear operator in (1.10b) by including the appropriate boundary conditions in $\mathcal{D}(A^*)$. Then, we see that p is in fact an element of $\ker A^*$. Third, we argue that we can use these results to obtain a numerical solution to the discretized problem. Thus, other researchers can adopt this method when solving continuous-time heterogeneous-agent models.

Theorem 1.3.1. *Let A be the linear operator associated to* (1.10a)*,*

$$Av = \psi v_r + s v_a + \mu v_z + \frac{\eta^2 r}{2} v_{rr} + \frac{\sigma^2 k^2}{2} v_{aa} + \frac{\nu^2 z^2}{2} v_{zz}, \quad v \in \mathcal{D}(A).$$

Then, the solution to the stationary FP equation (1.10b) *is given by the eigenfunction to the eigenvalue* 0 *of* A^**, the adjoint operator of* A*; that is,*

$$A^* p = 0.$$

Consequently, if $\mathcal{A}$ *is a discretization of* A*, then the discretization* $\mathcal{P}$ *of* p *is the eigenvector corresponding to the* 0 *eigenvalue of the matrix* $\mathcal{A}^\top$*.*

Proof. See Appendix A.4. □

This result offers a neat approach to calculating a stationary distribution from the stationary HJB equation analytically. Additionally, it can be used to efficiently compute the stationary distribution of a sparse matrix numerically. These type of matrices naturally appear in an upwind scheme, as stated below, and every good software package has pre-implemented algorithms to quickly compute their eigenvalues and eigenvectors.[7]

Using the notation of Theorem 1.3.1, note that the sparsity of the matrix $\mathcal{A}$ corresponding to the discretization of A plays an important role, since the size of the matrix increases by N^3 for an N points mesh in each of the variables (a, r, z). If the matrix were not sparse, there would be no hope of carrying out this process numerically in a reasonable amount of time. This sparsity property is one of the main advantages of our continuous-time framework. In a discrete framework, this property disappears, and the number of computations needed to solve a model in three dimensions increases substantially. To overcome this issue, Bayer and Luetticke (2018) recently developed a perturbation method to adapt the sparsity property in a discrete-time setup.[8] However, they need to linearize their model around a stationary equilibrium to apply their algorithm. Thus, they lose the nonlinear dynamics, which can be quite rich, as we show in Section 1.4.

Combining the previous results, we now have collected all the ingredients to state our numerical algorithm:

[7] For example, we use the `eigs(AT,1,0)` command in Matlab, where AT denotes the transposed matrix of the stationary, discretized operator of the HJB equation; the option 1 tells Matlab to return one eigenvalue; and 0 denotes the initial guess for the eigenvalue, invoking Theorem 1.3.1.

[8] They also discuss the issue of computational resources needed to solve discrete-time heterogeneous-agent models with aggregate risk and idiosyncratic states much deeper than we do.

Algorithm 1: Solving Model (1.10)

1 Choose a tolerance $\varepsilon > 0$, a maximum value of iterations it_{max}, and initial guesses v^0, θ^0, w^0, K^0.
2 **for** $i = 1, \ldots, \text{it}_{max}$ **do**
3 Given $v^{i-1}, K^{i-1}, \theta^{i-1}, w^{i-1}$ solve the HJB equation (1.10a) using an implicit upwind scheme until the error is smaller than ε, and set the solution to v^i.
4 Use these values to solve the FP equation (1.10b) using Theorem 1.3.1, and normalize the solution according to (1.10c). Set p^i as the solution of (1.10b).
5 With the results, update K^i, θ^i, w^i using (1.10d).
6 **if** *the relative error between* $(K^{i-1}, \theta^{i-1}, w^{i-1})$ *and* (K^i, θ^i, w^i) *is smaller than* ε **then**
7 ,
8 **break**

If Algorithm 1 converges before it_{max} is reached, we call the last computed update $(v^i, p^i, K^i, \theta^i, w^i)$ a stationary equilibrium of (1.10). Importantly, we use the supremum norm for the convergence between two consecutive iterations in line **3**. Hence, Theorem 2.1 in Barles and Souganidis (1991) guarantees that by refining the mesh, the approximate solution converges to the viscosity solution[9] of the HJB equation (1.10a)

The exact steps are presented in Appendix A.5. If Algorithm 1 converged, we obtain the optimal controls for the consumption and savings policy functions and the steady-state values for aggregate capital, K, the mean level of wages, w, and interest rates, θ, as well as the stationary distributions for earnings, interest rates, and wealth. Hence, we still need to compute the stationary distribution of overall income. However, invoking Corollary 1.2.2, this is straightforward. The following code snippet sketches how to recover the income distribution (1.11):

1.3.2 Calibration and results

We continue calibrating the model and show that it can replicate some key facts of the empirically observed distributions. To this end, we use parameters that can be found in the existing literature. Table 1.1 presents the exact calibration of our structural parameters.

Using Algorithms 1 and 2, we now solve the model and describe the results.

Table 1.2 summarizes the key facts of our model. It is able to replicate all empirically observed Gini indices. In fact, we match the Gini index of earnings precisely because we calibrate the earnings process directly. Nevertheless, our model also mirrors the Gini indices of income and wealth as a result of the optimal decisions of agents in a stationary general equilibrium, which are matched indirectly.

[9] For an introduction to the topic of viscosity solutions of second-order equations, see Lions (1983), and for an overview on the topic, check Lions and Souganidis (1988).

Algorithm 2: Obtaining the income distribution (1.11)

1 Let $(v^i, p^i, K^i, \theta^i, w^i)$ be the solution obtained in Algorithm 1.

2 Compute $J \coloneqq 1 + (r - r^f)k_z(a, r, z)$ and check that $J > 0$.

3 Define $\iota \coloneqq z + ar^f + k(r - r^f)$ and interpolate $(v^i, p^i, K^i, \theta^i, w^i)$ and (a, r, z, ι).

4 Find $z(a, r, \iota)$ as in (1.12) numerically using the interpolated variables; that is, the following holds

$$\iota = z(a, r, \iota) + ar^f + k(a, r, z(a, r, \iota))(r - r^f).$$

5 Compute $q(a, r, \iota) \coloneqq p(a, r, z(a, r, \iota))$ and redefine

$$J(a, r, \iota) \coloneqq 1 + (r - r^f)k_z(a, r, z(a, r, \iota)).$$

6 Calculate $q(\iota)$ as in (1.11) by using

$$q(\iota) = \int_{-a_{min}}^{\infty} \int_0^{\infty} \frac{q(a, r, \iota)}{J(a, r, \iota)} \, \mathrm{d}r \, \mathrm{d}a \, .$$

Table 1.1: Calibration of the θ-w-model.

Parameter	Explanation	Value	Target/Source
ρ	Discount factor	0.05	Achdou et al. (2022)
γ	Coefficient of relative risk aversion	2	Achdou et al. (2022)
r^f	Risk-free rate	0.01	Safe assets' return
σ	Volatility of wealth	0.2	Vol. of stock ind.
κ	Speed of mean reversion investments	0.72	
η	Volatility of investment	0.2	Vol. of stock ind.
ζ	Speed of mean reversion wages	0.2	To match the data
ν	Volatility of wages	2	To match the data
α	Capital's share of output	1/3	
δ	Depreciation rate	0.025	

The model also captures the bottom tail of the wealth distribution almost exactly. It only fails to fully account for the thick, right tail of the wealth distribution. The top 30% share is matched well, but the very top of the distribution generates too little mass. One could possibly fix this by including jumps in the earnings process, obtaining some "superstars." For instance, Wang (2007) proposes a jump-diffusion process for income and obtains a fat, right tail of the income distribution, which propagates through to the wealth distribution.

Table 1.2: Results of the calibrated model. (Data Source: Kuhn and Rios-Rull (2016, Tab. 3 & Appendix B2)).

		Quintile					Top Groups (Percentile)		
Economy	Gini	First	Second	Third	Forth	Fifth	90th–95th	95–99th	99th–100th
		A. Distribution of earnings							
United States	.67	1.4	5	10.6	19.6	63.4	12	17.2	17.9
Benchmark	.67	52.7	0	0	24.8	22.4	4.9	4.3	1
		B. Distribution of incomes							
United States	.58	3	6.5	10.9	18.1	61.4	10.8	16.5	19.7
Benchmark	.57	0.7	9	7.1	20.9	69.4	17.4	21.6	8.8
		C. Distribution of wealth							
United States	.85	2.8	4.1	6.5	12.9	73.8	11	24.8	26.1
Benchmark	.88	-7.7	6.1	5.7	13.8	87.8	19	31.7	15.5

1.4 The effects of a minimum wage on wealth inequality

Having a model available that explains the distributions of earnings, income, and wealth, we can now focus on the implications of different policy experiments. One of the most critical questions in any policy decision is "Who gains and who loses?" Standard representative agent models cannot take such considerations into account, since they only feature one agent by construction. Hence, we need a model like ours to evaluate all the distributional impacts of a policy.

We illustrate the mechanisms in our model by analyzing a rise in the minimum wage. Hence, we change the behavior of our model at the lower bound of wages. By implementing a minimum wage, the process (1.2) now becomes an instantaneous reflected Wiener process. Thus, it becomes important to use a Neumann boundary condition at the lower boundary, i.e., $f'(z_{min}) = 0$, as described by Engelbert and Peskir (2014).

When we think of inequality, we always have to answer the question "inequality in what?" simultaneously. First, it is no surprise that raising the lower bound of earnings reduces the Gini coefficient in overall earnings, lowering income inequality. Countryman (1999) documents this effect empirically for Canada between 1975 and 1996, and Koeniger et al. (2007) present similar results for eleven OECD countries between 1973 and 1998. Our model also features this mechanism, and Figure 1.1 shows the quantitative results. However, the more interesting questions relate to the effects on income and

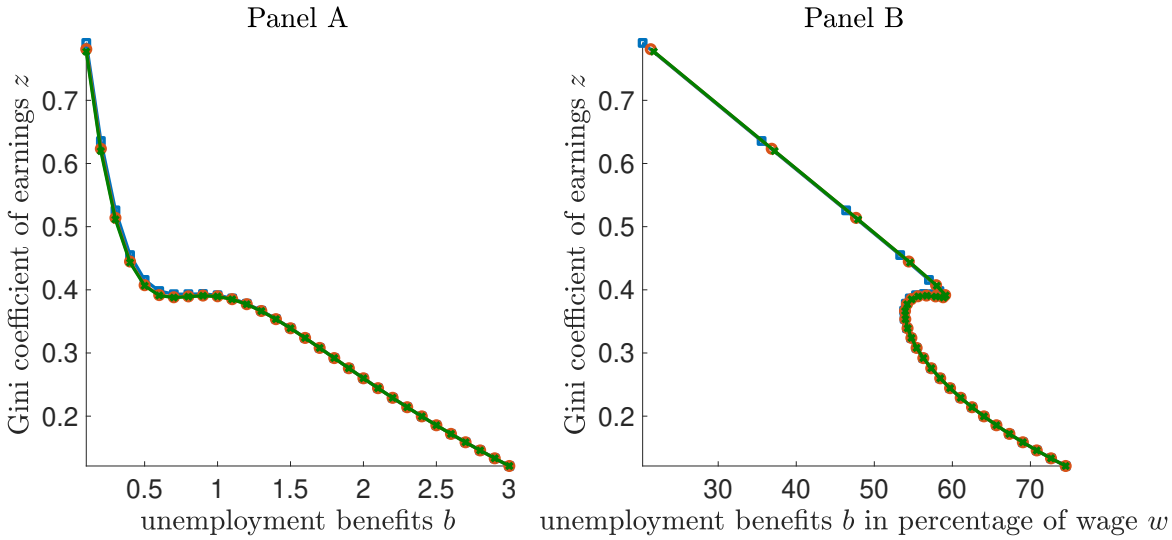


Figure 1.1: **Gini coefficient of earnings for different levels of unemployment benefits**
Panel A shows the Gini coefficient of earnings for different levels of unemployment benefits, revealing that the Gini coefficient is decreasing in all specifications. Panel B shows the Gini coefficient of earnings for different levels of unemployment benefits in relation to the equilibrium wage for different levels of the risk-free rate r^f. The blue-$\square$-line is connected to a risk-free rate of 1%, the orange-$\circ$-line to a risk-free rate of 2%, and the green-$\times$-line to a risk-free rate of 3%. Ceteris paribus, inequality is falling as unemployment benefits grow.

wealth inequality.

To analyze these redistributional effects, we first calibrate our model according to Table 1.1 again. Afterwards, we vary the risk-free rate and the variable of interest, i.e., the lowest possible value of z. In our model, we interpret the minimum as the minimum grid point of the z-process. Hence, the lowest possible value of z represents agents who are working for the minimum wage.

Figure 1.2 displays the results for income inequality. Our general equilibrium model features highly nonlinear effects. The nonlinear feedback effect on income is unclear. However, for U.S. unemployment benefits of 50%, income inequality is rising.[10] But with slightly higher values, income inequality decreases in our calibration.

We present the results on wealth inequality in Figure 1.3. The model implies that a rise in unemployment benefits leads to an increase in the Gini index of wealth, and hence to an increase in overall wealth inequality. Again, the equilibrium effects are highly nonlinear. As with income the income distribution, there is a window where inequality decreases, but it vanishes with a rising risk-free rate. Further, by comparing Figures 1.2 and 1.3, we see that the sharp fall in income inequality is

[10] Nickell (1997, Tab. 4) summarizes "replacement rates," which show what share of income is replaced by unemployment benefits, and the duration of these benefits, for different OECD countries.

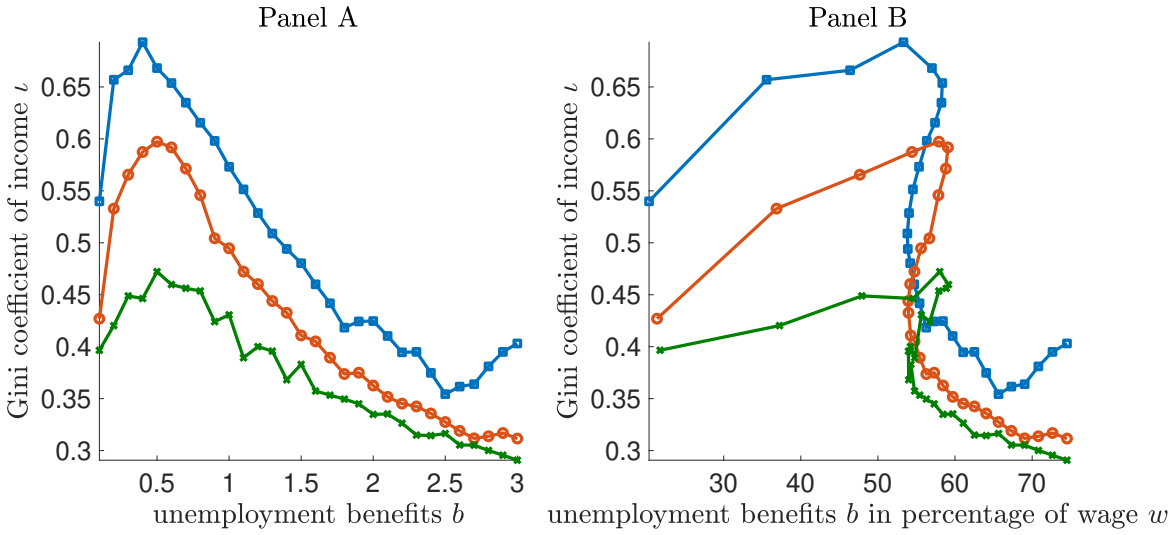


Figure 1.2: **Gini coefficient of income for different levels of unemployment benefits**
Panel A shows the Gini coefficient of income for different levels of unemployment benefits. Panel B shows the same Gini coefficient but displays the unemployment benefits in relation to the equilibrium wage for different levels of the risk-free rate r^f. The blue-□-line is based on a risk-free rate of 1%, the orange-∘-line on a risk-free rate of 2%, and the green-×-line on a risk-free rate of 3%. Ceteris paribus, for low levels of unemployment benefits, income inequality grows but then sharply declines. For unrealistically high values of b, income inequality starts rising again.

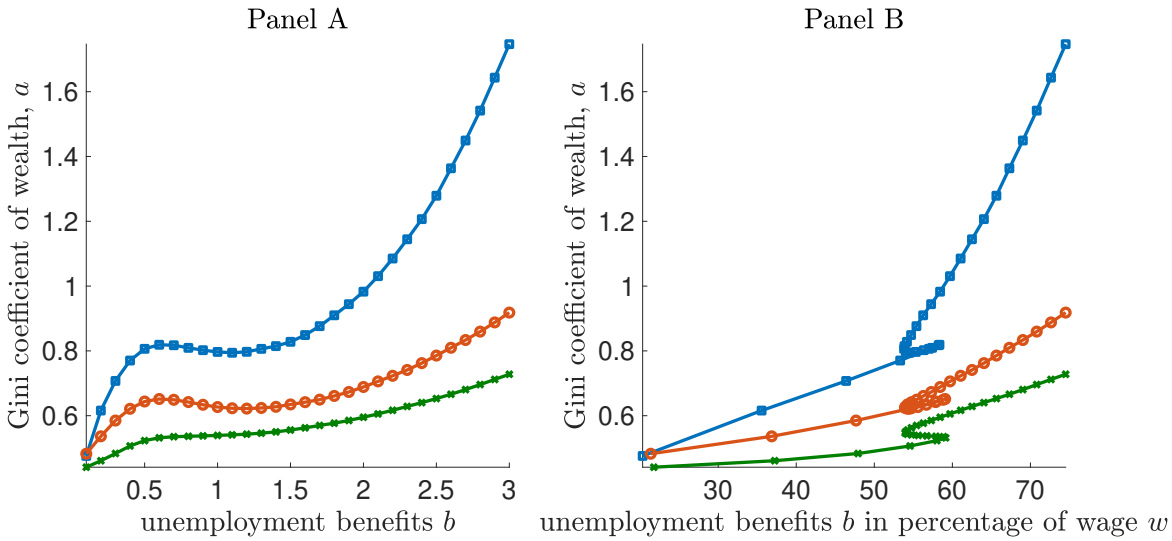


Figure 1.3: **Gini coefficient of wealth for different levels of unemployment benefits**
Panel A shows the Gini coefficient for different levels of unemployment benefits. Panel B shows b as a percentage of wages in relation to different levels of the risk-free rate r^f. The blue-□-line denotes a risk-free rate of 1%, the orange-∘-line a risk-free rate of 2%, and the green-×-line a risk-free rate of 3%. Ceteris paribus, inequality grows alongside an increasing level of unemployment benefits. Despite a small window of unemployment benefits where inequality decreases, we only observe this effect for unemployment benefits higher than 55% of the mean wage, a figure much higher than U.S. unemployment benefits.

accompanied by an explosion of wealth inequality.

The increase in wealth inequality is a self-fulfilling prophecy. Within a Bewley framework, one can think of unemployment benefits as insurance against losing a job.[11] Hence, if policymakers raise the lower bound of earnings, agents with low incomes will, under optimal conditions, choose lower savings. Consequently, in general equilibrium, they will reduce their overall wealth. With a low level of unemployment benefits in relation to the mean wage, low-income households have to save more to smooth their consumption intertemporally.

1.5 Channels of inequality

We now shed light on the different drivers of inequality in our model. The model features two main sources of heterogeneity. First, agents receive different wages from their labor. Second, agents differ in their ability to invest. In our setting, it is straightforward to set each process constant and elaborate on the effects on the different distributions.

To begin with, we switch off the optimal portfolio choice problem by giving each agent the same rate of return on investment, which is equal to the risk-free rate. Hence, the only source of inequality arises from heterogeneity in earnings. This setting is close to the standard Bewley-Aiyagari model. The only difference is that our model still features a continuous state space of possible labor earnings. However, we expect the standard results of these models. More precisely, we lose the flexibility to obtain a thick upper tail of the wealth distribution.

We display the results in column three of Table 1.3. We are still able to replicate the Gini coefficient of earnings. The model also matches the lower half of the wealth distribution. As expected, we lose the capability of fit the very top of the wealth distribution. However, this is no surprise. As aforementioned in the Bewley-Aiyagari model, wealthy agents have a lower propensity to save, since their earnings from income are already huge in relation to their labor earnings. Thus, they have no incentive to save more to guard against labor market fluctuations.

The table below summarizes the channels of inequality. The full model refers to our benchmark model (1.10), which we compare to the model with a constant labor process. Hence, in this specification, every agent receives the same labor income. Second, we compare it with a model without a portfolio problem. In this specification, all agents generate the same returns on their investments. We calibrate all parameters according to Table 1.1 again.

Next, we switch off the earnings process. This move causes some inconveniences. First, setting z constant, we do not obtain an earnings distribution by construction. Second, neither $r^f + k_a(r - r^f)$ nor $k_r(r - r^f) + k(r > r^f)$ is strictly positive/negative definite. Thus, we cannot use the techniques of Corollary 1.2.2 to infer an income distribution. Third, the resulting wealth distribution is no longer

[11] In our diffusion process framework, one can think of "losing" as multiple negative realizations of a normal random variable, reaching the lower bound of earnings.

Table 1.3: **Channels of inequality**

	Full Model	r constant	z constant	Data
Wealth Top 1%	15.27%	4.8%	385%	34%
Wealth Top 5%	39.21%	16.9%	457%	53%
Wealth Top 10%	52.12 %	29.1%	446%	67%
Wealth Top 20%	72.13%	46.6%	400%	85%
Wealth Top 30%	85.26%	68.44%	400%	88%
Wealth Bottom 50%	0.61%	0.96%	-300%	0.7%
Gini Earnings	0.39	0.39	—	0.39
Gini Income	0.64	0.43	—	0.58
Gini Wealth	0.82	0.63	5.14	0.85

useful. Nevertheless, we include the results in Table 1.3 for completeness.

1.6 Conclusion

In this chapter, we propose a new heterogeneous-agent model that explains the different shapes of earnings, income, and wealth distributions in general equilibrium in interaction with each other. We develop the model in continuous time and provide new methods to obtain solutions. Further, we analyze a policy change in our benchmark model. We demonstrate that an increase in the lower bound for wages reduces inequality in labor earnings, but the trade-off is higher inequality in wealth. We then shed light on the different drivers of our model. We argue that we need both inequality in labor earnings and returns on investment for an accurate description of the empirically observed distributions.

Our model can be used in several ways. On the one hand, we now have a model that displays the most important distributions of inequality in an economy. Hence, researchers can use our core to analyze richer models that also include a government sector. This incorporation would be especially useful to analyze the distributional effects of policies. As shown in Section 1.4, it is often unclear which distributions are affected by a policy change and in which form. For example, a government can use different tools in the form of taxes to redistribute wealth in an economy. But what are the effects of such policies on earnings and the distribution of income? More broadly, should governments use taxes on labor earnings, capital earnings, or overall wealth to achieve their goals?

On the other hand, our framework can be used for estimation. There is still no established technique for estimating a continuous-time heterogeneous-agent model, though Parra-Alvarez et al. (2017) re-

cently made some progress by using maximum-likelihood estimation. However, they use the standard Aiyagari model of Achdou et al. (2022) as an example for their estimation procedure. Thus, they are not able to recover realistic parameters because there simply is no parameter combination of the Aiyagari model that can explain the observed wealth distribution.[12] Our model can solve this problem because we have shown that there is a calibration that fits the observed distributions.

A.1 Derivation of the HJB equations

We derive the *Hamilton-Jacobi-Bellman equations* corresponding to the $\theta - w$-model. To this end, we recapitulate it in concise form as an Itô diffusion. The model is given by the stochastic processes

$$\begin{aligned} \mathrm{d}a_t &= \Big(r_t^f a_t + k_t(a_t, r_t, z_t)(r_t - r_t^f) + z_t - c_t(a_t, r_t, z_t)\Big)\,\mathrm{d}t + \sigma k_t\,\mathrm{d}W_t^1\,, \\ \mathrm{d}r_t &= \psi(t, r_t)\,\mathrm{d}t + \eta\sqrt{r_t}\,\mathrm{d}W_t^2\,, \\ \mathrm{d}z_t &= \mu(t, z_t)\,\mathrm{d}t + \nu z_t\,\mathrm{d}W_t^3\,, \end{aligned} \tag{1.13}$$

where $\sigma, \eta, \nu > 0$, $r^f \in C(\mathbb{R}_+, \mathbb{R})$, $\psi, \mu \in C({\mathbb{R}_+}^2, \mathbb{R})$, and (W_t^1, W_t^2, W_t^3) denotes a 3-dimensional Brownian motion. Here, $(c, k) : \mathbb{R}_+ \times \mathbb{R}^3 \to {\mathbb{R}_+}^2$ denotes our *Markov-control*. Further, we wish to extend this system by the additional constraints

$$0 \le k_t(a_t, r_t, z_t) \le a_t + a_{min}, \quad 0 < c_t(a_t, r_t, z_t) \tag{1.14}$$

for some fixed $a_{min} \in \mathbb{R}_+$. Accordingly, we define

$$D_a \coloneqq \left\{\ (c,k) \in {\mathbb{R}_+}^2 \ \middle|\ 0 \le k \le a + a_{min},\ 0 < c\ \right\}.$$

The set of such solutions is denoted by

$$\mathcal{S} \coloneqq \{(a, r, z, c, k) \text{ as described above satisfies (1.13) and (1.14)}\}$$

and

$$\mathcal{S}_{t_0,a_0,r_0,z_0} \coloneqq \{(a, r, z, c, k) \in \mathcal{S} | (a_{t_0}, r_{t_0}, z_{t_0}) = (a_0, r_0, z_0)\}\,.$$

For the formulation of the optimal control problem, assume that

$$\forall (a, r, z, c, k) \in \mathcal{S}, \tau \in \mathbb{R}_+ : \quad E_\tau \int_\tau^\infty e^{-\int_\tau^t \rho(\xi)\,\mathrm{d}\xi}\,|u(t, c_t(a_t, r_t, z_t))|\,\mathrm{d}t < \infty, \tag{H1}$$

for some *instantaneous utility function* $u \in C^1(\mathbb{R}_+ \times (0,\infty), \mathbb{R})$ and *time preference rate* $\rho \in C(\mathbb{R}_+, \mathbb{R}_+)$. Our optimal control problem is now described by the *optimal cost function*, which is defined by

$$v(t_0, a_0, r_0, z_0) \coloneqq \sup_{(a,r,z,c,k) \in \mathcal{S}_{t_0,a_0,r_0,z_0}} E_{t_0} \int_{t_0}^\infty e^{-\int_{t_0}^t \rho(\xi)\,\mathrm{d}\xi}\,u(t, c_t(a_t, r_t, z_t))\,\mathrm{d}t\,, \tag{1.15}$$

for $(t_0, a_0, r_0, z_0) \in \mathbb{R}_+ \times \Omega$.

For the treatment of this classic and well-known optimal control problem, we follow the rigorous approach of Øksendal (2014) in order to derive the Hamilton-Jacobi-Bellman equation.

[12] The authors would like to thank Olaf Posch for pointing this out.

Theorem A.1.1 (The Hamilton-Jacobi-Bellman equation).
Suppose that $V \in C^2(\mathbb{R}_+ \times \Omega, \mathbb{R})$.[13] *Then,*

$$\begin{aligned}\rho(\tau)v(\tau,a,r,z) = \sup_{(c,k)\in D_a} \Big\{ & u(\tau,c) + v_\tau(\tau,a,r,z) \\ & + (ar_\tau^f + k(r-r_\tau^f) + z - c)v_a(\tau,a,r,z) + \psi(\tau,r)v_r(\tau,a,r,z) + \mu(\tau,z)v_z(\tau,a,r,z) \\ & + \frac{\sigma^2k^2}{2}v_{aa}(\tau,a,r,z) + \frac{\eta^2 r}{2}v_{rr}(\tau,a,r,z) + \frac{\nu^2z^2}{2}v_{zz}(\tau,a,r,z)\Big\},\end{aligned} \tag{1.16}$$

for all $(\tau,a,r,z) \in \mathbb{R}_+ \times \Omega$.

Proof. In order to derive the Hamilton-Jacobi-Bellman equation (2.2), we aim to use Øksendal (2014, Theorem 11.2.1). To this end, we first present an optimization problem fitting this framework, which will yield a Hamilton-Jacobi-Bellman equation, allowing us to derive the statement of this theorem. Consider the following system of equations

$$\begin{aligned}\mathrm{d}a_t &= \Big((a_t + a_{min})k_t'(a_t,r_t,z_t)(r_t - r_t^f) - c_t(a_t,r_t,z_t) + r_t^f a_t + z_t\Big)\,\mathrm{d}t \\ &\quad + \sigma k_t'(a_t,r_t,z_t)\cdot(a_t+a_{min})\,\mathrm{d}W_t^1 \\ \mathrm{d}r_t &= \psi(t,r_t)\,\mathrm{d}t + \eta\sqrt{r_t}\,\mathrm{d}W_t^2, \\ \mathrm{d}z_t &= \mu(t,z_t)\,\mathrm{d}t + \nu z_t\,\mathrm{d}W_t^3,\end{aligned} \tag{1.17}$$

with $(c,k') : \mathbb{R}_+ \times \mathbb{R}^3 \to \mathbb{R}_+ \times [0,1]$ subject to the constraint

$$0 \le k_t'(a_t,r_t,z_t) \le 1, \quad 0 \le a_t + a_{min}, \tag{1.18}$$

for all $t \in \mathbb{R}_+$ for which the set of solutions $S'_{\tau,a,r,z}$ is defined as for the original system of equations (1.13). Here we define the optimal value function v' at (s,a_0,r_0,z_0) as

$$\sup_{\substack{(a,r,z,c,k)\in \\ \mathcal{S}'_{\tau,a_0,r_0,z_0}}} E_\tau \int_\tau^\infty e^{-\int_0^t \rho(\xi)\,\mathrm{d}\xi}\, u\left(t, c_t(a_t,r_t,z_t)\right)\,\mathrm{d}t\,,$$

i.e., we conducted the change of variables

$$\begin{aligned}k_t(a,r,z) &= k_t'(a,r,z)\cdot(a+a_{min}), \\ v(\tau,a,r,z) &= e^{\int_0^\tau \rho(\xi)\,\mathrm{d}\xi}\cdot v'(\tau,a,r,z).\end{aligned}$$

Note that if $a + a_{min} = 0$, then (1.14) and (1.18) guarantee that this assignment is indeed one-to-one.

[13] Indices on v represent partial derivatives and not time-dependency, as with other parameters.

Applying Øksendal (2014, Theorem 11.2.1),[14] we obtain

$$\begin{aligned}0 = &\\ \sup_{(c,k')\in\mathbb{R}_+\times[0,1]} &\left\{\frac{1}{2}\left[\eta^2 r v'_{rr}(\tau,a,r,z) + \nu^2 z^2 v'_{zz}(\tau,a,r,z) + \sigma^2 k'^2 (a+a_{min})^2 v'_{aa}(\tau,a,r,z)\right]\right.\\ &+ e^{-\int_0^\tau \rho(\xi)\,\mathrm{d}\xi} u(\tau,c) + \left[(a+a_{min})k'(r-r^f_\tau) - c + r^f_\tau a + z\right] v'_a(\tau,a,r,z)\\ &\left.+ v'_\tau(\tau,a,r,z) + \psi(\tau,r)v'_r(\tau,a,r,z) + \mu(\tau,z)v'_z(\tau,a,r,z)\right\}.\end{aligned}$$

Together with $v'_\tau(\tau,a,r,z) = e^{-\int_0^\tau \rho(\xi)\,\mathrm{d}\xi} v_\tau(\tau,a,r,z) - \rho(\tau)e^{-\int_0^\tau \rho(\xi)\,\mathrm{d}\xi} v(\tau,a,r,z)$ and undoing the change of variables yields the first statement of the theorem. □

Proposition A.1.2. *In the context of Theorem A.1.1, assume that the supremum in* (2.2) *is attained for some* $(\tau,a,r,z)\in\mathbb{R}_+\times\Omega$ *at* $(c^*,k^*)\in D_a$*. If* (c^*,k^*) *lies in the interior of* D_a*,* $\mathring{D}_a$*, then*

$$\begin{aligned}0 =& u_c(\tau,c^*) - v_a(\tau,a,r,z),\\ 0 =& \sigma^2 k^* v_{aa}(\tau,a,r,z) + (r-r^f_\tau)v_a(\tau,a,r,z).\end{aligned}$$

More generally, if $u_c(\tau,\cdot)$ *is invertible,* $v_{aa}(\tau,a,r,z)\neq 0$*,* $v_a(\tau,a,r,z)\lim_{c\to 0}u(\tau,c) = -\infty$ *and* $\lim_{c\to\infty}\frac{u(\tau,c)}{c} = 0$*, then*

$$\begin{aligned}c^* &= u_c(\tau,\cdot)^{-1}\left(v_a(\tau,a,r,z)\right),\\ k^* &\in \arg\max_{k\in\Lambda}\left\{\tfrac{1}{2}\sigma^2 k^2 v_{aa}(\tau,a,r,z) + k(r-r^f_\tau)v_a(\tau,a,r,z)\right\},\\ \Lambda &\coloneqq \left\{0, -\frac{(r-r^f_\tau)}{\sigma^2}\frac{v_a(\tau,a,r,z)}{v_{aa}(\tau,a,r,z)}, a+a_{min}\right\}\cap[0,a+a_{min}].\end{aligned}$$

Proof. For the given $(\tau,a,r,z)\in\mathbb{R}_+\times\Omega$, define

$$\begin{aligned}A(c) =& u(\tau,c) - c v_a(\tau,a,r,z),\\ B(k) =& \frac{1}{2}\sigma^2 k^2 v_{aa}(\tau,a,r,z) + k(r-r^f_\tau)v_a(\tau,a,r,z),\\ R =& v_\tau(\tau,a,r,z) + z v_a(\tau,a,r,z) + \psi(\tau,r)v_r(\tau,a,r,z) + \mu(\tau,z)v_z(\tau,a,r,z)\\ &+ \frac{1}{2}\left[\eta^2 r v_{rr}(\tau,a,r,z) + \nu^2 z^2 v_{zz}(\tau,a,r,z)\right],\end{aligned}$$

for $(c,k)\in D_a$. Then (2.2) reads

$$\begin{aligned}\rho(\tau)v(\tau,a,r,z) &= \sup_{(c,k)\in D_a} A(c)+B(k)+R\\ &= \max_{(c,k)\in \bar{D}_a} A(c)+B(k)+R\end{aligned}$$

[14] Øksendal (2014, Theorem 11.2.1) makes more technical assumptions than those displayed in our theorem. These are, of course, necessary and implicitly assumed, but omitted for better exposure.

$$= \max_{(c,k)\in \mathring{D}_a} A(c) + \max_{(c,k)\in \mathring{D}_a} B(k) + R.$$

Since c^*, k^* are maxima of A and B, respectively, we know by the first order condition that

$$\begin{aligned} 0 = A_c(c^*) =& u_c(s, c^*) - v_a(\tau, a, r, z), \\ 0 = B_k(k^*) =& \sigma^2 k^* v_{aa}(\tau, a, r, z) + (r - r^f_\tau) v_a(\tau, a, r, z). \end{aligned}$$

For the second statement of the theorem, note that these conditions read

$$\begin{aligned} c^* =& u_c(\tau, \cdot)^{-1} \left(v_a(\tau, a, r, z)\right), \\ k^* =& -\frac{(r - r^f_\tau)}{\sigma^2} \frac{v_a(\tau, a, r, z)}{v_{aa}(\tau, a, r, z)}. \end{aligned}$$

Since $\lim_{c\to\infty} \frac{u(\tau,c)}{c} = 0$ and $v_a(\tau, a, r, z) > 0$, we derive $\lim_{c\to\infty} u(\tau, c) - c v_a(\tau, a, r, z) = -\infty$. Together with $\lim_{c\to 0} u(\tau, c) = -\infty$, we ensure that the supremum of $A(c)$ is indeed a maximum in c^*. Note that if the supremum of $B(k)$ is not achieved in k^*, it is in either 0 or $a + a_{min}$. □

Remark A.1.3. Sometimes stationary systems arise, i.e., $\rho, u, r^f, u, \gamma, \mu$ in (1.13) are time-independent functions. An example of this situation is (1.10a). This additional structure can be exploited in order to simplify the characterization of the value function. In the following, we discuss how the lack of time-dependency can be reflected in the solution of the Hamilton-Jacobi-Bellman equation (2.2). Since the system described by (1.13) is autonomous and we consider an infinite time horizon, there is a trivial one-to-one correspondence between the solution sets $\mathcal{S}_{t_0,a_0,r_0,z_0}$ and $\mathcal{S}_{t_1,a_0,r_0,z_0}$ for every $t_0, t_1 \in \mathbb{R}_+$, namely by means of the time shift $t \mapsto t - t_0 + t_1$. This mapping being a homeomorphism of a derivative equal to 1, the transformation formula directly yields

$$\begin{aligned} v(t_0, a_0, r_0, z_0) =& \sup_{(a,r,z,c,k)\in\mathcal{S}_{t_0,a_0,r_0,z_0}} E_{t_0} \int_{t_0}^{\infty} e^{\rho t} u(c_t(a_t, r_t, z_t))\, \mathrm{d}t \\ =& \sup_{(a,r,z,c,k)\in\mathcal{S}_{t_1,a_0,r_0,z_0}} E_{t_1} \int_{t_1}^{\infty} e^{\rho t} u(c_t(a_t, r_t, z_t))\, \mathrm{d}t \\ =& v(t_1, a_0, r_0, z_0), \end{aligned}$$

for all $(t_0, a_0, r_0, z_0), (t_1, a_0, r_0, z_0) \in \mathbb{R}_+ \times \Omega$. In the context of Theorem A.1.1, this translates to $v_t = 0$ and we may rewrite (2.2) as

$$\begin{aligned} &\rho v(r, a, z) = \\ &\sup_{(c,k)\in D_a} \Big\{ u(c) + \gamma(r) v_r(r, a, z) + \mu(z) v_z(a, r, z) + (a r^f + k(r - r^f) + z - c) v_a(a, r, z) \\ &\qquad + \frac{1}{2}\Big[\eta(r)^2 v_{rr}(a, r, z) + \nu(z)^2 v_{zz}(a, r, z) + \sigma^2 k^2 v_{aa}(a, r, z)\Big]\Big\}. \end{aligned} \tag{1.19}$$

We also indicate that similar simplifications can be performed for the results of Proposition A.1.2. In this light, we also see that the optimal Markov control may be time-invariant, even if we did not explicitly require it in the formulation of the time-invariant optimal control problem.

A.2 Derivation of the FP equation

Since c and k are time-dependent, the whole problem becomes non-autonomous, and the usual derivation of the FP equation by means of Dynkin's formula is not valid, see Knäble (2012) for a case in which the derivation is done for a particular class of non-autonomous stochastic processes. For the general derivation of the FP equation with non-autonomous coefficients and transition probabilities, we refer to Öttinger (1996, Section 3), Arnold (1974, Section 5), Kallianpur (1980, Section 9), and Stroock and Varadhan (1997, Section 5). This calculation requires the notion of evolution operators, in particular two-parameter semigroups.

Since we are dealing with $\Omega \neq \mathbb{R}^3$, for $p \in W^{1,1}_{\text{loc}}(0,\infty; W^{2,2}(\Omega))$,[15] it is highly convenient to define the current $j = (j^a, j^r, j^z)$ by

$$\begin{aligned} j^a(t,a,r,z) &\coloneqq \partial_a \frac{\sigma^2 k(t,a,r,z)^2}{2} p(t,a,r,z) - s(t,a,r,z)p(t,a,r,z), \\ j^r(t,a,r,z) &\coloneqq \partial_r \frac{\eta^2 r}{2} p(t,a,r,z) - \alpha(t,r)p(t,a,r,z), \\ j^z(t,a,r,z) &\coloneqq \partial_z \frac{\nu^2 z^2}{2} p(t,a,r,z) - \mu(t,z)p(t,a,r,z), \end{aligned}$$

which will play a role in the proof of the next result.

Theorem A.2.1. *Assume that the stochastic differential equations* (1.2), (1.3), *and* (1.6) *have a unique solution.*[16] *Let* $p_0 \in L^1(\Omega) \cap L^2(\Omega)$ *be given with* $p_0 \geq 0$ *and*

$$\int_\Omega p_0(a,r,z)\,\mathrm{d}(a,r,z) = 1.$$

Then the PDF p *associated to* (1.2), (1.3), *and* (1.6) *satisfies*

$$\begin{aligned} \partial_t p &= \partial_a \left(\partial_a \tfrac{\sigma^2 k^2}{2} p - sp \right) + \partial_r \left(\partial_r \tfrac{\eta^2 r}{2} p - \alpha p \right) + \partial_z \left(\partial_z \tfrac{\nu^2 z^2}{2} p - \mu p \right), \quad t > 0, \\ p(0) &= p_0, \end{aligned} \tag{1.20}$$

with

$$\int_\Omega p\,\mathrm{d}(a,r,z) = 1, \quad \forall t \geq 0.$$

Proof. Let us define the family of time-dependent second-order differential operators $A^*(t)$ by

$$A^*(t)p \coloneqq \partial_a \left(\partial_a \frac{\sigma^2 k(t)^2}{2} p - s(tp) \right) + \partial_r \left(\partial_r \frac{\eta^2 r}{2} p - \alpha(t)p \right) + \partial_z \left(\partial_z \frac{\nu^2 z^2}{2} p - \mu(t)p \right),$$

where we have omitted the dependence on (a,r,z) for clarity.

[15] For the definition of the Lebesgue spaces $L^p(\Omega)$ and Sobolev spaces $W^{k,p}(\Omega)$ see Adams (1975, Chapters II–IV). For the Bochner spaces $W^{k,p}(0,T;B)$, where B is a Banach space, see Adams (1975, Chapter VII) and Evans (2010, Section 5.9.2).

[16] This is well known for the process r_t; see Cox et al. (1985). Otherwise, it is enough to assume Lipschitz continuity on the drifts and diffusion terms as well as affine-linear growth, both uniformly in time; see Øksendal (2014, Theorem 5.2.1).

Denote by $p_{t_0}(a,r,z;t,a',r',z')$ the PDF associated to the transition probability function from t_0 to t of (1.2), (1.3), and (1.6). From Kallianpur (1980, Theorem 9.4.4) and subsequent remarks, we see that $p_{t_0}(a,r,z;t,a',r',z')$ satisfies

$$\partial_t p_{t_0}(t,a,r,z;a',r',z') = A^*(t,a,r,z)p_{t_0}(t,a,r,z;a',r',z') \tag{1.21}$$

with

$$p_{t_0}(t_0,a,r,z;a',r',z') = \delta(a-a')\delta(r-r')\delta(z-z').$$

Let $p(t,a,r,z) \coloneqq p_{t_0}(t,a,r,z;\cdot,\cdot,\cdot) * p_0$. Then, $p(t_0,a,r,z) = p_0(a,r,z)$. Moreover, by using the properties of the convolution, if we convolute p_0 with (1.21), we obtain (1.20).
In order to show that the mass is conserved—that is, for all $t \geq 0$, it holds that

$$\int_\Omega p \,\mathrm{d}(a,r,z) = 1.$$

Note that the FP equation can be regarded as a continuity equation that expresses the conservation of the probability when we introduce the probability current j; see Öttinger (1996, Section 3.3.5). Since the processes a_t, r_t, z_t are confined in the region Ω, there is no flow probability across the boundary of Ω; that is,

$$\begin{aligned} j_a(t,-a_{min},r,z) = 0, &\quad \forall (t,r,z) \in (0,\infty)\times(0,\infty)\times(0,\infty) \\ j_r(t,a,0,z) = 0, &\quad \forall (t,a,z) \in (0,\infty)\times(-a_{min},\infty)\times(0,\infty) \\ j_z(t,a,r,0) = 0, &\quad \forall (t,a,r) \in (0,\infty)\times(-a_{min},\infty)\times(0,\infty). \end{aligned}$$

We can integrate the FP equation in Ω to obtain

$$\int_\Omega \partial_t p \,\mathrm{d}(a,r,z) = \int_\Omega A^*(t)p \,\mathrm{d}(a,r,z)\,.$$

By using the Stokes formula—integration by parts in higher dimension—we have that for $p \in \mathcal{D}(A^*(t))$, the following holds:

$$\begin{aligned} \int_\Omega A^*(t)p \,\mathrm{d}(a,r,z) &= \int_\Omega \nabla \cdot j \,\mathrm{d}(a,r,z) \\ &= \int_0^\infty \int_0^\infty j_a(t,-a_{min},r,z)\,\mathrm{d}r\ \mathrm{d}z \\ &\quad + \int_0^\infty \int_{-a_{min}}^\infty j_r(t,a,0,z)\,\mathrm{d}a\ \mathrm{d}z \\ &\quad + \int_0^\infty \int_{-a_{min}}^\infty j_z(t,a,r,0)\,\mathrm{d}a\ \mathrm{d}r \\ &= 0. \end{aligned}$$

Hence,

$$\int_\Omega \partial_t p \,\mathrm{d}(a,r,z) = 0.$$

If we have enough regularity, integration and differentiation commute so that

$$\partial_t \int_\Omega p \,\mathrm{d}(a,r,z) \,= 0.$$

Integrating from 0 to t leads

$$\int_\Omega p \,\mathrm{d}(a,r,z) \,- \int_\Omega p_0 \,\mathrm{d}(a,r,z) \,= 0.$$

Since

$$\int_\Omega p_0 \,\mathrm{d}(a,r,z) \,= 1,$$

we have that

$$\int_\Omega p \,\mathrm{d}(a,r,z) \,= 1, \quad \forall t \geq 0.$$

□

Remark A.2.2. Note that the FP equation can be written as an abstract, non-autonomous Cauchy problem of the form

$$\begin{aligned} \dot{p}(t) &= A^*(t)p(t), \quad t > 0, \\ p(0) &= p_0, \end{aligned}$$

with $A^*(t) : \mathcal{D}(A^*(t)) \subset L^2(\Omega) \to L^2(\Omega)$ defined by

$$A^*(t)p \coloneqq \partial_a \left(\partial_a \frac{\sigma^2 k(t)^2}{2} p - s(t)p \right) + \partial_r \left(\partial_r \frac{\eta^2 r}{2} p - \alpha(t)p \right) + \partial_z \left(\partial_z \frac{\nu^2 z^2}{2} p - \mu(t)p \right),$$

and

$$\mathcal{D}(A^*(t)) = \{p \in W^{2,2}(\Omega) \mid j_a|_{a=-a_{min}} = j_r|_{r=0} = j_z|_{z=0} = 0\}.$$

The notion of vanishing currents at the boundary of Ω is often called *reflecting boundary*.

Remark A.2.3. The stationary Fokker-Planck equation arises under the assumption that the coefficients become time-independent and there is no time change, so that the time derivative vanishes; that is:

$$\begin{aligned} 0 &= \partial_a \left(\partial_a \tfrac{\sigma^2 k^2}{2} p - sp \right) + \partial_r \left(\partial_r \tfrac{\eta^2 r}{2} p - \alpha p \right) + \partial_z \left(\partial_z \tfrac{\nu^2 z^2}{2} p - \mu p \right), \\ 1 &= \int_\Omega p \,\mathrm{d}(a,r,z)\,. \end{aligned}$$

A.3 Income distribution

Here, we give a proof of Corollary 1.2.2. First, observe that $z + ar^f + k(r - r^f)$ is the income ι. The ratio of this quantity per earning is then given by

$$J \coloneqq \partial_z \iota = 1 + (r - r^f)k_z.$$

From an economic viewpoint, it is clear that in most cases, the quantity J is larger than 1, since the more one earns, the more one invests. However, there might be a region where the capital is so large that it does not pay off to invest more in the risky asset k, and J is then slightly smaller than 1. Hence, we can assume that $J > 0$ and, by the Inverse Function Theorem, $z + ar^f + k(a, r, z)(r - r^f)$ is globally locally invertible with respect to z, and we denote the inverse by $(I + ar^f + k(a, r, \cdot)(r - r^f))^{-1}$. We now define the transformation $\Psi : \Sigma \to \Omega$ by

$$\Psi(a, r, \iota) \coloneqq \begin{pmatrix} a \\ r \\ (I + ar^f + k(a, r, \cdot)(r - r^f))^{-1}(\iota) \end{pmatrix}.$$

Note that Ψ is invertible, and the inverse is precisely

$$\Psi^{-1}(a, r, z) = \begin{pmatrix} a \\ r \\ z + ar^f + k(r - r^f) \end{pmatrix}.$$

Hence, we can compute the distribution as a function of (a, r, ι) by using the change of variables formula

$$q(a, r, \iota) = p(\Psi(a, r, \iota))|\det(D\Psi(a, r, \iota))|.$$

Set $z(a, r, \iota) \coloneqq \Psi_3(a, r, \iota)$. From the Inverse Function Theorem, we also have that

$$D\Psi(a, r, \iota) = ([D\Psi^{-1}](\Psi(a, r, \iota)))^{-1} = ([D\Psi^{-1}](a, r, z(a, r, \iota)))^{-1}.$$

Using the property of the determinant $\det(M^{-1}) = (\det(M))^{-1}$, we further have that

$$\det(D\Psi(a, r, \iota)) = \frac{1}{\det([D\Psi^{-1}](a, r, z(a, r, \iota)))}.$$

It is an easy computation to see that

$$D\Psi^{-1}(a, r, z) = J.$$

Thus,

$$q(a, r, \iota) = \frac{p(a, r, z(a, r, \iota))}{1 + (r - r^f)k_z(a, r, z(a, r, \iota))}.$$

A.4 Stationary distribution as an eigenvector

Here, we give a proof of Theorem 1.3.1. We consider the case in which z_t is modeled by a CIR process, since the mathematical description is simpler, as we only have one PDE. Nevertheless, the same argumentation can be used when considering the two weakly coupled PDEs that arise from the two-state Poisson process $z_t \in \{z_1, z_2\}$, as in Achdou et al. (2022).

Recall that $\Omega = [-a_{min}, \infty) \times [0, \infty) \times [0, \infty)$. Observe that Ω can be trivially identified with $\mathbb{R}^3_+$ by implementing the change $a \mapsto a + a_{min}$. Moreover, under such an affine transformation, the differential operator ∂_a remains unchanged, so that without loss of generality, we could work with $\Omega \equiv \mathbb{R}^3_+$. Points of Ω will be denoted by (a, r, z).

Consider the stationary HJB equation:

$$\rho v = \max_{\substack{0<c \\ 0\leq k\leq a+a_{min}}} \left\{ u(c) + \alpha v_r + s v_a + \mu v_z + \frac{\eta^2 r}{2} v_{rr} + \frac{\sigma^2 k^2}{2} v_{aa} + \frac{\nu^2 z^2}{2} v_{zz} \right\}. \tag{1.22}$$

Assume that (1.22) is *well-posed*; that is, there exists a function v such that $v_a > 0$, $v_{aa} < 0$ and the optimal problem has a solution c, k, where c is the optimal cost and k the optimal capital. Note that by using the definition of s, we have

$$s v_a + \frac{\sigma^2 k^2}{2} v_{aa} = (z + r^f a - c) v_a + (r - r^f) k v_a + \frac{\sigma^2 k^2}{2} v_{aa}.$$

Hence, if $r \leq r^f$, it turns out that the optimal capital is $k \equiv 0$. For $r > r^f$,

$$k = \min \left\{ -\frac{(r - r^f) v_a}{\sigma^2 v_{aa}}, a + a_{min} \right\}.$$

Also, for $a = -a_{min}$ it follows from (1.22) that $k \equiv 0$. Note that if v is regular enough, k is continuous at $r = r^f$ and hence k^2 is continuously differentiable at $r = r^f$. Thus, we will further assume that $c \in C^{0,1}(\Omega)$ and $k^2 \in C^{1,1}(\Omega)$, which is plausible if we take into account the prior considerations.

Consider the linear operator associated to (1.22), namely,

$$\mathfrak{A} v = \alpha v_r + s v_a + \mu v_z + \frac{\eta^2 r}{2} v_{rr} + \frac{\sigma^2 k^2}{2} v_{aa} + \frac{\nu^2 z^2}{2} v_{zz}.$$

We take the formal adjoint operator $A^* : \mathcal{D}(A^*) \subset L^2(\Omega) \to L^2(\Omega)$ to be

$$A^* p = \partial_a \left(\partial_a \frac{\sigma^2 k^2}{2} p - s p \right) + \partial_r \left(\partial_r \frac{\eta^2 r}{2} p - \alpha p \right) + \partial_z \left(\partial_z \frac{\nu^2 z^2}{2} p - \mu p \right)$$

for $p \in \mathcal{D}(A^*)$.

Recall the current $j = (j^a, j^r, j^z)$, which is given by

$$\begin{aligned} j^a &\coloneqq \partial_a \frac{\sigma^2 k^2}{2} p - s p, \\ j^r &\coloneqq \partial_r \frac{\eta^2 r}{2} p - \alpha p, \\ j^z &\coloneqq \partial_z \frac{\nu^2 z^2}{2} p - \mu p, \end{aligned}$$

so that $A^* p = \nabla \cdot j$. The problem $A^* p = 0$ describes the well-known stationary Fokker-Planck equation, and the domain of the operator A^*, $\mathcal{D}(A^*)$, inherits the boundary conditions of the dynamic FP equation described in Appendix A.2, namely, that the currents vanish at the boundary,

$$j^a|_{a=-a_{min}} = 0,$$

$$j^r|_{r=0} = 0,$$
$$j^z|_{z=0} = 0,$$

c.f. Risken (1989, Section 4.7.1) and Öttinger (1996, Section 3.3.5).

Since Ω is unbounded in a, r, and z, the functions need to decay at infinity. Moreover, $p \geq 0$ with

$$\int_\Omega p(a,r,z)\,\mathrm{d}(a,r,z) = 1.$$

Hence, we are looking for nontrivial solutions of $A^*p = 0$, that is, $p \neq 0$. For that to be possible, the assumption that there is $p \geq 0$ in $\ker A^*$ naturally arises. The existence of solutions and uniqueness—also for the degenerate case, where the diffusion coefficients are not strictly positive—of divergence type FP equations can be found in Bogachev et al. (2015, Chapter 2), in particular Bogachev et al. (2015, Corollary 2.4.4) for the case $\Omega = \mathbb{R}^3$. The results can be adapted to our case by considering the vanishing currents at the boundary.

Let

$$\mathcal{D}(A^*) \coloneqq \{p \in W^{2,2}(\Omega) \mid j^a|_{a=-a_{min}} = j^r|_{r=0} = j^z|_{z=0} = 0\} \tag{1.23}$$

and note that $\mathcal{D}(A^*)$ is not the maximal domain of A^*. The operator $A^* : \mathcal{D}(A^*) \subset L^2(\Omega) \to L^2(\Omega)$ is *closed* in the sense of Goldberg (1966, Chapter II) and *densely defined*; that is, $\mathcal{D}(A^*)$ is dense in $L^2(\Omega)$, since $C_0^\infty(\overline{\Omega})$ is dense in $W^{2,2}(\Omega) \subset \mathcal{D}(A)$ by Grisvard (1985, Theorem 1.4.2.1). Hence, the adjoint operator of A^* is again closed, $(A^*)^* = A$ and $\mathcal{D}(A)$ is dense in $L^2(\Omega)$, as in Tucsnak and Weiss (2009, Proposition 2.8.1). Moreover, $[\mathcal{R}(A)]^\perp = \ker A^*$ and $[\mathcal{R}(A^*)]^\perp = \ker A$; see Tucsnak and Weiss (2009, Remark 2.8.2). A further computation also shows that $Av = \mathfrak{A}|_{\mathcal{D}(A)}v$, where

$$\mathcal{D}(A) = \{v \in W^{2,2}(\Omega) \mid \sigma^2 k^2 V_a|_{a=-a_{min}} = \eta^2 r V_r|_{r=0} = \nu^2 z^2 V_z|_{z=0} = 0\}. \tag{1.24}$$

Hence, the assumption $p \in \ker A^*$ implies $p \in [\mathcal{R}(A)]^\perp$. Note that the boundary conditions included in $\mathcal{D}(A)$ and $\mathcal{D}(A^*)$ impose decays of the functions at the boundaries $a = -a_{min}$, $r = 0$, and $z = 0$. The nontrivial question of which boundary conditions to impose on v arises when trying to solve (1.22). It seems reasonable to use the boundary conditions in $\mathcal{D}(A)$ to solve the HJB equation, so that we solve

$$\begin{aligned} \rho v &= \max_{\substack{0<c \\ 0\leq k\leq a+a_{min}}} \{u(c) + Av\}, \\ 0 &= A^*p. \end{aligned} \tag{1.25}$$

We will further assume that 0 is not an accumulation point of the spectrum of A^*. Otherwise, we could not really guarantee that an approximation of p converges to p and not to another eigenfunction with an eigenvalue arbitrarily close to 0.

After discretizing the HJB, we end up with an equation of the form

$$\rho\mathcal{V} = u(\mathcal{C}) + \mathcal{A}\mathcal{V},$$

where $\mathcal{V}, \mathcal{C}$ are column vectors with the values of v, c at the nodes, and $\mathcal{A}$ is a matrix that corresponds to the discretization of A. Note that a discretization of A^* is in fact $\mathcal{A}^\top$. Hence, by $[\mathcal{R}(A)]^\perp = \ker A^*$, it follows that there is a row vector $\mathcal{V} \neq 0$ such that $\mathcal{V}\mathcal{A} = 0$. If we transpose the former, we have $\mathcal{A}^\top \mathcal{V}^\top = 0$. Thus, $\mathcal{P} = \mathcal{V}^\top$ is a discretization of p, which is certainly the eigenvector corresponding to the eigenvalue 0 of $\mathcal{A}^\top$.

A.5 Numeric scheme for the equilibrium model

Here, we elucidate how to solve the equilibrium model (1.10) numerically. The first step is to solve the HJB equation. More precisely, we wish to solve (1.25). As in Appendix A.4, we will discuss the CIR case, but the method can be performed analogously with other models.

We begin by deriving a finite-difference scheme using the upwind method. To this end, let us consider a partition of Ω with M, N, L points in a, r, z, respectively. The resulting domain is denoted by $\Omega_{NML} \coloneqq [a_1, a_M] \times [r_1, r_N] \times [z_1, z_L]$. The points in the nodes will be denoted by a_i, r_j, z_l. Similarly, the evaluation of a function f at a nodal point will be denoted by $f_{ijl} = f(a_i, r_j, z_l)$. Sometimes, in order to avoid any confusion, we will separate the subscripts with a comma, e.g., $f_{1,2,3}$ instead of f_{123}. In order to fully appreciate the subtleties of the model, we will use a nonhomogeneous scheme; that is, the partition of Ω will not be assumed to be regular. For two consecutive points in $[a_1, a_M]$, $[r_1, r_n]$, and $[z_1, z_L]$, we set $\mathrm{d}a_i = a_i - a_{i-1}$, $\mathrm{d}r_j = r_j - r_{j-1}$ and $\mathrm{d}z_l = z_l - z_{l-1}$, respectively.

As in Algorithm 1, guess v^0, K^0, θ^0, w^0. Define

$$\begin{array}{rclrcl}
\psi_j^+ & \coloneqq & \max\{\psi(r_j), 0\}, & \psi_j^- & \coloneqq & \min\{\psi(r_j), 0\}, \\
s_{ijl}^+ & \coloneqq & \max\{s(a_i, r_j, z_l; k_{ijl}, c_{ijl}), 0\}, & s_{ijl}^- & \coloneqq & \min\{s(a_i, r_j, z_l; k_{ijl}, c_{ijl}), 0\}, \\
\mu_l^+ & \coloneqq & \max\{\mu(z_l), 0\}, & \mu_l^- & \coloneqq & \min\{\mu(z_l), 0\},
\end{array}$$

so that

$$\begin{aligned}
F_{ijl} \coloneqq\ & u(c_{ijl}) - \rho v_{ijl} + \frac{\psi_j^+}{\mathrm{d}r_{j+1}}(v_{ij+1l} - v_{ijl}) + \frac{\psi_j^-}{\mathrm{d}r_j}(v_{ijl} - v_{ij-1l}) \\
& + \frac{s_{ijl}^+}{\mathrm{d}a_{i+1}}(v_{i+1jl} - v_{ijl}) + \frac{s_{ijl}^-}{\mathrm{d}a_i}(v_{ijl} - v_{i-1jl}) \\
& + \frac{\mu_l^+}{\mathrm{d}z_{l+1}}(v_{ijl+1} - v_{ijl}) + \frac{\mu_l^-}{\mathrm{d}z_l}(v_{ijl} - v_{ijl-1}) \\
& + \frac{\eta^2 r_j}{2} \frac{\frac{v_{ij+1l} - v_{ijl}}{\mathrm{d}r_{j+1}} - \frac{v_{ijl} - v_{ij-1l}}{\mathrm{d}r_j}}{\mathrm{d}r_{j+1}} \\
& + \frac{\sigma^2 k_{ijl}^2}{2} \frac{\frac{v_{i+1jl} - v_{ijl}}{\mathrm{d}a_{i+1}} - \frac{v_{ijl} - v_{i-1jl}}{\mathrm{d}a_i}}{\mathrm{d}a_{i+1}} \\
& + \frac{\nu^2 z_l^2}{2} \frac{\frac{v_{ijl+1} - v_{ijl}}{\mathrm{d}z_{l+1}} - \frac{v_{ijl} - v_{ijl-1}}{\mathrm{d}z_l}}{\mathrm{d}z_{l+1}}.
\end{aligned}$$

A possible way of rewriting the former is

$$\begin{aligned} F_{ijl} = u(c_{ijl}) &+ \zeta_j v_{ij+1l} + \chi_j v_{ij-1l} \\ &+ z_{ijl} v_{i+1jl} + x_{ijl} v_{i-1jl} \\ &+ \Theta_l v_{ijl+1} + \Psi_l v_{ijl-1} \\ &+ y_{ijl} v_{ijl} - \rho v_{ijl}, \end{aligned}$$

where

$$\zeta_j \coloneqq \frac{\psi_j^+}{\mathrm{d}r_{j+1}} + \frac{\eta^2 r_j}{2\,\mathrm{d}r_{j+1}{}^2}, \quad \chi_j \coloneqq -\frac{\psi_j^-}{\mathrm{d}r_j} + \frac{\eta^2 r_j}{2\,\mathrm{d}r_{j+1}\,\mathrm{d}r_j},$$

$$z_{ijl} \coloneqq \frac{s_{ijl}^+}{\mathrm{d}a_{i+1}} + \frac{\sigma^2 k_{ijl}^2}{2\,\mathrm{d}a_{i+1}{}^2}, \quad x_{ijl} \coloneqq -\frac{s_{ijl}^-}{\mathrm{d}a_i} + \frac{\sigma^2 k_{ijl}^2}{2\,\mathrm{d}a_{i+1}\,\mathrm{d}a_i},$$

$$\Theta_l \coloneqq \frac{\mu_l^+}{\mathrm{d}z_{l+1}} + \frac{\nu^2 z_l^2}{2\,\mathrm{d}z_{l+1}{}^2}, \quad \Psi_l \coloneqq -\frac{\mu_l^-}{\mathrm{d}z_l} + \frac{\nu^2 z_l^2}{2\,\mathrm{d}z_{l+1}\,\mathrm{d}z_l},$$

and

$$y_{ijl} \coloneqq -(\zeta_j + \chi_j + z_{ijl} + x_{ijl} + \Theta_l + \Psi_l).$$

By using the upwind method, we can guarantee that $y_{ijl} < 0$ independent of the drifts ψ, s, μ.
Since we solve the equation in the bounded domain Ω_{NML}, we need boundary conditions for the extremes of the domain. For $j = 1$ and $j = N$, we have that

$$\frac{v_{i,1,l} - v_{i,0,l}}{\mathrm{d}r_1} = 0, \quad \frac{v_{i,N+1,l} - v_{i,N,l}}{\mathrm{d}r_{N+1}} = 0$$

, so that

$$\begin{aligned} F_{i,1,l} = u(c_{i,1,l}) &+ \zeta_1 v_{i,2,l} \\ &+ x_{i,1,l} v_{i-1,1,l} + z_{i,1,l} v_{i+1,1,l} \\ &+ \Theta_l v_{i,1,l+1} + \Psi_l v_{i,1,l-1} \\ &+ (y_{i,1,l} + \chi_1) v_{i,1,l} - \rho v_{i,1,l} \end{aligned}$$

and

$$\begin{aligned} F_{i,N,l} = u(c_{i,N,l}) &+ \chi_N v_{i,N-1,l} \\ &+ x_{i,N,l} v_{i-1,N,l} + z_{i,N,l} v_{i+1,N,l} \\ &+ \Theta_l v_{i,N,l+1} + \Psi_l v_{i,N,l-1} \\ &+ (y_{i,N,l} + \zeta_N) v_{i,N,l} - \rho v_{i,N,l}, \end{aligned}$$

respectively.
For $l = 1$ and $l = L$, we have that

$$\frac{v_{i,j,1} - v_{i,j,0}}{\mathrm{d}z_1} = 0, \quad \frac{v_{i,j,L+1} - v_{i,j,L}}{\mathrm{d}z_{L+1}} = 0,$$

which leads to

$$\begin{aligned} F_{i,j,1} = u(c_{i,j,1}) &+ \zeta_j v_{i,j+1,1} + \chi_j v_{i,j-1,1} \\ &+ z_{i,j,1} v_{i+1,j,1} + x_{i,j,1} v_{i-1,j,1} \\ &+ \Theta_1 v_{i,j,2} \\ &+ (y_{i,j,1} + \Psi_1) v_{i,j,1} - \rho v_{i,j,1} \end{aligned}$$

and

$$\begin{aligned} F_{i,j,L} = u(c_{i,j,L}) &+ \zeta_j v_{i,j+1,L} + \chi_j v_{i,j-1,L} \\ &+ z_{i,j,L} v_{i+1,j,L} + x_{i,j,L} v_{i-1,j,L} \\ &+ \Psi_L v_{i,j,L-1} \\ &+ (y_{i,j,L} + \Theta_L) v_{i,j,L} - \rho v_{i,j,L}. \end{aligned}$$

Finally, for $i = 1$ and $i = M$, we use the following conditions: Since we know that at $a = -a_{min}$, capital satisfies $k(-a_{min}, r, z) = 0$, and we want the consumption to be bounded. Thus, for $i = 1$ we set $x_{1,j,l} = 0$, which implies that $v_{0,j,l}$ does not play a role in the system. Given the complexity of the system and the relationship between v_{aa} and k, it is rather complicated to find a boundary condition for $i = M$. In order to prescribe a behavior for k at $i = M$, we consider the asymptotic behavior of k in the case of constant z and r, so that

$$k = \frac{r - r^f}{(\gamma + 1)\sigma^2} a$$

for large values of a. Hence

$$k_{M,j,l} = \min \left\{ \frac{r - r^f}{(\gamma + 1)\sigma^2} a_M, a_M + a_{min} \right\}.$$

This implies that

$$\begin{aligned} x_{Mjl} &= -\frac{s^-_{M,j,l}}{\mathrm{d}a_M} + \frac{\sigma^2 k^2_{Mjl}(\gamma + 1)}{2\,\mathrm{d}a_M\ \mathrm{d}a_{M+1}}, \\ z_{Mjl} &= \frac{s^+_{Mjl}}{\mathrm{d}a_{M+1}} \end{aligned}$$

Note that the boundary conditions at $i = j = l = 1$ comply with the ones in (1.24).
We want to solve $F_{ijl} = 0$ using an implicit scheme. We begin with a guess v^0 and compute the following magnitudes.
We begin with the forward and backward first derivative with respect to a, namely,

$$v^f_a \coloneqq \frac{v_{i+1jl} - v_{ijl}}{\mathrm{d}a_{i+1}}, \quad v^b_a \coloneqq \frac{v_{ijl} - v_{i-1jl}}{\mathrm{d}a_i}.$$

At the boundaries we use for $i = 1$

$$v^b_a \coloneqq u(z_l - a_{min} r^f)$$

and at $i = M$

$$v_a^f \coloneqq u(z_l + a_M r^f + k_{Mjl}(r_j - r^f))$$

and, with that, compute the forward and backward cost from the optimality condition

$$c^f \coloneqq \dot{u}^{-1}(v_a^f), \quad c^b \coloneqq \dot{u}^{-1}(v_a^b).$$

With that, we can define the forward and backward savings s^f and s^b by using c^f and c^b, respectively. Finally, we set

$$v_a \coloneqq \phi^f v_a^f + \phi^b v_a^b + \phi^0 \dot{u}(z_l + a_i r^f + k_{ijl}(r_j - r^f)),$$

where ϕ^f is 1 if $s^f > 0$ and 0 otherwise, ϕ^b is 1 if $s^b < 0$ and 0 otherwise and $\phi^0 = 1 - \phi^f - \phi^b$. Now, we can compute the optimal cost c as

$$c \coloneqq \dot{u}^{-1}(v_a),$$

and the savings using the definition. For capital, we use the following expression given by the optimality condition

$$k_{ijl} = \min \left\{ \frac{-(r_j - r^f)(v_a^f + v_a^b)\,\mathrm{d}a_{i+1}}{2\sigma^2(v_a^f - v_a^b)}, a_i + a_{min} \right\}.$$

Consider a column vectorization of $F_{ijl}, v_{ijl}, c_{ijl}, k_{ijl}$, so that

$$\mathcal{F} = \mathcal{U} + \mathcal{A}\mathcal{V} - \rho I_{MNL}\mathcal{V},$$

where

$$\mathcal{A} = \begin{bmatrix} A_1 + \Psi_1 I_{MN} & \Theta_1 I_{MN} & 0 & \cdots & 0 \\ \Psi_2 I_{MN} & A_2 & \Theta_2 I_{MN} & \ddots & \vdots \\ 0 & \Psi_3 I_{MN} & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & A_{L-1} & \Theta_{L-1} I_{MN} \\ 0 & \cdots & 0 & \Psi_L I_{MN} & A_L + \Theta_L I_{MN} \end{bmatrix} \in \mathbb{R}^{MNL \times MNL},$$

$$A_l = \begin{bmatrix} D_{1,l} + \chi_1 I_M & \zeta_1 I_M & 0 & \cdots & 0 \\ \chi_2 I_M & D_{2,l} & \zeta_2 I_M & \ddots & \vdots \\ 0 & \chi_3 I_M & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & D_{N-1,l} & \zeta_{N-1} I_M \\ 0 & \cdots & 0 & \chi_N I_M & D_{N,l} + \zeta_N I_M \end{bmatrix} \in \mathbb{R}^{MN \times MN},$$

$$D_{j,l} = \begin{bmatrix} y_{1,j,l} & z_{1,j,l} & 0 & \cdots & 0 \\ x_{2,j,l} & y_{2,j,l} & z_{2,1,l} & \ddots & \vdots \\ 0 & x_{3,1,l} & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & y_{M-1,j,l} & z_{M-1,j,l} \\ 0 & \cdots & 0 & x_{M,j,l} & y_{M,j,l} \end{bmatrix} \in \mathbb{R}^{M \times M}.$$

The column vector corresponding to v is then

$$\mathcal{V} = \begin{bmatrix} v_{1,1,1} \\ \vdots \\ v_{M,1,1} \\ \vdots \\ v_{1,N,1} \\ \vdots \\ v_{M,N,1} \\ \vdots \\ v_{1,1,L} \\ \vdots \\ v_{M,1,L} \\ \vdots \\ v_{1,N,L} \\ \vdots \\ v_{M,N,L} \end{bmatrix}$$

and similar to $\mathcal{U} = u(\mathcal{C})$. Let $\Delta > 0$ be given. We now solve the implicit problem

$$\frac{\mathcal{V}^{n+1} - \mathcal{V}^n}{\Delta} = \mathcal{U}^n + \mathcal{A}^n \mathcal{V}^{n+1} - \rho I_{MNL} \mathcal{V}^{n+1};$$

that is, beginning with $\mathcal{V}^0$, we find $\mathcal{V}^1$ using the former equation, find the new forward and backward variables, update $\mathcal{A}$ and $\mathcal{U}$, and repeat the process until $\|\mathcal{V}^{n+1} - \mathcal{V}^n\|_\infty < \varepsilon$ for some error $\varepsilon > 0$, which implies

$$\|\mathcal{F}^n\|_\infty = \|\mathcal{U}^n + \mathcal{A}^n \mathcal{V}^{n+1} - \rho I_{MNL} \mathcal{V}^{n+1}\|_\infty < \frac{\varepsilon}{\Delta}.$$

For the iteration in which the latter is achieved, we jump into finding the solution of the FP equation. Omitting the n where $\|\mathcal{V}^{n+1} - \mathcal{V}^n\|_\infty < \varepsilon$, we consider $c, k, \mathcal{A}$ at this iteration.

Now, using Appendix A.4, we have that there is a $\mathcal{P} \in \ker \mathcal{A}^\top$ which is in fact a discretization of the solution to the stationary FP equation. For the discretization of p in the mesh Ω_{NML}, we set p_{ijl} so that $\mathcal{P}$ is a column vectorization of p_{ijl}. Having obtained the density function, we can now adjust the parameters. In the end, we wish to find θ and w. To this end, we begin with K^0, θ^0, w^0 and choose a relaxation parameter $\omega \in (0, 1)$. In order to update K, compute with a numerical method the following quantity

$$S = \int_{\Omega_{NML}} p(a_i, r_j, z_l) k(a_i, r_j, z_l) \, \mathrm{d}(a_i, r_j, z_l) \, .$$

Now update K as

$$K = \omega K + (1 - \omega) S$$

and use (1.10d) to update the new parameters, namely,

$$\theta = \frac{1}{3}K^{-2/3} - \delta,$$
$$w = \frac{2}{3}K^{1/3}.$$

One needs to repeat the process of solving the HJB with these new K^1, θ^1, w^1, obtain the new stationary density of the FP equation, and again update K, θ, w. The stationary equilibrium is reached when the relative error between two consecutive iterations is smaller than a given $\varepsilon > 0$; that is,

$$\left|\frac{\theta^{m+1}}{\theta^m} - 1\right| + \left|\frac{w^{m+1}}{w^m} - 1\right| < \varepsilon.$$

Once we have concluded this process, we are able to find the income distribution $q(\iota)$. To this end, we compute k_z by using the forward derivative in z and check the condition

$$J_{ijl} \coloneqq 1 + (r_j - r^f)(k_z)_{ijl} \neq 0.$$

If this holds, we interpolate p, k, and k_z. Next, we define the income

$$\iota_{ijl} = z_l + a_i r^f + k_{ijl}(r_j - r^f)$$

and set $\iota_{min} \coloneqq \min\{\iota_{ijl}\}$ and $\iota_{max} \coloneqq \max\{\iota_{ijl}\}$. Next, we interpolate a, r, z, ι and p, k_z. For the sake of simplicity, we denote the new number of points by the same letters as before and assume that the partition in $\iota_n \in [\iota_{min}, \iota_{max}]$ has P points. We denote by q_{ijn} and J_{ijn} the points of

$$q(a_i, r_j, \iota_n) \coloneqq p(a_i, r_j, z(a_i, r_j, \iota_n))$$

and

$$J(a_i, r_j, \iota_n) \coloneqq 1 + (r_j - r^f)k_z(a_i, r_j, z(a_i, r_j, \iota_n)),$$

respectively. In order to find the former and the latter, we proceed as follows

for $n = 1, \ldots, P$:
  for $i = 1, \ldots, N$:
    for $j = 1, \ldots, N$:
      for $l = 1, \ldots, L$:
        if $|(z_l + a_i r^f + k_{ijl}(r_j - r^f))/\iota_n - 1| < \varepsilon$:
          $q_{ijn} = p_{ijl}$
          $J_{ijn} = 1 + (r_j - r^f)(k_z)_{ijl}$.

With that, we proceed by computing $q_n \coloneqq q(\iota_n)$ as in (1.11), so that

$$q_n = \int_{-a_{min}}^{a_{max}} \int_{r_{min}}^{r_{max}} \frac{q(a_i, r_j, \iota_n)}{J(a_i, r_j, \iota_n)} \, \mathrm{d}r_j \, \mathrm{d}a_i \, .$$

Bibliography Chapter 1

Achdou, Y., Han, J., Lasry, J.-M., Lions, P.-L., and Moll, B. (2022). Income and wealth distribution in macroeconomics: A continuous-time approach. *The Review of Economic Studies*, 89(1):45–86.

Adams, R. A. (1975). *Sobolev Spaces*. Number 65. New York, London.

Aiyagari, S. R. (1994). Uninsured idiosyncratic risk and aggregate saving. *The Quarterly Journal of Economics*, 109(3):659–684.

Angeletos, G.-M. (2007). Uninsured idiosyncratic investment risk and aggregate saving. *Review of Economic Dynamics*, 10(1):1–30.

Arnold, L. (1974). *Differential Equatinos: Theory and Applications*. Wiley-Interscience, New York London Sydney Toronto.

Aït-Sahalia, Y. (1996). Nonparametric pricing of interest rate derivative securities. *Econometrica*, 63(3):527–560.

Barles, G. and Souganidis, P. (1991). Convergence of approximation schemes for fully nonlinear second order equations. *Asymptotic Analysis*, (4):271–283.

Bayer, C. and Luetticke, R. (2018). Solving discrete time heterogeneous agent models with aggregate risk and many idiosyncratic states by perturbation. Working paper, CEPR Discussion Paper No. DP13071. Available at SSRN: https://ssrn.com/abstract=3222582.

Benhabib, J. and Bisin, A. (2018). Skewed wealth distributions: Theory and empirics. *Journal of Economic Literature*, 56(4):1261–91.

Benhabib, J., Bisin, A., and Zhu, S. (2011). The distribution of wealth and fiscal policy in economies with finitely lived agents. *Econometrica*, 79(1):123–157.

Benhabib, J., Bisin, A., and Zhu, S. (2015). The wealth distribution in bewley economies with capital income risk. *Journal of Economic Theory*, 159:489–515.

Benhabib, J., Cui, W., and Miao, J. (2024). Capital income jumps and wealth distribution. *Quantitative Economics*, 15(4):1197–1247.

Bewley, T. (1977). The permanent income hypothesis: A theoretical formulation. *Journal of Economic Theory*, 16:252–292.

Bogachev, V. I., Krylov, N. V., Röckner, M., and Shaposhnikov, S. V. (2015). *Fokker–Planck–Kolmogorov Equations*. American Mathematical Society.

Cagetti, M. and De Nardi, M. (2006). Entrepreneurship, frictions, and wealth. *Journal of political Economy*, 114(5):835–870.

Carroll, C. D. (1997). Buffer-stock saving and the life cycle/permanent income hypothesis. *The Quarterly journal of economics*, 112(1):1–55.

Castaneda, A., Diaz-Gimenez, J., and Rios-Rull, J.-V. (2003). Accounting for the us earnings and wealth inequality. *Journal of political economy*, 111(4):818–857.

Countryman, G. J. (1999). The effect of unemployment insurance benefits on income inequality in the canadian provinces. *Canadian Public Policy/Analyse de Politiques*, pages 539–556.

Courtadon, G. (1982). The pricing of options on default-free bonds. *Journal of Financial and Quantitative Analysis*, 17(1):75–100.

Cox, J. C., Ingersoll Jr, J. E., and Ross, S. A. (1985). An intertemporal general equilibrium model of asset prices. *Econometrica: Journal of the Econometric Society*, pages 363–384.

Daminato, C. and Pistaferri, L. (2024). Returns heterogeneity and consumption inequality over the life cycle. Technical report, National Bureau of Economic Research.

Du, Z. (2025). Solving heterogeneous agent models in continuous time with adaptive sparse grids. Working paper.

Duffie, D., Pan, J., and Singleton, K. (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68(6):1343–1376.

Engelbert, H.-J. and Peskir, G. (2014). Stochastic differential equations for sticky brownian motion. *Stochastics An International Journal of Probability and Stochastic Processes*, 86(6):993–1021.

Evans, L. (2010). *Partial Differential Equations*. 2nd edition.

Fischer, T. (2018). Modeling inequality and mobility with stochastic processes. *Available at SSRN 3090904*.

Gabaix, X. and Landier, A. (2008). Why has ceo pay increased so much? *The Quarterly Journal of Economics*, 123(1):49–100.

Garcke, J. and Ruttscheidt, S. (2025). Finite differences on sparse grids for continuous-time heterogeneous agent models. *Algorithms*, 18(1).

Garicano, L. and Rossi-Hansberg, E. (2006). Organization and inequality in a knowledge economy. *The Quarterly Journal of Economics*, 121(4):1383–1435.

Geerolf, F. (2016). A theory of pareto distributions. *UCLA manuscript*.

Goldberg, S. (1966). *Unbounded Linear Operators*. McGraw Hill.

Grisvard, P. (1985). *Elliptic problems in nonsmooth domains*. Pitman Advanced Publishing Program.

Han, J., Yang, Y., et al. (2021). Deepham: A global solution method for heterogeneous agent models with aggregate shocks. *arXiv preprint arXiv:2112.14377*.

Huang, J. (2023). Breaking the curse of dimensionality in heterogeneous-agent models: A deep learning-based probabilistic approach. *Available at SSRN 4649043*.

Huggett, M. (1993). The risk-free rate in heterogeneous-agent incomplete-insurance economies. *Journal of economic Dynamics and Control*, 17(5-6):953–969.

Kallianpur, G. (1980). *Stochastic Filtering Theory*, volume 13 of *Applications of Mathematics*. Springer, New York Heidelberg Berlin.

Kaplan, G. and Violante, G. L. (2022). The marginal propensity to consume in heterogeneous agent models. *Annual Review of Economics*, 14(1):747–775.

Khieu, H. and Wälde, K. (2018). Capital income risk and the dynamics of the wealth distribution. Working paper, Institute of Labor Economics (IZA), Bonn. IZA Discussion Papers, No. 11840.

Knäble, F. (2012). *Analysis of non-autonomous Ornstein-Uhlenbeck Processes with Càd-Làg paths*. PhD thesis, Universität Bielefeld.

Koeniger, W., Leonardi, M., and Nunziata, L. (2007). Labor market institutions and wage inequality. *ILR Review*, 60(3):340–356.

Krueger, D., Li, F., and Uhlig, H. (2024). Neoclassical growth transition dynamics with one-sided commitment. Technical report, National Bureau of Economic Research.

Krusell, P. and Smith, Jr, A. A. (1998). Income and wealth heterogeneity in the macroeconomy. *Journal of Political Economy*, 106(5):867–896.

Kuhn, M. and Rios-Rull, J.-V. (2016). 2013 update on the us earnings, income, and wealth distributional facts: A view from macroeconomics. *Quarterly Review, Federal Reserve Bank of Minneapolis, April*, pages 1–75.

Lasry, J.-M. and Lions, P.-L. (2007). Mean field games. *Japanese journal of mathematics*, 2(1):229–260.

Lions, P.-L. (1983). Optimal control of diffusions processes and hamilton-jacobi equations, part ii: Viscosity solutions and uniqueness. *Comm. Partial Differential Equations*, (8):1229–1276.

Lions, P.-L. and Souganidis, P. (1988). Viscosity solutions of second-order equations, stochastic control and stochastic differential games. In *Stochastic differential systems, stochastic control theory and applications*, pages 293–309. Springer.

Longstaff, F. A. and Schwartz, E. S. (1992). Interest rate volatility and the term structure: A two-factor general equilibrium model. *The Journal of Finance*, 47(4):1259–1282.

Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: The continuous-time case. *The review of Economics and Statistics*, pages 247–257.

Nickell, S. (1997). Unemployment and labor market rigidities: Europe versus north america. *Journal of Economic perspectives*, 11(3):55–74.

Ocampo, S. and Robinson, B. (2024). Computing longitudinal moments for heterogeneous agent models. *Computational Economics*, 64(3):1891–1912.

Øksendal, B. (2014). *Stochastic Differential Equations*. Springer-Verlag GmbH, sixth edition.

Parra-Alvarez, J. C., Posch, O., and Wang, M.-C. (2017). Identification and estimation of heterogeneous agent models: A likelihood approach. Working paper, Center for Economic Studies and Ifo Institute (CESifo), Munich. CESifo Working Paper, No. 6717.

Quadrini, V. (2000). Entrepreneurship, saving, and social mobility. *Review of economic dynamics*, 3(1):1–40.

Quadrini, V. (2009). Entrepreneurship in macroeconomics. *Annals of Finance*, 5(3-4):295–311.

Quadrini, V. and Rıos-Rull, J.-V. (1997). Dimensions of inequality: Facts on the us distribution of earnings, income and wealth. *Federal Reserve Bank of Minneapolis Quarterly Review*, 21(2):3–21.

Reiter, M. (2023). State reduction and second-order perturbations of heterogeneous agent models. Technical report, IHS Working Paper.

Risken, H. (1989). *The Fokker-Planck Equation*. Springer Verlag.

Rosen, S. (1981). The economics of superstars. *The American economic review*, 71(5):845–858.

Smith, M., Zidar, O. M., and Zwick, E. (2021). Top wealth in america: New estimates and implications for taxing the rich. Technical report, National Bureau of Economic Research.

Stroock, D. W. and Varadhan, S. (1997). *Multidimensional diffusion processes*. Springer.

Tervio, M. (2008). The difference that ceos make: An assignment model approach. *American Economic Review*, 98(3):642–68.

Tucsnak, M. and Weiss, G. (2009). *Observation and Control for Operator Semigroups*. Birkhäuser Verlag AG.

Vasicek, O. (1977). An equlibrium characterization of the term structure. *Journal of Financial Economics*, 5:177–188.

Wang, N. (2007). An equilibrium model of wealth distribution. *Journal of Monetary Economics*, 54(7):1882–1904.

Öttinger, H. C. (1996). *Stochastic Processes in Polymeric Fluids: Tools and Examples for Developing Simulation Algorithms*. Springer-Verlag Berlin Heidelberg.

Chapter 2

Sequence-Space Jacobians in Continuous Time

This paper describes a method for solving transition dynamics in heterogeneous-agent models with aggregate risk in continuous time. It extends the methods proposed in Achdou et al. (2022) and complements recent work by Auclert et al. (2021) on how to solve for the transition dynamics of heterogeneous-agent models in discrete time using sequence-space Jacobians. Our approach solves for the stationary equilibrium of the model without aggregate risk. Then, it invokes the Jacobian as a sufficient statistic, which describes the transition dynamics of the model with aggregate shocks. We apply our method to perform likelihood estimations of model parameters in a continuous-time framework.

2.1 Introduction

The field of macroeconomics has increasingly recognized the importance of diverse heterogeneity in dynamic general equilibrium models. However, the computation of transition dynamics and impulse responses following aggregate shocks remains a significant challenge. This paper presents a novel approach to this problem, demonstrating that the sequence-space Jacobian method proposed by Auclert et al. (2021) for discrete-time models can be effectively applied to continuous-time heterogeneous-agent models using the infinitesimal generator of the Hamilton-Jacobi-Bellman (HJB) equation to compute expectation vectors.

Building on the approach proposed by Reiter (2009), the sequence-space Jacobian method offers significant computational advantages. Most importantly, it allows for the efficient computation of transition dynamics, particularly in models with large state spaces, by solving the model in the sequence space rather than the state space. The approach requires solving for the steady state only

once, with subsequent transition dynamics efficiently computed in the sequence space, which rarely exceeds 400 time steps (Auclert et al., 2021).

This paper extends the sequence-space Jacobian method to a continuous-time setting, demonstrating its application in solving transition dynamics in a neoclassical heterogeneous-agent model in the spirit of Krusell and Smith (1998) and in a one-asset New Keynesian model in the spirit of McKay et al. (2016). The models can be solved when idiosyncratic shocks to labor productivity follow either a Poisson process or a diffusion process. The computational acceleration is substantial—at least 30 times faster than the previous approach of iterating over a guess of the transition path and updating it using market-clearing conditions, as used in Achdou et al. (2022) and Kaplan et al. (2018).

For discrete-time models, a substantial body of research has led to significant improvements in solution techniques since the early breakthrough of Krusell and Smith (1998). For example, the sequence-space Jacobians approach follows Reiter (2009) by linearizing regarding aggregates but preserving nonlinearities regarding idiosyncratic shocks. The Reiter method can be used to solve models that do not feature approximate aggregation and instead capture the rich interactions between the distribution of agents and macroeconomic outcomes, which are central to recent advances in modeling rich heterogeneity (e.g., Kaplan et al., 2018). However, the method involves solving a linear system that grows according to the dimension of the state space. For many complex models, decompositions such as Schur or singular value decomposition (SVD), which are required to solve these models, become too costly. This concern has led researchers to turn to "model reduction" methods, which involve approximating the equilibrium distribution and sometimes also the value function (Bayer and Luetticke, 2020; Winberry, 2018). Alternatively, Auclert et al. (2021) build on Boppart et al. (2018) and solve the transition in the sequence space, which decouples the computation from the rapidly growing state space.

Our contribution is to show how continuous-time approaches benefit from such developments and how the stochastic calculus avoids discretization of numerical expectations in discrete-time models. Predating our continuous-time approach, Ahn et al. (2018) propose one way of efficiently computing transition dynamics. They also suggest model reduction and, further, using automatic differentiation to obtain a sparse Jacobian. This adjustment helps with memory requirements and computational time for both the SVD of the Jacobian and solving the difference equation of the linearized model itself. In addition, they suggest perturbing the deviations of value and distribution functions from their stationary equilibrium counterparts instead of perturbing the functions themselves. This approach allows for different parametric classes for deviations and stationary equilibrium functions. As a result, it decouples the number of perturbed parameters from the number of parameters used for the function approximation in the stationary equilibrium (which can potentially be richer). Compared to their method, ours has the advantage of avoiding the calculation of a very large Jacobian because we work in the sequence space, which is much smaller than the state space for most models. Another advantage is that this allows us to calculate second-order (or higher) perturbations because the number of

(higher-order) derivatives to be calculated does not increase as fast as in their setting.

Complementary to our approach, Bilal and Goyal (2025) propose an analytic representation of sequence-space Jacobians for heterogeneous-agent models in continuous time. Their key innovation is to analytically differentiate individual policy functions with respect to price changes before any discretization, using a "pen-and-paper" perturbation. This process yields linear partial differential equations for the response of policy functions, which can then be solved efficiently. However, their "linearize first, discretize next" strategy necessitates the analytical derivation of these specific PDEs for each model configuration, which can become increasingly complex and require significant upfront analytical investment as model features evolve. In contrast, our method provides a robust framework by more directly extending the Auclert et al. (2021) methodology. By discretizing the model first and then utilizing the infinitesimal generator of the Hamilton-Jacobi-Bellman equation to compute the necessary expectation vectors, our approach offers a more standardized and potentially more adaptable numerical engine. This reliance on the generator, a common tool for steady-state solutions in continuous time, allows for consistent application across various model specifications and idiosyncratic shock processes without the need for repeated analytical derivations of policy function responses. Our method, therefore, emphasizes a numerically robust and generalizable pathway to constructing sequence-space Jacobians, achieving substantial computational speed-ups over traditional iterative techniques for solving transition dynamics.

The remainder of the paper is organized as follows: Section 2.2 defines the generic model we attempt to solve with our method. Section 2.3 reviews the sequence-space Jacobian approach by Auclert et al. (2021) in the first half and then proceeds with our main results, providing an efficient way to compute the Jacobian in continuous time. 2.4 shows how a model representation as a directed acyclical graph aids in transitioning from partial equilibrium Jacobians to general equilibrium Jacobians, which include all impulse responses after an aggregate shock. In Section 2.5, we present some numerical results and computational times of our method and compare the discrete and continuous-time settings. Section 2.6 concludes the paper. Example codes are provided on GitHub.[1]

2.2 Model environment

We illustrate our computational method by solving for the transition dynamics of the one-asset heterogeneous-agent New Keynesian (HANK) model in continuous time. In Appendix B.1, we show how one can solve the continuous-time version of the Krusell-Smith model presented in Ahn et al. (2018). Furthermore, Auclert et al. (2021) describe how one can use sequence-space Jacobians to solve a two-asset version of the HANK model, as well as various other model frameworks.

[1] See `https://github.com/reneglawion/SequenceSpaceJacobiansCT`.

2.2.1 A simple HANK model in continuous time

There is a continuum of agents $i \in [0,1]$ normalized to one who maximize their lifetime utility

$$\mathbb{E}_0 \left[\int_0^\infty e^{-\rho t} \frac{c^{1-\gamma}}{1-\gamma} \, \mathrm{d}t \right],$$

where ρ is the rate of time preference and γ is the coefficient of relative risk aversion. At each instant t, a household's idiosyncratic labor productivity is e_{it}, which follows an Ornstein-Uhlenbeck (Vasicek) process[2]

$$\mathrm{d}e_{it} = \kappa(\mu - e_{it}) \, \mathrm{d}t + \sigma \, \mathrm{d}W_t \,, \tag{2.1}$$

where $\mathrm{d}W_t$ is a Brownian motion. We assume sticky wages, and all workers work the same number of hours, i.e., $n_{it} = N_t$. Real pre-tax income is thus $y_{it} = W_t/P_t e_{it} n_{it}$, and real after-tax income is $z_{it} = (1-\tau) y_{it}$. Agents are subject to the budget constraint

$$\begin{aligned} c_{it} &\leq r_t \underline{a} + z_{it} \\ a_{it} &\geq \underline{a}, \end{aligned}$$

where $\underline{a}$ is an ad-hoc borrowing constraint (above the natural borrowing limit). Additionally, there exists a representative firm with an aggregate production function

$$Y_t = \exp(X_t) \, N_t,$$

where X_t is a TFP shock. Furthermore, assume flexible prices, i.e., $P_t = W_t/X_t$, which leads to $X_t = W_t/P_t$. Since the real wage is exogenous, the firm makes no profits. Goods inflation π_t equals wage inflation π_t^W minus TFP growth.

The government sets its fiscal policy, consisting of choosing the amount of government spending G_t, setting taxes $T_t = \tau_t Y_t$, and issuing the amount of government bonds B_t, subject to the government budget constraint $B_t = (1 + r_t) B_t + G_t - T_t$. For tractability, monetary policy follows a real rate rule $i_t = r + \pi_t + \varepsilon_t$.

2.2.2 Definition of equilibrium

Assume, for the moment, that there are no aggregate shocks. In equilibrium, the decisions made by households are determined by both individual and aggregate state variables. The individual state variables for a given household include capital holdings a and labor productivity z. Additionally, all markets clear, i.e.,

$$G_t + C_t = Y_t$$

[2]Note that one can also model the idiosyncratic shocks in logs as presented in Appendix F of Achdou et al. (2022) or as a Cox-Ingersoll-Ross process by multiplying the diffusion term by $\sqrt{e_{it}}$ to avoid negative levels of productivity. However, here, we hew as closely as possible to the discrete-time version in Auclert et al. (2021).

$$A_t = B_t.$$

We will later use the market-clearing conditions to compute the transition dynamics using sequence-space Jacobians. As shown in chapter 1, one can summarize the dynamics through a coupled system of the HJB equation and the Fokker-Planck equation (FPE)[3]

$$\rho v_t(a,z) = \partial_t v_t(a,z) + \max_{c\geq 0} u(c) + s_t(a,z)\partial_a v_t(a,z) + \lambda_t(z)\partial_z v_t(a,z) + \frac{\sigma^2}{2}\partial_{zz} v_t(a,z), \tag{2.2}$$

$$\partial_t g_t(a,z) = -\partial_a\left(s_t(a,z)g_t(a,z)\right) - \partial_z\left(\lambda_t(z)g_t(a,z)\right) + \frac{1}{2}\partial_{zz}\left(\sigma^2 g_t(a,z)\right), \tag{2.3}$$

$$B_t = \int a g_t(a,z)\,\mathrm{d}a\;\mathrm{d}z\,, \tag{2.4}$$

where we use the short-hand notation $\partial_a v = \frac{\partial a}{\partial v}$, s denotes the saving policy function $s \equiv z + r_t a - c$, and λ the drift of the diffusion process for labor income $\lambda \equiv \kappa(\mu - e)$. In continuous time, we denote the density by $g(a,z)$ and maintain it. In contrast, the discrete-time notation of Auclert et al. (2021) uses $D(a,z)$. In steady state, the system reduces to

$$\rho v(a,z) = \max_{c\geq 0} u(c) + s(a,z)\partial_a v(a,z) + \lambda(a,z)\partial_z v(a,z) + \frac{\sigma^2}{2}\partial_{zz}, \tag{2.5}$$

$$0 = -\partial_a\left(s(a,z)g(a,z)\right) - \partial_z\left(\lambda(z)g(a,z)\right) + \frac{1}{2}\partial_{zz}\left(\sigma^2 g(a,z)\right), \tag{2.6}$$

$$B = \int a g(a,z)\,\mathrm{d}a\;\mathrm{d}z\,. \tag{2.7}$$

To derive the key notion of expectation vectors later, we define the infinitesimal operator analogous to Theorem 1.3.1.

Definition 2.2.1 (Infinitesimal operator). Let A be the linear operator (*infinitesimal operator* or *generator of the semi-group*) associated with (2.5),

$$A(a,z)v(a,z) = s(a,z)\partial_a v(a,z) + \lambda(z)\partial_z v(a,z) + \frac{\sigma^2}{2}\partial_{zz} v(a,z), \quad v \in \mathcal{D}(\mathcal{A}).$$

Suppressing notational dependence on a and z using the infinitesimal operator, Theorem 1.3.1 shows that the solution to the stationary FPE (2.6) is given by the eigenfunction to the eigenvalue 0 of A^*, the adjoint operator of A; that is,

$$A^* g = 0.$$

Consequently, if $\mathcal{A}$ is a discretization of A, then the discretization of g, denoted by $\mathcal{G}$, is the eigenvector corresponding to the 0 eigenvalue of the matrix $\mathcal{A}^\top$. Using that notion, Achdou et al. (2022) show that if the finite-difference approximation is chosen correctly, i.e., one uses the upwind scheme, the discretized steady state is the solution to the following system of matrix equations:

$$\rho\mathcal{V} = u(\mathcal{V}) + \mathcal{A}\mathcal{V}$$

[3]The Fokker-Planck equation is also known as the Kolmogorov Forward equation.

$$0 = \mathcal{A}^\top \mathcal{G} \tag{2.8}$$
$$Y = F(\mathcal{G}),$$

where $\mathcal{V}$ and $u(\mathcal{V})$ are the discretized value and the maximized utility function over the grid, respectively. The transition matrix $\mathcal{A}^\top(a, z)$ is simply the transpose of the matrix from the discretized HJB equation because it encodes how households move around the individual state space. Finally, the third equation defines the prices/outcomes $h = (r, w)^\top$ as a function of aggregates (e.g., capital in the Aiyagari model) through the distribution $\mathcal{G}$.

To obtain the transition dynamics for the model, we now have to perform the following steps: First, solve for the steady-state equilibrium using any appropriate method without aggregate shocks. Here, we use the upwind scheme of Achdou et al. (2022). Next, shock the economy with any aggregate shock. In the one-asset HANK model, we use a government expenditure shock for illustration. In the Krusell-Smith model, we use a TFP shock. Finally, we use the sequence-space Jacobians to compute the path from the shocked steady state back to the original steady state. In what follows, we will describe the sequence-space Jacobian approach in detail.

2.2.3 Certainty equivalence

The sequence-space Jacobian solution method exhibits certainty equivalence regarding aggregate shocks, which is a typical feature of linearization techniques (Ahn et al., 2018). The standard deviation of aggregate TFP shock does not impact households' decision rules in the HJB. However, the solution method does not exhibit certainty equivalence regarding idiosyncratic shocks; thus, it captures the impact of aggregate uncertainty to the extent that aggregate shocks affect the distribution of idiosyncratic shocks.

Imposing certainty equivalence is not innocuous if we are interested in the effects of risk, particularly the effects of precautionary savings. Hence, the approach is less suitable for certain asset-pricing applications that require researchers to account for the direct effect of aggregate uncertainty on individual decisions. To address this issue, the continuous-time approach can be extended to first-order perturbation methods to allow decision rules to depend nonlinearly on relevant low-dimensional aggregate state variables. As Parra-Alvarez et al. (2021) show, a first-order perturbation in continuous time incorporates second-order terms to account for risk effects, which makes it advantageous to work in continuous time compared to discrete-time settings, where one would need second-order perturbation methods to capture risk effects. Alternatively, we may assume that individuals are averse to ambiguity, which would cause risk premia to survive linearization.

2.2.4 Incorporating Poisson shocks

To switch from a diffusion process driving the labor productivity process (2.1) to Poisson innovations, it is only necessary to alter the infinitesimal operator. Following Achdou et al. (2022, Proof of

Proposition 5), one can alternatively define the infinitesimal operator for the diffusion process as the generator of the idiosyncratic productivity process, i.e., $\mathcal{A}(a,z)\mathcal{V}(a,z) = \mu\partial_z v + \frac{1}{2}\sigma^2\partial_{zz}v$, where μ denotes the mean of the diffusion process and σ denotes the variance. With that, the model formulation (2.9) in steady state changes to

$$\begin{aligned} \rho v &= u(v) + (z + ra - c)\partial_a v + \mathcal{A}v \\ 0 &= -\partial_a(s(a,z)g) + \mathcal{A}^* g \\ Y &= F(g). \end{aligned} \tag{2.9}$$

To now incorporate a Poisson process with states $j = 1, 2$, one only need use the generator of the Poisson process $\mathcal{A}v = \lambda_j(v(a, z_{-j}) - v(a, z_j))$, where λ_j is the intensity of the Poisson process in state j, and the subscript $-j$ denotes $j = 2$ when $j = 1$ and vice versa.

2.3 Computing Jacobians in continuous time

In this section, we formulate the fake news algorithm of Auclert et al. (2021) for continuous-time models. To facilitate a direct comparison between their exposition and ours, we will adhere closely to their established notation and explanatory framework. To do so, we use the model formulation (2.9). In this general case, inputs are the aggregates relevant to the decision-making of individual agents, such as interest rates or wages, while outputs can include aggregate savings, consumption, investment, or other decisions by heterogeneous households or firms. Our main contribution is showing how to efficiently compute expectation vectors in continuous time.

2.3.1 General model representation

We start by formalizing the framework given in (2.9). Recall that the heterogeneous-agent model is a mapping between a time path of aggregate inputs $\boldsymbol{X}_t$ and a time path of outputs $\boldsymbol{Y}_t$. Assume that there are n_x inputs and n_y outputs and that the distribution is discretized on n_g grid points. Let $\boldsymbol{g}_t$ be the $n_g \times 1$ vector representing the distribution of agents at time t, and suppose that the aggregate outputs of interest are the average of individual outcomes against the distribution, i.e., $\boldsymbol{Y}_t = y_t^\top g_t$, with $\boldsymbol{y}_t$ denoting the $n_g \times n_y$ matrix of individual outcomes.[4] We assume that three functions—v, A, and y—exist, such that for a given initial distribution $\boldsymbol{g}_0$, $\boldsymbol{Y}_t$ is the solution to the system of matrix equations:

$$\rho\mathcal{V}_t = u(\mathcal{V}_t) + \mathcal{A}(\boldsymbol{X}_t)\mathcal{V}_t \tag{2.10}$$

$$\partial_t\mathcal{G}_t = \mathcal{A}(\boldsymbol{X}_t)^\top\mathcal{G}_t \tag{2.11}$$

$$\boldsymbol{Y}_t = y(\mathcal{V}_t, \boldsymbol{X}_t)^\top\mathcal{G}_t, \tag{2.12}$$

[4] Auclert et al. (2021) Appendix A shows how to extend the setting to address higher-order moments of outputs.

where $\boldsymbol{X}_t$ is the $n_x \times 1$ vector of aggregate inputs. The first equation is the approximated steady-state HJB equation for each point on the discretized grid, expressed in vector notation. The vector $u(\mathcal{V}_t)$ is the maximized utility function over the grid, and the matrix multiplication $A(\boldsymbol{X}_t)\mathcal{V}_t$ captures the remaining terms. The second equation is the discretized version of the steady-state FPE. The $n_g \times n_g$ transition matrix $\mathcal{A}(\boldsymbol{X}_t)^\top$ is simply the transposition of the matrix from the discretized HJB equation, which encodes how households move around the individual state space (Achdou et al., 2022).[5] Finally, the third equation defines the $n_y \times 1$ vector of aggregate outputs $\boldsymbol{Y}_t$, with the $n_g \times n_y$ matrix of individual outcomes $\boldsymbol{y}_t$ defined by $y(\mathcal{V}_t, \boldsymbol{X}_t)$. For any given $\boldsymbol{X}_{ss}$, the steady state of the model is the fixed point $(\boldsymbol{Y}_{ss}, \mathcal{V}_{ss}, \mathcal{G}_{ss})$ of (2.10)–(2.12) that is obtained when $\boldsymbol{X}_t = \boldsymbol{X}_{ss}$ at all times. For convenience, we write $A_{ss} \equiv A(\boldsymbol{X}_{ss})$ and $\boldsymbol{y}_{ss} = y(\mathcal{V}_{ss}, \boldsymbol{X}_{ss})$. We consider transitions of length T that end at this steady state, so that the terminal values are $\boldsymbol{X}_T = \boldsymbol{X}_{ss}$ and $\mathcal{V}_T = \mathcal{V}_{ss}$. The initial distribution $\mathcal{G}_0$ is given, and the algorithm of Auclert et al. (2021) later assumes that it is also equal to $\mathcal{G}_{ss}$. Hence, this setting allows us to study transitory shocks around a steady state.
Given this setup, we define a mapping from the $T \times n_x$ stacked vector of inputs $\boldsymbol{X}$ to the $T \times n_y$ stacked vector $\mathbf{Y}$ of outputs, which we write as

$$\boldsymbol{Y} = h(\boldsymbol{X}). \tag{2.13}$$

The goal is to characterize the Jacobian $\mathcal{J}$ of h evaluated at the steady state $\boldsymbol{X} = \boldsymbol{X}_{ss}$.

2.3.2 Structure of the Jacobian $\mathcal{J}$

Using Equation (2.12), define the output at time t in response to the shock at time s as

$$\boldsymbol{Y}_t^s = (\boldsymbol{y}_t^s)^\top \mathcal{G}_t^s. \tag{2.14}$$

Discretizing Equation (2.11) regarding time using an implicit scheme with step size Δt, we obtain

$$\mathcal{G}_{t+1} = (I - \Delta t \mathcal{A}_t^\top)^{-1} \mathcal{G}_t \tag{2.15}$$

which provides us with an update of the distribution in $t+1$ given the shock at time s as

$$\mathcal{G}_{t+1}^s = \Lambda_t^s \mathcal{G}_t^s \equiv (I - \Delta t \mathcal{A}_t^{*,s})^{-1} \mathcal{G}_t^s.$$

Auclert et al. (2021) first show how to efficiently obtain the policy functions $\boldsymbol{y}_t^s$ and transition matrices Λ_t^s. This approach makes use of the following implication of dynamic programming

Lemma 2.3.1 (Lemma 1 in Auclert et al. (2021)). *For any $t \geq 1, s \geq 1$, we have:*

$$\boldsymbol{y}_t^s = \begin{cases} \boldsymbol{y}_{ss} & \text{for } s < t \\ \boldsymbol{y}_{T-1-(s-t)}^{T-1} & \text{for } s \geq t \end{cases} \quad \text{and} \quad \Lambda_t^s = \begin{cases} \Lambda_{ss} & \text{for } s < t \\ \Lambda_{T-1-(s-t)}^{T-1} & \text{for } s \geq t. \end{cases} \tag{2.16}$$

[5] One can consider that to be the discretized law of motion.

Now define for the aggregate outcomes

$$\mathcal{F}_{t,s} = \mathrm{d}\boldsymbol{Y}_t^s - \mathrm{d}\boldsymbol{Y}_{t-1}^{s-1} \tag{2.17}$$

as the difference between the aggregate response of the output at t to a shock at date s and its response at $t-1$ to a shock at date $s-1$. Since the distribution in Equation (2.14) changes over time, $\mathcal{F}_{t,s}$ follows suit, and the following lemma describes $\mathcal{F}_{t,s}$ for a time discretization.

Lemma 2.3.2 (Lemma 2 in Auclert et al. (2021)). *Assume that $\mathcal{G}_0 = \mathcal{G}_{ss}$. For infinitesimal dx and any $s \geq 1, t \geq 1$, we have:*

$$\mathcal{F}_{t,s}\,\mathrm{d}x = \boldsymbol{y}_{ss}^\top \Lambda_{ss}^{t-1}\,\mathrm{d}\mathcal{G}_1^s\,. \tag{2.18}$$

The intuition for Equation (2.18) is as follows. From Lemma 2.3.1, we have that the path of the aggregate output $\boldsymbol{Y}_t$ at all dates $t = 0, \ldots T-1$ in response to a shock at date $s-1$ is the same as the path of $\boldsymbol{Y}_t$ in response to a shock at date s, from date $t = 1$ onward. Thus, the only difference is that the initial distribution in the second case is $\boldsymbol{g}_1^s$ rather than $\boldsymbol{g}_{ss}$. In terms of the first order, this difference in initial distribution affects aggregates at all dates as if agents followed their steady-state behavior, which is precisely what Equation (2.18) captures.

For a given s, $\mathcal{F}_{t,s}$ can be interpreted as the impulse response to a "date-s fake news shock," i.e., a shock to date s announced at date 0 and retracted at date 1. For that reason, Auclert et al. (2021) call their computational approach the fake news algorithm. Essentially, at date 0, agents react to the announcement, which leads to the distribution $\mathcal{G}_1^s$. After the announcement is retracted, they revert to their steady-state policies, so the effect on output at all dates $t \geq 1$ is $\boldsymbol{y}_{ss}\Lambda_{ss}^{t-1}\,\mathrm{d}\mathcal{G}_1^s$. This expression can usefully be rewritten with the notion of expectation vectors.

Definition 2.3.3 (Definition 1 in Auclert et al. (2021)). The expectation vector for outcome $\boldsymbol{Y}$ at time t is defined by

$$\mathcal{E}_t = \Lambda_{ss}^t \boldsymbol{y}_{ss}. \tag{2.19}$$

In the control theory literature, the matrix with rows $\mathcal{E}_0^\top, \mathcal{E}_1^\top, \ldots$ is called the *observability matrix* (Ahn et al., 2018). With Definition 2.3.3, Equation (2.18) can be rewritten as $\mathcal{F}_{t,s}dx = \mathcal{E}_{t-1}^\top\,\mathrm{d}\mathcal{G}_1^s$. Thus, for each grid point, the time path of $\mathcal{E}_t$ represents the expected time path of outcome $\boldsymbol{Y}$ in the steady state for an agent starting at that grid point. Following Auclert et al. (2021), one can use Lemmas 2.3.1 and 2.3.2 to arrive at the following proposition.

Proposition 2.3.4 (Proposition 1 in Auclert et al. (2021)). *Assume that $\mathcal{G}_0 = \mathcal{G}_{ss}$. For infinitesimal $\mathrm{d}x$, define the (t,s)-th element of the fake news matrix $\mathcal{F}$ as*

$$\mathcal{F}_{t,s}dx \equiv \begin{cases} \mathrm{d}\boldsymbol{Y}_0^s & \text{for } t = 0 \\ \mathcal{E}_{t-1}^\top\,\mathrm{d}\mathcal{G}_1^s & \text{for } t \geq 1, \end{cases} \tag{2.20}$$

where $\mathrm{d}\boldsymbol{Y}_0^s = \mathrm{d}\boldsymbol{y}_0^s \mathcal{G}_{ss}$ *and* $\mathrm{d}\mathcal{G}_1^s = \mathrm{d}\Lambda_0^s \mathcal{G}_{ss}$*. Then, the Jacobian* $\mathcal{J}$ *of* h *satisfies the recursion* $\mathcal{J}_{t,s} = \mathcal{J}_{t-1,s-1} + \mathcal{F}_{t,s}$ *for* $t, s \geq 1$*, with* $\mathcal{J}_{t,s} = \mathcal{F}_{t,s}$ *for* $t = 0$ *or* $s = 0$*, and therefore, is given by*

$$\mathcal{J}_{t,s} = \sum_{k=0}^{\min\{t,s\}} \mathcal{F}_{t-k,s-k}. \tag{2.21}$$

Having recapitulated the method developed by Auclert et al. (2021), the missing piece for the continuous-time framework is how to efficiently compute expectation vectors. In the next section, we will show how to do this for the one-asset HANK model with labor productivity following a diffusion process. In Appendix B.1, we will show that the approach is applicable to almost all reasonable continuous-time models and provide an implementation for the Krusell-Smith model, where labor productivity follows a Poisson process.

2.3.3 Efficiently computing expectation vectors

Suppose that we have some function $x(a, z)$. If the state of an individual household is (a_t, z_t) today, what do we expect (a_s, z_s) to be, for $s > t$, assuming that the household follows its steady-state policy rule? This mapping is called the expectation vector $\mathcal{E}(a, z)$.

Proposition 2.3.5 (Expectation vector in continuous time). *Let* $0 \leq s < t \leq T$*, define* $\mathrm{d}t = t - s$*, let* $x(a_t, z_t)$ *be some function of the state at time* s*, and assume that agents follow their steady-state policy rules. Then, the* t*th expectation vector* $\mathcal{E}_t = \mathbb{E}[x(a_t, z_t)|(a_s, z_s)]$ *is given by*

$$\mathcal{E}_t(a_s, z_s) = x(a_s, z_s) + \mathcal{A}(x(a_s, z_s))x(a_s, z_s)\,\mathrm{d}t\,. \tag{2.22}$$

Proof. Start with the definition of the expectation vector

$$\begin{aligned}\mathcal{E}_t(a_s, z_s) &= \mathbb{E}[x(a_t, z_t)|(a_s, z_s)] \\ &= x_s(a_s, z_s) + \frac{\mathbb{E}[\,\mathrm{d}x\,(a_s, z_s)]}{\mathrm{d}t},\end{aligned}$$

where the latter term is the expected drift over the (short) time interval $\mathrm{d}t$. Since we use a diffusion process, it follows from Itô's lemma that

$$\begin{aligned}\mathrm{d}x \;=\; &\left(\partial_a x(a_s, z_s)(z_s + ra_s - c_s) + \partial_z x(a_s, z_s)\kappa(\mu - e) + \partial_{zz} x(a_s, z_s)\frac{\sigma^2}{2}\right)\mathrm{d}t \\ &+ \partial_z x(a_s, z_s)\sigma(z_s)\,\mathrm{d}W_s\,.\end{aligned}$$

Therefore, using that the expectation of the increment of a standard Brownian motion is zero, $\mathbb{E}[\,\mathrm{d}W_s\,] = 0$,

$$\mathbb{E}[\,\mathrm{d}x\,(a_s, z_s)] = \left(\partial_a x(a_s, z_s)(z_s + ra_s - c_s) + \partial_z x(a_s, z_s)\kappa(\mu - e) + \partial_{zz} x(a_s, z_s)\frac{\sigma^2}{2}\right)\mathrm{d}t\,.$$

Thus, we obtain as the expectation vector over a time interval $\mathrm{d}t$

$$\mathcal{E}_t(a, z) = x_s(a_s, z_s) + \left(\partial_a x(a_s, z_s)(z_s + ra_s - c_s) + \partial_z x(a_s, z_s)\kappa(\mu - e) + \partial_{zz} x(a_s, z_s)\frac{\sigma^2}{2}\right)\mathrm{d}t\,.$$

Using the notation of the generator $\mathcal{A}$ from Definition 1.3.1 yields

$$\mathcal{E}_t(a,z) = x_s(a_s,z_s) + \mathcal{A}\left(x(a_s,z_s)\right) x(a_s,z_s)\,\mathrm{d}t\,. \qquad \square$$

The result also holds for processes different from diffusions. One essentially only needs that $\lim_{t\searrow 0} \frac{\mathbb{E}_x[h(x(t)]-h(x)}{t} = \mathcal{A}$, for a function h. In Appendix B.1.6, we provide conditions under which this holds and apply the method for a Krusell-Smith type of model, where labor productivity follows a continuous-time Poisson process. To increase computational efficiency, we make use of the following result.

Corollary 2.3.6 (Compute expectation vectors via recursion). *To compute a numerical solution, discretize the time domain $[0,T]$ using $N+1$ evenly spaced points $t_j, j = 0,\dots,N$,*

$$t_0 = 0, t_1 = t_0 + \Delta t, t_2 = t_0 + 2\Delta t, \dots, t_N = t_0 + N\Delta t = T,$$

where $\Delta t = \frac{T}{N}$. Then, the t_jth expectation vector can be computed via the recursion

$$\mathcal{E}_{t_j}(a_{t_{j-1}}, z_{t_{j-1}}) = \mathcal{E}_{t_{j-1}}(a_{t_{j-1}}, z_{t_{j-1}}) + \mathcal{A}\left(x(a_{t_{j-1}}, z_{t_{j-1}})\right) x(a_{t_{j-1}}, z_{t_{j-1}})\,\mathrm{d}t\,,$$

starting with $\mathcal{E}_{t_0}(a_{t_0}, z_{t_0}) = x(a_0, z_0)$.

Proof. The proof follows directly from the law of iterated expectations. $\square$

Therefore, as in the discrete-time version, it follows that as long as we know how to evaluate one-period-ahead expectations, we can calculate expectation vectors at any horizon. In the numerical implementation, we can evaluate $\mathcal{A}\left(x(a_t,z_t)\right)$ using the discretized version of the generator $\mathcal{A}$ in Definition 1.3.1 by computing the matrix product $A \cdot x$.

2.4 The model as a DAG

To move from partial equilibrium Jacobians to general equilibrium transitions, one can represent the models in the form of directed acyclical graphs (DAGs).[6] Here, the model is represented as a combination of blocks, where a block is mapped from sequences of inputs to sequences of outputs.

2.4.1 Fiscal policy DAG for the one-asset HANK model

As an example, to analyze transition paths after a fiscal policy shock in the one-asset HANK model, one can start with the household block, which maps from the interest rate r and after-tax income Z to the aggregate consumption C and aggregate asset holdings A, and the fiscal policy block, which maps interest rates r, the number of outstanding bonds B, government expenditures G, and output Y to taxes T and after tax-income Z. Finally, there is a goods market clearing block, which takes

[6] See Appendix C.2 in Auclert et al. (2021).

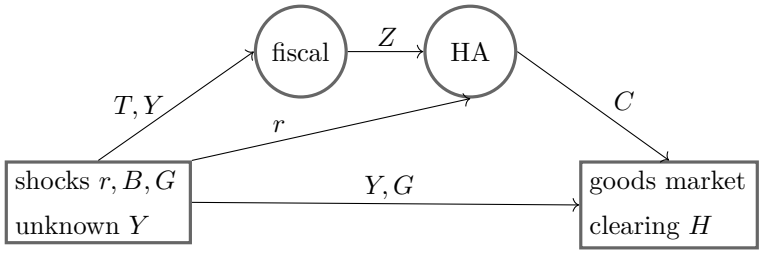


Figure 2.1: **DAG for the one-asset HANK model**
The graph shows one possible DAG of the one-asset HANK model for the analysis of a fiscal policy shock. The model has three possible shocks ($\mathbf{X}$) and one unknown ($\mathbf{Y}$). The illustration is based on Auclert et al. (2021).

Y, C, and G as inputs and maps to market clearing $H \equiv C + G - Y$. One can also consider the other blocks and examine the asset market clearing condition. Furthermore, note that the DAG is not unique. Figure 2.1 illustrates the idea of representing the model as a DAG. We only show the part of the model that we will need to analyze a government expenditure shock in what follows.

2.4.2 Taking derivatives along the DAG

With the notion of a DAG, one can easily analyze a government spending shock $\mathrm{d}G$.[7] We are interested in computing how Y is affected by a change in government spending. One needs to find Y such that

$$H(Y, G) = 0. \tag{2.23}$$

For a first-order shock $\mathrm{d}G$, one can use the implicit function theorem to obtain

$$\mathrm{d}Y = -(H_Y)^{-1} H_G \, \mathrm{d}G \,. \tag{2.24}$$

Thus, all we need are the Jacobians H_Y and H_G. Using the chain rule, we have

$$\begin{aligned} H_Y &= \underbrace{J^{H,Y}}_{=-I} + \underbrace{\mathcal{J}^{H,C}}_{=I} \cdot \underbrace{\mathcal{J}^{C,Z}}_{=\text{Fake News algorithm}} \cdot \underbrace{\mathcal{J}^{Z,Y}}_{=I}, \\ H_G &= \underbrace{\mathcal{J}^{H,G}}_{=I}, \end{aligned}$$

where I is a $T \times T$ identity matrix. Hence, the only difficult part is to compute $J^{C,Z}$ using the fake news algorithm, which we described earlier, since we can analytically compute all the other Jacobians.

[7] The idea is similar to learning via backward propagation in the neural networks literature.

Table 2.1: **Calibration of the one-asset HANK model**
We fix most of the parameters according to Auclert et al. (2021). In both models, the time preference rate and the amount of government debt are used to target asset market clearing and a weighted MPC of .25. Due to Walras's law, the goods market also clears.

Parameter	Description	Discrete time	Continuous time
β	Time preference rate	.7958	
ρ	Time preference rate		.0806
γ	Risk aversion	2	2
μ	Mean of idiosyncratic productivity shock	1	1
ρ_e	Persistence of idiosyncratic productivity shocks	.9	
κ	Speed of mean reversion		-log(.9)
σ	Standard deviation of idiosyncratic productivity shocks	.92	.92
G	Government spending	.2	.2
B	Government debt	.7872	.4938
Y	Output	1.0	1.0
r	interest rate	.03	.03
n_a	Number of asset grid points	200	200
n_e	Number of productivity grid points	10	10

2.5 Computing the transition path after a shock to government expenditures

In this section, we demonstrate the approach by computing impulse responses to AR(1) government expenditure shocks. We first calibrate the continuous-time version of the model from Section 2.2 to match the discrete-time version in Auclert et al. (2021). Next, we present and discuss the results of the analysis in detail by modeling a balanced budget shock and a deficit-financed shock using the DAG representation from Section 2.4.

2.5.1 Calibration procedure

We calibrate the continuous-time version to match the discrete-time version as closely as possible to enable comparison. For some parameters, this matching is not possible on a one-to-one basis; for example, the discrete-time discount rate β differs from the continuous-time rate ρ. However, these differences are minimal. The complete discrete-time version of the model is summarized in Appendix B.2.

Furthermore, in both versions, some parameters are calibrated internally to match market-clearing

conditions or MPCs. For example, we use the time preference rates to achieve market clearing and government debt to match a weighted MPC of .25.

Additionally, we normalize the income processes in both cases to a mean of one. In the discrete-time version, we use the Rouwenhorst (1995) method, so we sample on a grid of 10 points, and in continuous time, we use the finite-difference method of Achdou et al. (2022) to convert the Ornstein–Uhlenbeck process into a discrete-state Poisson process. In both cases, we use uniform grids and set the minimum level of assets to $\underline{a} = 0$ and the upper bound to $\overline{a} = 100$.

Moreover, we focus on matching the outcomes of both models for pedagogical reasons and not on matching any observed data. However, while we have specified and parameterized the continuous-time and discrete-time versions of the one-asset HANK models to be as comparable as possible, these are still different models. For example, while the OU process (2.1) is the continuous-time analogue of the discrete-time AR(1) process and we parameterize the two processes in an analogous fashion, these are still different stochastic processes: one process moves continuously over time, whereas the other only moves at discrete-time intervals. For the interested reader, we refer to Appendix F.1.1 in Achdou et al. (2022) for a thorough discussion and comparisons between the computation of steady states in a discrete-time and continuous-time version of a Hugget-style model. We present the full calibration in Table 2.1.

2.5.2 Resulting transition paths

We compute two kinds of government expenditure shocks to illustrate the method. The first is a balanced budget spending shock, where the spending increase is defined as an increase in both government spending and taxes by the same amount. The second is a deficit-financed expenditure shock, where only government spending is increased and taxes remain unchanged, i.e., $\mathrm{d}G = \mathrm{d}Y$.[8] In both versions, $\mathrm{d}G$ is a 1% AR(1) shock with persistency $\rho_G = 0.8$, but note that once the general Jacobian has been computed, one can compute all kinds of persistencies by a simple matrix multiplication.

Figure 2.2 plots the comparison between the discrete-time version of the model and the continuous-time version of the model in the case of a balanced budget government spending shock for the first 50 quarters. The discrete-time solution was computed with the SSJ toolkit,[9] whereas the continuous-time version was computed using the approach from this paper. Both models lead to similar results. Due to the discrete-time nature of the model, the market-clearing condition fluctuates considerably more than in the continuous-time version. However, at a scale of 10^{-16}, this is merely machine precision. Overall, we confirm the Haavelmo (1945) result and obtain $\mathrm{d}G = \mathrm{d}Y$ for all periods.

More interesting is the case of a deficit-financed government spending shock. We compare the results

[8] Details of the discrete-time implementation can be found on the SSJ GitHub repository: `https://github.com/shade-econ/nber-workshop-2022/blob/main/Tutorials/Tutorial%201%20Fiscal%20Policy.ipynb`.

[9] See `https://github.com/shade-econ/sequence-jacobian`.

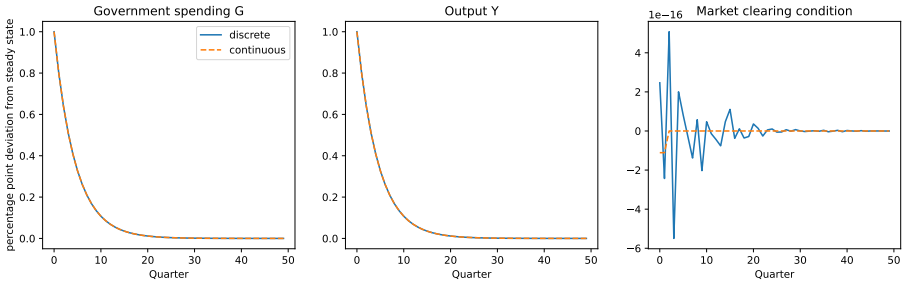


Figure 2.2: **Comparison of the discrete-time and continuous-time transition path**
The figure displays the discrete-time transition path of the one-asset HANK model after a government spending shock taken from Auclert et al. (2021) (blue, solid) and compares it with the continuous-time transition path (orange, dashed). Both models were calibrated using Table 2.1, and in both cases, the 1% balanced budget government spending shock follows an AR(1)-process with persistency $\rho_G = 0.8$.

for the continuous-time implementation in Figure 2.3. We see that both government spending shocks stimulate output, but the deficit-financed shock generates an increase of $\mathrm{d}Y > \mathrm{d}G$, since taxes are not increased, which also leads to the observed increase in the government deficit. The complete responses were computed using the DAG approach from Section 2.4, where we computed the Jacobian of the heterogeneous-agent block with the fake news algorithm. The complete code is hosted on GitHub. In terms of computational performance, we juxtapose our method with that presented in Achdou et al. (2022) and elaborated in Appendix B.3. The initial step in both scenarios entails calculating the steady state, which takes approximately .2 seconds. Subsequently, we proceed to compute the transition dynamics for deficit-financed government spending. This task demands 14.44 seconds when employing the technique of Achdou et al. (2022), while the sequence-space Jacobian method requires merely .0094 seconds. This results in a computational speed increase of a factor of approximately 1,500. Moreover, in our approach, the computation of additional impulses is reduced to a matrix multiplication operation, which is executable in a fraction of a millisecond on modern computing hardware. The rapid computation times per iteration even permit parameter estimation for the model, akin to the discrete-time framework of Auclert et al. (2021). Appendix B.1.4 contains the methodology for estimating the variance of the persistent component of a continuous-time Krusell-Smith model, and the requisite code is available in the associated GitHub repository.

2.6 Conclusion

In this paper, we present a new approach for computing heterogeneous-agent models in continuous time. The approach is based on the use of sequence-space Jacobians as sufficient statistics for summarizing the heterogeneity of agents to determine general equilibrium dynamics. This approach was

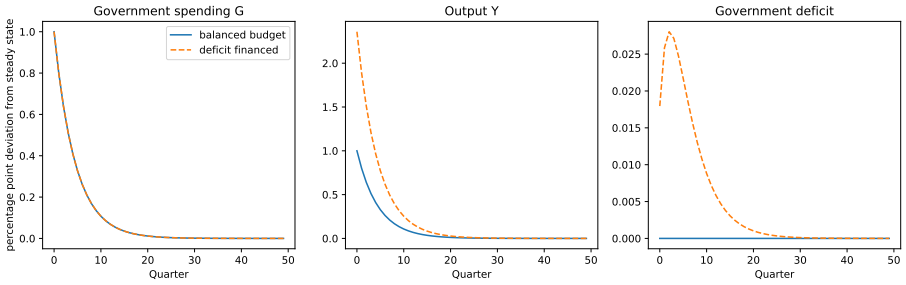


Figure 2.3: **Transition paths of balanced budget and deficit-financed government spending shocks**

Comparison between balanced budget (blue, solid) and deficit-financed (orange, dashed) government spending shocks. The left panel shows the 1% government spending shock, which follows an AR(1)-process with persistency $\rho_G = 0.8$. The panel in the middle shows the different effects on output, and the right panel displays the government deficit.

originally developed by Auclert et al. (2021) and has been widely used in the literature to study the effects of aggregate shocks on discrete-time heterogeneous-agent economies. Our modified approach builds upon the original by allowing for efficient and accurate computation in continuous-time settings. We demonstrate the effectiveness of our approach by computing the response of a fiscal policy shock in a continuous-time setting. Our results show that our modified approach is a valuable tool for studying the effects of policy shocks on heterogeneous-agent economies.

B.1 Krusell-Smith model in continuous time

We present the continuous-time version of the Krusell-Smith model, based on the version in Ahn et al. (2018).

B.1.1 Model environment

There is a continuum of agents $i \in [0, 1]$ normalized to one who maximize lifetime utility

$$\mathbb{E}_0 \left[\int_0^\infty e^{-\rho t} \frac{c^{1-\gamma}}{1-\gamma} \, \mathrm{d}t \right],$$

where ρ is the rate of time preference and γ is the coefficient of relative risk aversion. At each instant t, a household's idiosyncratic labor productivity is $z_{it} \in \{z_L, z_H\}$ with $z_L < z_H$. Households switch between the two states according to a Poisson process with arrival rates λ_L and λ_H. The aggregate supply of efficiency units of labor is exogenous and constant and denoted by $\bar{N} = \int_0^1 z_{jt} \, \mathrm{d}j$. A household with labor productivity z_{it} earns labor income $w_t z_{it}$. However, markets are incomplete, and households can only save in government bonds. They are subject to the borrowing constraint

$a_{it} \geq 0$, where a denotes their wealth, with which they can only trade in productive capital.
There is a representative firm with a Cobb-Douglas production function

$$Y_t = e^{Z_t} K_t^{\alpha} \bar{N}_t^{1-\alpha},$$

where Z_t is (the logarithm of) aggregate productivity, K_t is aggregate capital, and $\bar{N}_t$ is aggregate labor, which we fix to 1 without loss of generality. The logarithm of aggregate productivity follows the Ornstein-Uhlenbeck process:

$$\mathrm{d}Z_t = -\eta Z_t \,\mathrm{d}t + \sigma \,\mathrm{d}W_t \,,$$

where $\mathrm{d}W_t$ is a standard Brownian motion, η is the rate of mean reversion, and σ denotes the variance.

B.1.2 Equilibrium

Assume, for the moment, that there are no aggregate shocks, i.e., $Z_t = 0$, and solve for the steady state. In equilibrium, the decisions made by households are determined by both individual and aggregate state variables. The individual state variables for a given household include capital holdings a and labor productivity z. The aggregate state variables, which apply to all households, consist of aggregate productivity Z_t and the distribution of households across individual state variables $g_t(a, z)$.

$$\rho v_t(a,z) = \frac{\partial v_t}{\partial t} + \max_{c \geq 0} u(c) + \frac{\partial v_t}{\partial a}(w_t z_t + r_t a - c) + \lambda_z(v_t, (a, z_{-j}) - v_t(a,z)), \quad a \geq 0 \tag{2.25}$$

$$0 = -\frac{\partial}{\partial a}\left((w_t z_t + r_t a - c) g(a,z)\right) - \lambda_z g(a,z) - \lambda_{z_{-j}} g(a, z_{-j}), \tag{2.26}$$

$$K_t = \int a g_t(a,z) \,\mathrm{d}a \;\mathrm{d}z \,. \tag{2.27}$$

Again, we can rewrite the stationary system as

$$\begin{aligned} \rho v(a,z) &= u(v) + \mathcal{A} v \\ 0 &= \mathcal{A}^* g \\ Y &= F(g), \end{aligned}$$

where $u(v)$ is the maximized utility function over the grid, and the matrix multiplication $\mathcal{A}(v)$ captures the remaining terms in Equation (2.25). The transition matrix $\mathcal{A}^*(v)$ is simply the transposition of the matrix from the discretized HJB equation, which encodes how households move around the individual state space, i.e., $(Av)(a, z_j) = \lambda_j(v(a, z_{-j}) - v(a, z_j))$.

B.1.3 Linearized dynamics using Jacobians

Using a DAG representation, one can reduce the model to a single equation to find the t, s entry of H_K:

$$H_t(K, Z) \equiv \mathcal{K}_t\left(\left\{\alpha Z_s K_s^{\alpha-1} - \delta, (1-\alpha) Z_s K_s^{\alpha}\right\}\right) - K_t = 0,$$

which has to hold for all K and Z. This is the composition of the household capital supply function $\mathcal{K}$ with the interest rate $r(K_t, Z_t) = \alpha Z_t K_t^{\alpha-1} - \delta$ and wage $w(K_t, Z_t) = (1-\alpha) Z_t K_t^{\alpha}$ functions. Using the implicit function theorem yields the impulse response for $\mathrm{d}K$ after a $\mathrm{d}Z$ shock

$$\mathrm{d}K = -H_K^{-1} H_Z \, \mathrm{d}Z \, .$$

We can use the chain rule to obtain the Jacobians and use the fake news algorithm to find the derivative of the heterogeneous-agent block:

$$\begin{aligned} H_K &= \mathcal{J}^{A,r} \cdot \mathcal{J}^{r,K} + \mathcal{J}^{A,w} \cdot \mathcal{J}^{w,K} - I_{T \times T}, \\ H_Z &= \mathcal{J}^{A,r} \cdot \mathcal{J}^{r,Z} + \mathcal{J}^{A,w} \cdot \mathcal{J}^{w,Z}. \end{aligned}$$

Once we have the general Jacobian matrix $G_K \equiv -H_K^{-1} H_Z$, we can also compute all other impulse responses to the variables r, w, and Y by decomposing the responses into direct effects due to the shock $\mathrm{d}Z$ and an indirect effect from the induced change in $\mathrm{d}K$. We apply the chain rule to obtain the matrices G_r, G_w, and G_Y mapping from any shock $\mathrm{d}Z$ to the resulting impulse responses $\mathrm{d}r$, $\mathrm{d}w$, $\mathrm{d}Y$:

$$\begin{aligned} G_r &= \mathcal{J}^{r,Z} + \mathcal{J}^{r,K} \cdot G_K, \\ G_w &= \mathcal{J}^{w,Z} + \mathcal{J}^{w,K} \cdot G_K, \\ G_Y &= \mathcal{J}^{Y,Z} + \mathcal{J}^{Y,K} \cdot G_K, \end{aligned}$$

where each G characterizes the full linearized dynamics of aggregate variables K, r, w, and Y in response to any perturbation of Z. If we want the linearized dynamics of other variables, e.g., consumption, all we need is to take the Jacobian from the block in which these are calculated and then compose it with the G of its inputs. Consumption was already reported in the Jacobian, which we obtained via the fake news algorithm, so we simply need to compute

$$G_C = \mathcal{J}^{C,r} \cdot G_r + \mathcal{J}^{C,w} \cdot G_w.$$

B.1.4 Estimation

From the impulse responses, one obtains the second moments. With that, it is also possible to conduct likelihood estimation, as Auclert et al. (2021) show. Here, we briefly demonstrate that their approach also carries over to the discretized version of the continuous-time Krusell-Smith model.

The linear impulse responses can be interpreted as the $MA(\infty)$ representation of the model. Here, all we need are the second moments of the shocks to fully characterize the second moments of endogenous outcomes. This computation can be done in a highly efficient way, without any need for simulation. Let us decompose the TFP shock into a persistent and a transitory component:

$$\mathrm{d}Z_t \, = \, \mathrm{d}Z_t^1 \, + \, \mathrm{d}Z_t^2$$

$$\begin{aligned} \mathrm{d}Z_t^1 &= \eta\, \mathrm{d}Z_{t-1}^1 + \varepsilon_t^1 \\ \mathrm{d}Z_t^2 &= \varepsilon_t^2, \end{aligned}$$

where ε^1 and ε^2 are i.i.d. normal with mean zero and variances σ_1^1 and σ_2^2, i.e., the Brownian motions before discretization w.l.o.g., since the sum of two Brownian motions can be represented as one Brownian motion in distribution.

Now, define $m^{x,1}$ and $m^{x,2}$ as the (truncated) impulse responses of the variable X to these two structural shocks. As Boppart et al. (2018) note, the linearized impulse responses reduce to a $MA(T-1)$ representation:

$$X_t = \sum_{z=1}^{2} \sum_{s=0}^{T-1} m_s^{x,z} e_{t-s}^z.$$

From that, it follows that the covariances between two outcomes Y and X take the form

$$\mathrm{Cov}(Y_t, X_{t+l}) = \sum_{z=1}^{2} \sigma_z^2 \sum_{s=0}^{T-1} m_s^{x,z} m_{s+l}^{y,z}.$$

Since we work with Brownian motions (and their discretized versions), we know that, first, the means of all series are zero, and second, shocks are multivariate normal.[10] Thus, all we need to evaluate the log-likelihood are the second moments and the data. Specifically, if our data observations are stacked in a vector $\boldsymbol{y}$ and the covariances at all lags are stacked in a matrix $\boldsymbol{V}$, then the log-likelihood is (up to a constant):

$$\mathcal{L} = -\frac{1}{2}\log(\det(\boldsymbol{V})) - \frac{1}{2}\boldsymbol{y}^\top \boldsymbol{V}^{-1}\boldsymbol{y}.$$

We use that to estimate the parameter σ_{persist}, which measures the persistence of the TFP shock. For demonstration purposes, we simulate 100 data points with $\sigma_{\text{persist}} = .1$ and estimate the parameter via likelihood estimation.

B.1.5 Numerical results

Figure 2.4 shows the impulse responses for output Y, aggregate consumption C, and the real interest rate r for various persistencies of the government spending shock ρ. Computing additional impulse responses for varying persistencies in our approach is equivalent to a matrix multiplication. In fact, it is sufficient to perform one large matrix multiplication to simultaneously compute all impulse responses after obtaining the general Jacobians from Section B.1.3. The computation time is thus just .42 seconds to compute the first impulse response and .102 milliseconds for the additional five impulse responses for different persistencies.

Next, Figure 2.5 shows the results of the estimation. The mode of the log-likelihood function is near the value of σ_{persist} with which the data were simulated, although not exactly at .1, since we are simulating a finite sample with only 100 periods. Since our formulation closely follows the method

[10] That can also be imposed through some kind of preliminary demeaning or detrending procedure on the data.

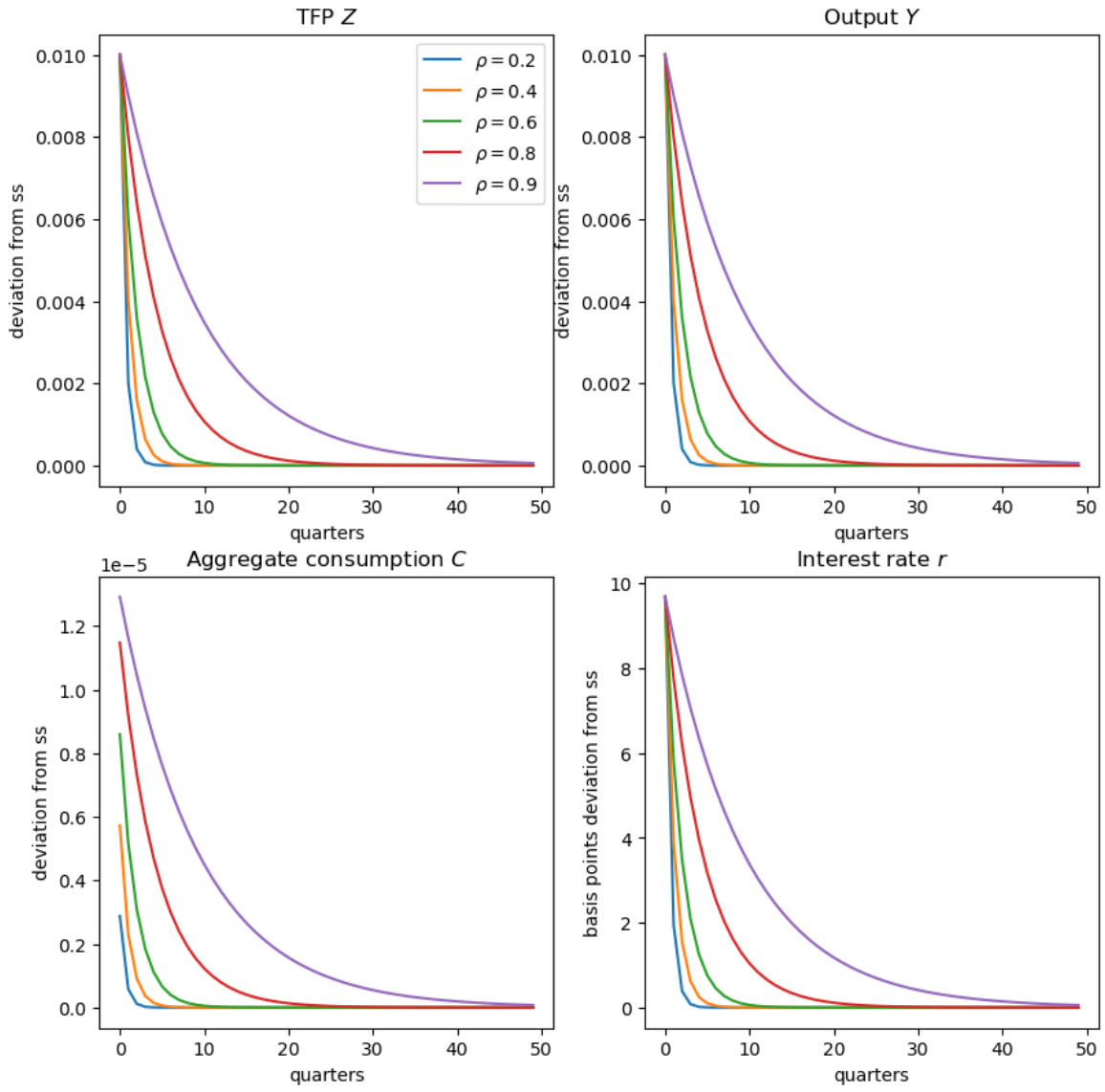


Figure 2.4: **Impulse responses in the Krusell-Smith model**

The graphs show the 1% TFP shock in the Krusell-Smith model for various persistencies $\rho = 0.5, \ldots, 0.9$ of the government spending shock. Computing the additional impulse responses takes virtually no time, since it breaks down to just one matrix multiplication.

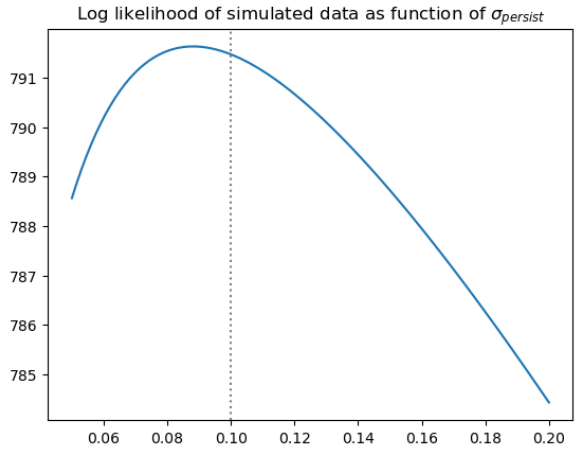


Figure 2.5: **Log-Likelihood estimation in the Krusell-Smith model**
The mode of the log-likelihood function is close to the true value of σ_{persist}, although not exactly at .1, since we only use a finite sample for the simulation.

of Auclert et al. (2021), it is possible to even use their SSJ Toolkit to perform the estimation, after computing the Jacobians with our approach. This makes it easy to perform them for other models. We provide the relevant snippets from their toolkit within our GitHub codebase.

B.1.6 Existence and uniqueness of infinitesimal generator

The existence of the infinitesimal generator $\mathcal{A}$ is not guaranteed by the definition of a continuous-time Markov chain. In fact, it is possible to construct continuous-time Markov chains with diagonal entries of A being -1, which intuitively means that the transition out of a state may be instantaneous. However, this possibility is eliminated in the context of economic analysis, as instantaneous transitions are not modeled in the Markov chains of interest (Poisson processes and diffusion processes). For more information on the existence and uniqueness of infinitesimal generators, see Chapter 1 of Pazy (1992). Essentially, we need the linear operator to be bounded.

B.2 Discrete-time one-asset HANK model

Here, we present a simplified discrete-time one-asset model based on the Appendix in Auclert et al. (2021), specifically the complementary code notebooks. Assume there is a mass of heterogeneous households in discrete time that maximize their lifetime utility

$$\max_{c_{it}} \mathbb{E}_o \left[\sum_{t=0}^{\infty} \beta^t \frac{c^{1-\gamma}}{1-\gamma} \right],$$

subject to the borrowing constraints

$$\begin{aligned} c_{it} + a_{it} &\leq (1 + r_{t-1})a_{it-1} + w_t n_{it} e_{it} - \tau_t, \\ a_{it} &\geq \underline{a}, \end{aligned}$$

where a_{it} denotes the wealth of a household; w_t denotes the wage, which is the same across households; τ_t is the time-varying tax rate; and e_{it} evolves according to an AR(1) process with persistence ρ_e and standard deviation σ. Further, assume all agents work the same number of hours, i.e., $n_{it} = N_t$. Technology is linear in labor, prices are sticky à la Rotemberg, and monetary policy follows a Taylor rule. The model can be summarized in sequence form as

$$\mathbf{F}_t(\mathbf{X}, \mathbf{Z}) \equiv \begin{pmatrix} Y_t - Z_t N_t \\ Y_t - w_t N_t \\ r_t B + G - \tau_t \\ \mathcal{A}_t(r_s, w_s, \tau_s) - B_t \\ \mathcal{C}_t(r_s, w_s, \tau_s) - C_t \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \qquad t = 0, 1, \ldots,$$

where the endogenous variables are $\mathbf{X} = (Y, N, r, w, \tau)$ and the exogenous variables are $\mathbf{Z} = (Z, G)$. The asset demand and consumption functions $\mathcal{A}$ and $\mathcal{C}$ follow from the household block with Bellman equation

$$\begin{aligned} V_t(e_{it}, a_{t-1,it}) = \max_{c_{it}, a_{it}} &\left\{ \frac{c_{it}^{1-\sigma}}{1-\sigma} + \beta \mathbb{E}_t \left[V_{t+1}(e'_{it}, a_{it}) | e_{it} \right] \right\}, \\ c_{it} + a_{it} &= (1 + r_t)a_{t-1,it} + w_t n_{it} e_{it} - \tau_t, \\ a_{it} &\geq 0. \end{aligned}$$

B.3 Comparison with the method used in Achdou et al. (2022)

Here, we follow Achdou et al. (2022) and Kaplan et al. (2018), especially numerical Appendix 3.11. Similar to our setup, their algorithm assumes that the economy starts in a steady state and then is hit by a one-time zero-probability shock. Furthermore, they also assume that the economy transitions back to steady state after an extended period of time. The algorithm proceeds as follows:

1. Compute the steady-state equilibrium of the model.

2. Guess the transition paths of the aggregate outcomes of interest. For example, in the Krusell-Smith economy, guess the path of aggregate capital K.

3. From market clearing conditions, compute the excess demand/supply of those aggregated variables and update them, i.e., lower the guess of capital if there is excess supply and vice versa.

4. Using the updated guess, compute the implied prices and equilibrium objects.

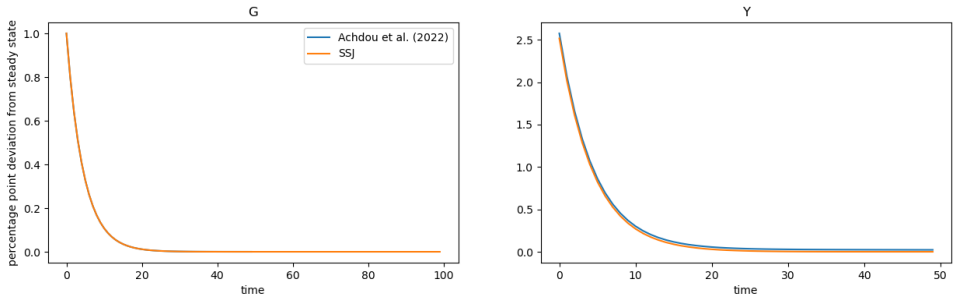


Figure 2.6: **Comparison of SSJ to Achdou et al. (2022)**
The right-hand panel shows the impulse response of output to a 1% deficit-financed government spending shock. The blue line displays the impulse response that was calculated with the method described in Achdou et al. (2022), and the orange line was computed by using the sequence-space Jacobian approach. Both methods obtain the same result, but the SSJ does so approximately 1,500 times faster in computing the impulse response.

5. Solve the households' transition problem, and find the path of distributions using the method described by Achdou et al. (2022). For example, in the one-asset HANK model, one has to use the time-dependent HJB and FPE from equations (2.2)–(2.4).

6. Using the path of distributions, compute the implied path for aggregate outcomes.

7. Compute the error ε between subsequent updates of K using an appropriate norm.

8. If the error is below a predefined threshold, stop, as an equilibrium has been found. Otherwise, adjust the guess using, for example, a Newton-type scheme, and return to step 2.

Figure 2.6 compares this approach to the sequence-space Jacobian approach for a deficit-financed government spending shock in the one-asset HANK model from Section 2.2. Both algorithms converge to the same transition path for output. However, the sequence-space Jacobian approach does this considerably faster. Computing the steady-state equilibrium takes 0.2 seconds in both approaches. Afterward, computing the transition paths with the method of Achdou et al. (2022) takes 14.44 seconds, whereas the sequence-space Jacobian approach requires only .0094 seconds, and computing additional impulse responses is even faster, since it is a matrix multiplication. For example, computing five additional impulse responses for different values of ρ, ranging from 0.5 to 0.9, takes 0.00009 seconds. All computations were performed using a MacBook Pro with an M1Max chip.

Bibliography Chapter 2

Achdou, Y., Han, J., Lasry, J.-M., Lions, P.-L., and Moll, B. (2022). Income and wealth distribution in macroeconomics: A continuous-time approach. *The Review of Economic Studies*, 89(1):45–86.

Ahn, S., Kaplan, G., Moll, B., Winberry, T., and Wolf, C. (2018). When inequality matters for macro and macro matters for inequality. *NBER Macroeconomics Annual*, 32(1):1–75.

Auclert, A., Bardóczy, B., Rognlie, M., and Straub, L. (2021). Using the sequence-space jacobian to solve and estimate heterogeneous-agent models. *Econometrica*, 89(5):2375–2408.

Bayer, C. and Luetticke, R. (2020). Solving discrete time heterogeneous agent models with aggregate risk and many idiosyncratic states by perturbation. *Quantitative Economics*, 11(4):1253–1288.

Bilal, A. and Goyal, S. (2025). Some pleasant sequence-space arithmetic in continuous time. Technical report, National Bureau of Economic Research.

Boppart, T., Krusell, P., and Mitman, K. (2018). Exploiting mit shocks in heterogeneous-agent economies: the impulse response as a numerical derivative. *Journal of Economic Dynamics and Control*, 89:68–92.

Haavelmo, T. (1945). Multiplier effects of a balanced budget. *Econometrica*, pages 311–318.

Kaplan, G., Moll, B., and Violante, G. L. (2018). Monetary policy according to hank. *American Economic Review*, 108(3):697–743.

Krusell, P. and Smith, Jr, A. A. (1998). Income and wealth heterogeneity in the macroeconomy. *Journal of Political Economy*, 106(5):867–896.

McKay, A., Nakamura, E., and Steinsson, J. (2016). The power of forward guidance revisited. *American Economic Review*, 106(10):3133–58.

Parra-Alvarez, J. C., Polattimur, H., and Posch, O. (2021). Risk matters: Breaking certainty equivalence in linear approximations. *Journal of Economic Dynamics and Control*, 133:104248.

Pazy, A. (1992). *Semigroups of linear operators and applications to partial differential equations*, volume 44. Springer Science & Business Media.

Reiter, M. (2009). Solving heterogeneous-agent models by projection and perturbation. *Journal of Economic Dynamics and Control*, 33(3):649–665.

Rouwenhorst, K. G. (1995). Asset pricing implications of equilibrium business cycle models. In *Frontiers of Business Cycle Research*, pages 294–330. Princeton University Press.

Winberry, T. (2018). A method for solving and estimating heterogeneous agent macro models. *Quantitative Economics*, 9(3):1123–1151.

Chapter 3

Projecting the Spread of COVID-19 for Germany

joint work with Jean Roch Donsimoni[(a,c)], Bodo Plachter[(a,d)] and Klaus Wälde[(a,c,e),1]

We model the evolution of the number of individuals that are reported to be sick with COVID-19 in Germany. Our theoretical framework builds on a continuous time Markov chain with four states: healthy without infection, sick, healthy after recovery or after infection but without symptoms and dead. Our quantitative solution matches the number of sick individuals up to the most recent observation and ends with a share of sick individuals following from infection rates and sickness probabilities. We employ this framework to study inter alia the expected peak of the number of sick individuals in a scenario without public regulation of social contacts. We also study the effects of public regulations. For all scenarios we report the expected end of the CoV-2 epidemic.
We have four general findings: First, current epidemiological thinking implies that the long-run effects of the epidemic only depend on the aggregate long-run infection rate and on the individual risk to turn sick after an infection. Any measures by individuals and the public therefore only influence the dynamics of spread of CoV-2. Second, predictions about the duration and level of the epidemic must strongly distinguish between the officially reported numbers (Robert Koch Institut, RKI) and actual numbers of sick individuals. Third, given the current (scarce) medical knowledge about long-run

[1]Jean Roch Donsimoni and Klaus Wälde are at the Johannes Gutenberg Universität Mainz, Gutenberg School of Management and Economics, Jakob-Welder-Weg 4, D-55131 Mainz, Telefon + 49.6131.39-20143, jdonsimo@uni-mainz.de, waelde@uni-mainz.de. Bodo Plachter is at the Institute for Virology of the University Medical Center Mainz, Mainz, Germany, plachter@uni-mainz.de. We are grateful to Matthias an der Heiden, Michael Bauer, Matthias Birkner, Christian Dormann, Oliver Emrich, Hoang Khieu, Dietmar Leisen, Kai Pastor, Olaf Posch, Julien Riou, Perikles Simon and members of the modelling initiative of the Deutsche Gesellschaft für Epidemiologie for comments and discussions. We are deeply indebted to Damir Stijepic for the initial numerical solutions and the initial calibration. Our biggest debt of gratitude is owed to Matthias Birkner who has been advising and teaching us on stochastic models for many years.

infection rate and individual risks to turn sick, any prediction on the length (duration in months) and strength (e.g. maximum numbers of sick individuals on a given day) is subject to a lot of uncertainty. Our predictions therefore offer robustness analyses that provide ranges on how long the epidemic will last and how strong it will be. Fourth, public interventions that are already in place and that are being discussed can lead to more and less severe outcomes of the epidemic. If an intervention takes place too early, the epidemic can actually be stronger than with an intervention that starts later. Interventions should therefore be contingent on current infection rates in regions or countries.

Concerning predictions about COVID-19 in Germany, we find that the long-run number of sick individuals (that are reported to the RKI), once the epidemic is over, will lie between 500 thousand and 5 million individuals. While this seems to be an absurd large range for a precise projection, this reflects the uncertainty about the long-run infection rate in Germany. If we assume that Germany will follow the good scenario of Hubei (and we are even a bit more conservative given discussions about data quality), we will end up with 500 thousand sick individuals over the entire epidemic. If by contrast we believe (as many argue) that once the epidemic is over 70% of the population will have been infected (and thereby immune), we will end up at 5 million cases.

Defining the end of the epidemic by less than 100 newly reported sick individuals per day, we find a large variation depending on the effectiveness of governmental pleas and regulations to reduce social contacts. An epidemic that is not influenced by public health measures would end mid June 2020. With public health measures lasting for few weeks, the end is delayed by around one month or two. The advantage of the delay, however, is to reduce the peak number of individuals that are simultaneously sick. When we believe in long-run infection rates of 70%, this number is equally high for all scenarios we went through and well above 1 million. When we can hope for the Hubei-scenario, the maximum number of sick individuals will be around 200 thousand "only".

Whatever value of the range of long-run infection rates we want to assume, the epidemic will last at least until June, with extensive and potentially future public health measures, it will last until July. In the worst case, it will last until end of August.

We emphasize that all projections are subject to uncertainty and permanent monitoring of observed incidences are taken into account to update the projection. The most recent projections are available at https://www.macro.economics.uni-mainz.de/corona-blog/.

3.1 Introduction

There is no need to stress the importance of the Coronavirus disease (COVID-19) for public health, economic consequences and well-being of individuals. Yet, there is a need for more knowledge about objective information and reliable predictions of how the pandemic will likely evolve in the months to come. How large is the risk of any individual to get infected on a particular day or within a week? How does this risk change over time? How large is the expected number of individuals ever infected

and how large is the number of individuals infected at any point in time? The latter question is the main concern of the public, politicians and health practitioners to ensure sufficient provision of health services. Finally, what are the effects of policy interventions that are already implemented and that are being discussed?

This paper starts by describing what we know quantitatively about the spread of the Corona epidemic in the Chinese province of Hubei and in South Korea. This offers insights from two episodes of epidemics that seem to be coming to an end and that help in making predictions for other countries. We then compare the time-series evidence from Hubei and South Korea with cross-sectional information in European countries.

Section 3 develops an epidemiological Markov model. We start from individuals that can be in four states: healthy (the initial state, state 1), infected and sick (state 2), infected and recovered or infected and never having displayed symptoms (state 3) and dead (state 4). Individuals move between these states with endogenous transition rates that depend on population characteristics. This model allows to make predictions, inter alia, about all the quantities of interest that appeared in the questions raised initially.

We apply this model to Germany. We choose parameter values such that the time series of the number of sick (and reported) individuals in Germany since 24 February is matched by our model. Our data source is the Robert Koch Institute (RKI, 2020). In order to predict the future evolution, we build on various evidence from epidemiology like individual sickness probability and long-run infection rates of the population as a whole. We emphasize the scarce medical and statistical knowledge on these parameters. Any predictions therefore must be subject to large variations. We therefore undertake robustness checks with respect to our central parameters.

Given our calibration for the epidemic without public intervention, we predict the number of sick individuals at each point in time. We compute when the peak will be reached and whether medical services are in sufficient supply. We also undertake counterfactual analysis to understand when drastic policy measures (like shutting down educational institutions or canceling big public events) are most effective.

There is an exploding literature on these issues, especially in medical science. The classic large-scale spatial model is described in Balcan et al. (2010). It is applied in Chinazzi et al. (2020) to study the effect of a travel ban on China and the world. Akbarpour and Jackson (2018) study a general model of infection, but cannot provide applications to the current epidemic. The widely discussed study by Ferguson et al. (2020) focuses on the US and the UK.[2] An insightful study on the usefulness of isolation strategies for COVID-19 based on a stochastic transition model similar to ours is by Hellewell et al. (2020). Wilder-Smith et al. (2020) discuss the similarities and differences between the SARS 2003 (severe acute respiratory syndrome) epidemic and COVID-19. One big issue is the large

[2] An essay by Simon (2020) argues that the death rate in Germany is much smaller than in other countries. This makes it difficult to apply the Ferguson et al. (2020) findings to Germany.

number of "quiet infections" in COVID-19 as opposed to SARS 2003, an aspect we will capture in our model below.

The analyses which come closest to ours are two notes by the Deutsche Gesellschaft für Epidemiologie (DGfE, 2020) and by an der Heiden and Buchholz (2020). DGfE (2020) offer predictions based on a model similar to ours (so called SEIR models, see e.g. Anderson et al., 1992). We present our model in detail, including the stochastic foundation, and discuss the implications of the modelling assumptions.[3] Modelling assumptions turn out to be crucial for evaluating public policy measures. We also implement in detail the effects of policy measures and discuss the trade-offs with respect to length and strength of the evolution of the epidemic. The match of the model with the data also differs. While DGfE (2020) seem to use summary statistics for their calibration, we determine the parameters of our model by fitting our model predictions to the data. This offers the advantage that we can update our predictions whenever new data is available.

The study by an der Heiden and Buchholz (2020) is, to the best of our knowledge, the most elaborate study at this point for Germany. We share their belief that mathematical models help in studying the effects of policy measures. Many parameters of their analysis are derived from knowledge on the epidemic in China and especially the city Shenzen. We calibrate many of our central parameters such that the model fits the observations in Germany. We also share with them the approach of using SEIR models. Our paper emphasises the implications of the fundamental model assumptions for the prediction. We find below that the evaluation of policy measures crucially depends on these model assumptions. We make these model assumptions explicit in our paper and discuss the effects of variations in these assumptions.

Why should economists work on an epidemic? The economic costs of COVID-19 are huge and seem to be larger than those of the financial crisis starting in 2007.[4] This paper provides a model that allows to understand the spread of the disease. We employ a model in the tradition of search and matching models originating from Diamond (1982), Mortensen (1982) and Pissarides (1985). The basic structure of these models (continuous time Markov chains) is identical to epidemiological models (as becomes clear very quickly when reading overviews like e.g. Hethcote, 2000). Compared to standard search and matching models, we do allow for four states, however. We extend the typical matching framework from a technical perspective by taking the stochastic nature of transitions and their prediction for the probability to be in various states into account. These probabilities are

[3] There are also various internet pages that offer projections for parameter models that can be chosen by the user. Examples include https://gabgoh.github.io/COVID/index.html, https://neherlab.org/covid19/ or http://covidsim.eu/. These pages usually do not show the details of the model and a comparison to our approach is therefore difficult.

[4] Eichenbaum et al. (2020) study the effects of the CoV-2 epidemic on economic decisions of households. They emphasize the household decisions to reduce labour supply exacerbate the recessionary effects of the epidemic. They apply their model to the US. Barro et al. (2020) employ mortalities and economic consequences of the "Spanish Flu" to provide estimates of the world wide consequences of COVID19. They do not offer a detailed projection for Germany. The costs for Germany are estimated by Dorn et al. (2020). Atkeson (2020) provides esitmates of economic consequences of COVID19 for the US.

described by forward Kolmogorov equations. As we work with a discrete number of states, we obtain an ordinary differential equation system.[5] The conclusion discusses various next steps that should be undertaken to understand economic consequences of COVID-19 better. We believe that our framework can form the basis for an understanding of economic costs of the pandemic.

3.2 Data

This section describes what we know quantitatively about the spread of reported infections. We start with Hubei and its capital Wuhan in China. We afterwards look at South Korea and very briefly at Japan. We focus on the evolution of the number of individuals that were ever reported to be sick during the current epidemic. This number, by construction, can only rise. We chose Hubei and South Korea as the epidemics seem to come to an end in these countries. Given the huge uncertainty about long-run infection rates and individual risk to get sick after an infection with CoV-2, we consider the long-run ratios of sick reported individuals to the overall population a benchmark that allows us to judge the credibility of our predictions for Germany. We are aware that we compare very different countries with very different cultural and political habits and medical systems. For reference, we also look at a cross section of infections for European countries. See the appendix for a description of data sources.[6]

- Hubei

Since the onset of the outbreak in the Hubei province in China, the number of confirmed cases has risen dramatically, peaking at 67,801 (WHO Situation Reports, 2020).[7] Reporting began in January 2020 and we note that in mid-February, the WHO changed its classification methodology by reporting both clinically and lab confirmed cases rather than only lab confirmed cases. In terms of cumulative numbers, the total reported cases of infected individuals rose continuously since January, since there have been continuous influxes of new cases and very little communication about individuals transitioning back to being healthy.

When reporting began, the ascension of new cases seemed to follow an exponential curve, with the number of new cases constantly increasing, however in early February, this trend stopped. Indeed, another trend then took over following a more S-shaped pattern, i.e. that of a sigmoidal function. The total number of reported cases over time is shown in Figure 3.1. The figure shows a clear explosive trend at the start of the outbreak before slowing down and approaching what appears to be an asymptote around 70,000.

[5] When working with continuous states like e.g. wealth or labour income, partial differential equations result. See e.g. Bayer and Wälde (2010a), Achdou et al. (2020), Kaplan et al. (2018) or Khieu and Wälde (2020).

[6] Note that all figures below are as of 26 March 2020 and come from the Johns Hopkins University dataset from Dong et al. (2020).

[7] There are attempts to correct supposedly under reported numbers from China by employing South Korea as a benchmark country (Lachmann, 2020).

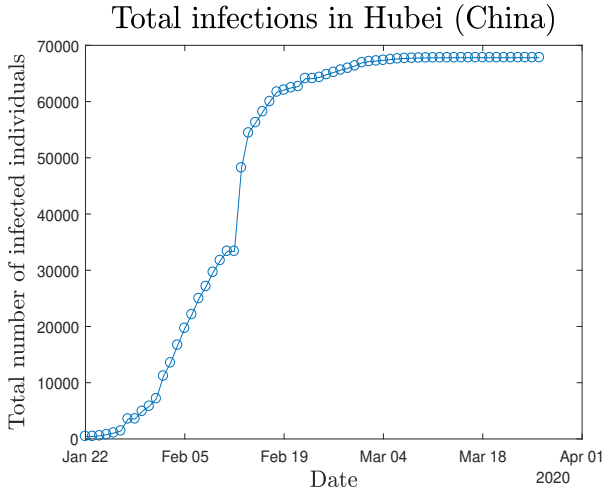


Figure 3.1: Total cases in Hubei seem to converge to a constant

To make this comparable to other countries, we take the population of around 59.17 million individuals in Hubei into account. This implies an infection rate of about 0.11% or 1 person in 873 being infected in the long-run. While this is only one province, it does provide us with a reference point for what to expect in other regions of the world.

- South Korea and Japan

Soon after China, South Korea, Japan, and many other countries all started seeing a rise in their number of confirmed cases. However, these countries have had very different experiences when it comes to infection outbreaks. Indeed, while South Korea has seen the number of confirmed cases rise to 9,241, Japan only has 1,307 confirmed cases (or 14% that of South Korea). The question remains to know whether these countries are also experiencing a convergence to a stable number of (cumulative) confirmed cases. Figure 3.2 shows the cumulative number of confirmed cases in South Korea (left) and Japan (right).
The data from South Korea shows a reduction in the slope of the curve for total confirmed cases, indicating a significant drop in the number of new infections reported. Assuming the tendency continues, this would point toward South Korea progressively converging to a constant, similarly to the Hubei province in China. Japan on the other hand (as well as all other countries at this point) does not appear to slow down significantly. Relating South Korean infections to population size, we get a long-run infection rate of 1 in 5,337, with a peak number of confirmed cases of ca. 9,606.

- Europe

In Europe, however, the situation continues to deteriorate, with new cases being reported by authorities on a daily basis numbering in the thousands (WHO Situation Reports, 2020). However, bearing in mind the lessons learned from the case of the Hubei province, and to a lesser extent from the case of

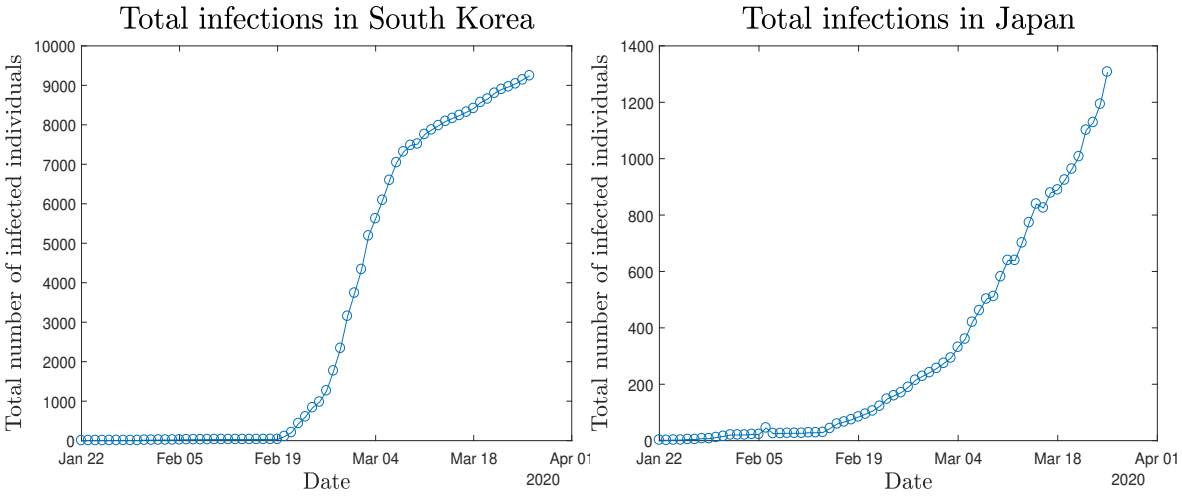


Figure 3.2: Total cases in South Korea exhibit a similar pattern to the Hubei province

South Korea, this current period of explosive growth should, in theory, be followed by a slow tapering leading to a plateauing of the total number of cases. We report in Table 3.1 the current infection rates in some European countries. The sample below represents 58% of the total EU population, including Norway and Switzerland. This sample is by no means intended to be representative and is for illustrative purposes only by considering some of the most affected economies.

As we can see from Table 3.1, the degree of infection per country varies from 1 in 4,050 to 1 in 784, a fivefold ratio. Other countries fare better, the UK for example has a risk of 1 in 6,914, however we can clearly see that these figures are not yet as low as those in Hubei, with the exception of Italy and Switzerland, which are now below the 1 in 873 rate observed in the Chinese province, and Spain, which is approaching it fast. If this case is to be repeated, then things will likely worsen before they improve. However, with measures taken by governments, and progress being made on the medical front for an effective vaccine, there is still room for governments to curb these rates despite the rapid rise in cases.

3.3 The model

3.3.1 The structure

The model builds on a continuous time Markov chain with 4 states. We look at health of an individual i that can be in four states $S = \{1, 2, 3, 4\}$. Healthy without infection ($s = 1$), sick ($s = 2$), dead ($s = 3$), healthy with or after infection ($s = 4$). State 1 is the initial state for all individuals, thus

Table 3.1: Current infection rates and risks in a sample of European countries

Country	Population (1)	Confirmed cases (2)	Infection rate (%) =(2)/(1)	Risk per 1 inhabitant =(1)/(2)
Sweden	10,151,866	2,526	0.025%	4,050
Netherlands	17,424,978	6,440	0.037%	2,684
France (metro)	64,906,000	25,600	0.039%	2,535
Belgium	11,455,519	4,937	0.043%	2,320
Germany	83,100,00	37,323	0.045%	2,227
Norway	5,515,736	3,100	0.058%	1,719
Spain	46,937,060	49,515	0.105%	948
Italy	60,359,546	74,386	0.123%	811
Switzerland	8,570,146	10,897	0.128%	784
Sample total	308,143,212	214,724	0.070%	1,435

$s_i(0) = 1$ for all $i = 1...N$ where N is initial population size.[8] There are individual transition rates λ_{rs} (with $r \in S$ and $s \in S$) between these states as illustrated in figure 3.3.

3.3.2 Transition rates

Starting with the transition from being healthy to sick, each individual in state 1 has a certain average number of social contacts per day. We capture this by an exogenous arrival or contact rate a. Once a contact takes place, there is a certain probability that the contact is with an individual that can infect the healthy individual. If we assumed random contacts with healthy and sick individuals, this probability would be $(N_2(t) + \eta N_4(t)) / (N_1(t) + N_2(t) + N_4(t))$, where $N_s(t)$ is the (expected) number of individuals in state s at t. This probability assumes that all $N_2(t)$ individuals that are sick can infect healthy individuals and a constant and exogenous share η of healthy individuals in state 4 can also infect healthy individuals from state 1. This parameter η captures the idea of "quiet infections" as emphasized e.g. by Wilder-Smith et al. (2020). As opposed to SARS 2003 (or the common flu), individuals infected with the CoV-2 do not necessarily display symptoms but can nevertheless infect others. There is also first evidence (Xing et al., 2020) that even recovered individuals (i.e. those that came from state 2) can carry the CoV-2, also supporting our use of a positive η.

In practice, individuals do not encounter sick individuals with this probability being equal to the share of sick individuals in society as sick individuals might stay at home or behave differently (not shaking hands or similar) than non-sick individuals. We do assume, however, that the probabil-

[8] In our quantitative solution below, the initial state follow from the number of reported sick individuals on 24 February 2020.

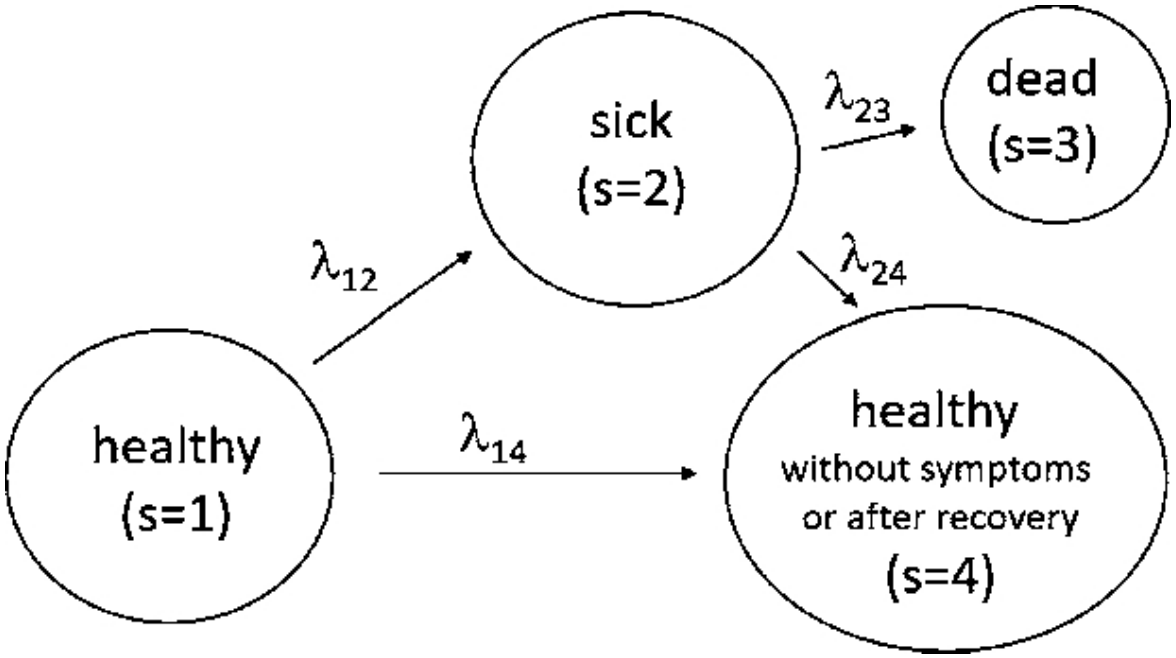


Figure 3.3: Transitions between the state of health (initial state), sickness, death and health despite infection or after recovery

ity $\pi(t)$ to encounter a sick individual rises in $N_2(t)$ and $\eta N_4(t)$ and falls in $N_1(t)$, i.e. $\pi(t) = \pi(N_2(t), \eta N_4(t), N_1(t))$ with $\partial\pi/\partial N_2 > 0, \partial\pi/\partial(\eta N_4) > 0$ and $\partial\pi/\partial N_1 < 0$.[9] These contacts can lead to infections. Hence, we have an individual sickness rate of $\lambda_{12}(t) \equiv a\pi(t)$. Finally, when there are $N_1(t)$ healthy individuals, the aggregate flow from state 1 to state 2 is $\phi_{12}(t) \equiv a\pi(t) N_1(t)$.[10] A further property an infection rate should reflect is the fact emphasized by many virologists and epidemiologists that once an epidemic is over, around two thirds of the entire population will be infected. This includes individuals that had symptoms at some point and asymptomatic cases. We capture this fact by introducing the infection rate which is simply the ratio of the number of infected individuals (sick and in state 2 or without symptoms in state 4) to individuals that are alive,

$$\rho(t) = \frac{N_2(t) + N_4(t)}{N_1(t) + N_2(t) + N_4(t)}. \tag{3.1}$$

The infection rate is zero initially at $t < 0$. Illustrated for Germany in a somewhat too simple way but perfectly fine for our projection, $\rho = 0$ on 23 February 2020 (and before) when the number of reported sick individuals was zero. On 24 February 2020, a number of $N_2(0) = 16$ are introduced into the system and infections and sickness start occurring. The sickness rate $\lambda_{12}(t)$ is zero also in the long-run when this infection rate equals the long-run infection rate, which we denote by $\bar{\rho}$. This

[9] Individuals that have died no longer play a role in social interactions. Hence, $N_3(t)$ does not appear as argument in the sickness rate.

[10] A parallel can be drawn between ϕ and a matching function, and λ_{12} and the job finding probability in search and matching frameworks in the tradition of Diamond (1982), Mortensen (1982) and Pissarides (1985).

captures the above mentioned fact that some individuals will never be infected during the epidemic and will always remain in state 1, $\lim_{t\to\infty} N_1(t) > 0$.

Summarizing our discussion, an example for the individual sickness rate that captures insights from epidemiology and virology would be

$$\begin{aligned}\lambda_{12}(t) &\equiv \lambda_{12}(a, N_1(t), N_2(t), \eta N_4(t), \rho(t)) \\ &= aN_1(t)^{-\alpha}(N_2(t) + \eta N_4(t))^{\beta}\left[\bar{\rho} - \frac{N_2(t) + N_4(t)}{N_1(t) + N_2(t) + N_4(t)}\right]^{\gamma}, \end{aligned} \tag{3.2}$$

where $0 < \alpha, \beta, \gamma < 1$ allows for some non-linearity in the process and $a > 0$. The first term $N_1(t)^{-\alpha}$ captures the idea that more healthy individuals reduce the individual sickness rate. The second term $(N_2(t) + \eta N_4(t))^{\beta}$ increases the sickness rate when there are more infectious individuals. The third term in squared brackets makes sure that the arrival rate is zero when a share ρ of society is sick (state 2) or healthy after infection (state 4). The sickness rate satisfies "no sickness without infected individuals", $\lambda_{12}(a, N_1, 0, 0, \rho) = 0$ and "end of spread at sufficiently high level", $\lambda_{12}(a, N_1, N_2, \eta N_4, \bar{\rho}) = 0$. In between these start- and endpoints, the infection rate will first rise and then fall. This specification makes sure that in the long run a share of around $1 - \bar{\rho}$ will not have left state 1, i.e. will never have been infected.[11]

As is widely documented and reported (see e.g. Nishiura et al., 2020), there is also a flow of individuals that are infected but do not display any symptoms. We denote this flow by $\phi_{14}(a, N_1(t), N_2(t), \eta N_4(t))$ and the corresponding individual rate by λ_{14}. The arguments can be rationalized in the same way as the flows into sickness.

Is it possible to obtain data for λ_{14}? One would need a cohort of inhabitants of a region (and not just those that contact medical services to get tested) to understand the share of individuals that are infected but do not display symptoms. Good data is not available (let alone time series) and might never be available, given the currently still large costs per individual tests (between 80 Euro and 120 Euro). We therefore assume that a constant share r of infected individuals display symptoms. This is a quantity many epidemiologists also work with. Hence, the transition rate λ_{14} is related to λ_{12} in the following sense,

$$r[\lambda_{14}N_1 + \lambda_{12}N_1] = \lambda_{12}N_1. \tag{3.3}$$

In words, the outflow of newly infected, $\lambda_{14}N_1 + \lambda_{12}N_1$ in squared brackets on the left-hand side, times the share r of individuals that show symptoms after infection gives the flow into sickness, $\lambda_{12}N_1$. We can therefore compute the transition rate from the state of being healthy ($s = 1$) to the state of healthy again or no symptoms ($s = 4$) as

$$r[\lambda_{14} + \lambda_{12}] = \lambda_{12} \Leftrightarrow \lambda_{14} = \frac{1 - r}{r}\lambda_{12}. \tag{3.4}$$

[11] We employ "around" as some individuals will have ended up in state 3 whose number does not enter the expression in (3.1).

Finally, there is an individual mortality rate and a recovery rate. Both rates depend on many individual characteristics. At the aggregate level, they rise in the stock of sick individuals. We specify them as $\phi_{23}(N_2(t))$ and $\phi_{24}(N_2(t))$ with corresponding rates λ_{23} and λ_{24}, respectively.[12] For our analysis below, we will assume that both rates are constants. Concerning recovery, we are aware that it strongly varies across age (e.g. Guan et al., 2020) and that recovery is not a state that is identical for each individual. As we abstract from ex-ante heterogeneity, we assume a recovery rate that is identical across individuals and that it takes on average n_{rec} days until recovery. Given the exponential distribution of duration in each state, this means that $1/\lambda_{24} = n_{\text{rec}}$ or

$$\lambda_{24} = n_{\text{rec}}^{-1}. \tag{3.5}$$

Death is an absorbing state. We also consider the final healthy state to be absorbing, i.e. we abstract from relapses.

3.3.3 Probabilities and means

- The evolution of probabilities, shares and expected numbers

We can express the transition across health states of an individual by a stochastic differential equation similar to the transition across states of employment (see e.g. appendix $B.1$ to Bayer et al., 2019 or Khieu and Wälde, 2020). We can also directly write down ordinary differential equations for an individual to be in a certain state s at t. Let us denote this probability by $p_s(t) = \text{Prob}(s_i(t) = s | s_i(0) = 1)$, i.e. as the probability to be in state s at t conditional on having started in state 1 at $t = 0$. The ODEs have the following general form (see Ross, 1993 for background),

$$\begin{aligned}
\dot{p}_1(t) &= -\lambda_{12}p_1(t) - \lambda_{13}p_1(t) - \lambda_{14}p_1(t) + \lambda_{21}p_2(t) + \lambda_{31}p_3(t) + \lambda_{41}p_4(t), \\
\dot{p}_2(t) &= -\lambda_{21}p_2(t) - \lambda_{23}p_2(t) - \lambda_{24}p_2(t) + \lambda_{12}p_1(t) + \lambda_{32}p_3(t) + \lambda_{42}p_4(t), \\
\dot{p}_3(t) &= -\lambda_{31}p_3(t) - \lambda_{32}p_3(t) - \lambda_{34}p_3(t) + \lambda_{13}p_1(t) + \lambda_{23}p_2(t) + \lambda_{43}p_4(t),
\end{aligned} \tag{3.6}$$

and are completed by the identity $p_4(t) = 1 - p_1(t) - p_2(t) - p_3(t)$. For brevity, individual transition rates are represented by λ_{rs}. We stress that these equations hold for arrival rates that can take any functional form with arguments N_1, N_2, N_3 and N_4.

As death is an absorbing state, $\lambda_{3i} = 0$ for all i. Given the assumption that infected means being immune to further infection and that the mortality rate is negligible in symptom-free states, we also set $\lambda_{21} = \lambda_{41} = \lambda_{13} = \lambda_{43} = 0$. Our final structure for the evolution of individual probabilities therefore reads

$$\dot{p}_1(t) = -(\lambda_{12} + \lambda_{14})\, p_1(t), \tag{3.7a}$$

[12] One could plausibly argue and find support in micro studies that the death and recovery rate also depend on the number of healthy individuals (that are required to take care of sick individuals). These considerations are not crucial for our macro analysis and are therefore neglected.

$$\dot{p}_2(t) = -(\lambda_{23} + \lambda_{24}) p_2(t) + \lambda_{12} p_1(t), \tag{3.7b}$$

$$\dot{p}_3(t) = \lambda_{23} p_2(t). \tag{3.7c}$$

A standard law of large numbers tells us that individual probabilities also give shares in the population. Hence, when we solve this system, we can think of the time paths of $p_s(t)$ also as the shares of society in the different states. When we compute the expected number of individuals in state s at t, denoted by $N_s(t)$, we would also obtain an identical ODE system where the probabilities are replaced by the expected numbers.[13] As the interest of our analysis lies in the (expected) number of individuals in certain states (are there sufficiently many beds in hospital for severe cases?), we will also refer to the variable $p_s(t)$ as the (expected) number $N_s(t)$ of individuals.

- The probability to get sick per day

For our quantitative predictions it will be useful to present the probability for an individual to get sick over a period of one day or one week. So far, we only talked about rates λ_{rs}. We denote the probability to get sick over a period of one day, starting in t, by $p^{\text{gs}}_{\text{day}}(t)$. This probability is one minus the probability to stay healthy. The probability to stay healthy is, given the exponential distribution of duration in a state, a simple exponential function of the transition rate. We therefore obtain

$$p^{\text{gs}}_{\text{day}}(t) = 1 - e^{-\int_t^{t+1} \lambda_{12}(s) ds}, \quad p^{\text{gs}}_{\text{week}}(t) = 1 - e^{-\int_t^{t+7} \lambda_{12}(s) ds}. \tag{3.8}$$

The second probability $p^{\text{gs}}_{\text{week}}(t)$ is the probability to get sick between today in t and the next 7 days to come. We will plot them below to provide a measure of how risky social interactions are during the evolution of the epidemic.

3.3.4 The number of individuals who get and ever got sick and deaths

As recovery is a process that enters additional measurement problems, a more reliable data source for our calibration purposes is the number of individuals that have been reported to be sick since the onset of the epidemic. This is also the time series that is typically reported by countries and regions (such as those reported by the RKI in Germany or the WHO more generally). In our model, the corresponding number follows from the inflows from state 1 to 2 and is given by

$$\dot{N}^{\text{ever}}_2(t) = \lambda_{12}(t) N_1(t) = \lambda_{12}(t) p_1(t) N. \tag{3.9}$$

When we are interested in the number of individuals that are newly reported to be sick at some point in time t (incidences at t in short), we should not employ $N_2(t)$. The change of the latter, described

[13] Let us denote the expected number of individuals in state s at t by $N_s(t) = E_0 \tilde{N}_s(t)$ where $\tilde{N}_s(t)$ is the stochastic process for the number of individuals in state s at t. As the expected share of individuals in state s equals the probability to be in this state, $N_s(t)/N = p_s(t)$, we can write $N_s(t) = p_s(t) N$ for the expected number of individuals in this state. Replacing the probabilities $p_s(t)$ by $N_s(t)$ would yield an ODE system in $N_s(t)$. Our system employed for predictions below in (3.15) is expressed in this way.

in (3.7b), is determined by the newly reported (inflow) minus the outflow, i.e. those that recovered (and those who died). Hence, when data provides incidences on a given day t, we need to compare this with the inflow between yesterday $t-1$ and t. Hence, our theoretical counterpart to incidences is

$$N_2^{\text{new}}(t) = \int_{t-1}^{t} \lambda_{12}(s)\, p_1(s)\, N ds. \tag{3.10}$$

Finally, the number $N_3(t)$ of deaths from Corona by t is determined by the constant death rate λ_{23} applied to those that are currently sick,

$$N_3(t) = \int_{0}^{t} \lambda_{23} p_2(s)\, N ds. \tag{3.11}$$

3.3.5 Model properties

Before we calibrate our model, we briefly discuss its theoretical predictions. This illustrates the plausibility of our assumptions and also the flexibility of the model. Our model consists of the ODE system (3.7) where we replace the probabilities $p_s(t)$ by the expected number $N_s(t)$ as described in footnote 13. The sickness rate λ_{12} is from (3.2), the other arrival rates λ_{14} and λ_{24} are from (3.4) and (3.5), respectively. The death rate λ_{23} is a constant. We start the solution of our model with $N_2^{\text{ever}}(0) = N_{\text{observed}}^{\text{ever}}(0)$. We set $N_3(0) = N_4(0) = 0$ and $N_1(0) = N - N_2^{\text{ever}}(0)$.

One can easily understand the plausible structure of our ODE system (3.7). The number $N_1(t)$ of healthy individuals in state 1 (i.e. the probability or share $p_1(t)$) can only fall. Individuals either get sick with rate λ_{12} or get infected without symptoms with rate λ_{14}. Both transitions imply an outflow from state 1. The same is true, mutatis mutandis, for state 3 and 4: For state 3, there are only inflows from state 2 implying that $N_3(t)$ increases over time. State 4 only experiences inflows from state 1 and 2 implying that $N_4(t)$ increases over time. The number $N_2(t)$ of sick individuals can rise or fall, depending on the difference between inflows with λ_{12} and outflows with λ_{23} or λ_{24}.

In the long-run, i.e. when the epidemic is over, individuals will be distributed across states 1, 3 and 4. Some individuals will never get infected. They therefore remain in state 1 all throughout the epidemic. This finding follows from the fact that the sickness rate λ_{12} from (3.2) falls to zero when the infection rate $\rho(t)$ from (3.1) equals its long-run value $\bar{\rho}$. Note that $\lambda_{12} = 0$ implies from (3.4) that $\lambda_{14} = 0$ as well, hence outflows from 1 to 4 end as well. The ratio $\bar{\rho}$ is a widely accepted quantity in epidemiology and will be quantified below. Some individuals that turn sick and are in state 2 die. They end up in state 3 in the long-run. Most individuals will be in state 4 in the long-run, either after having transitioned through state 2 or directly from state 1.

The model also displays a long-run property that is extremely useful to judge its quantitative property in a very simple way. With a population of N individuals and a long-run infection rate $\bar{\rho}$ discussed after (3.1), the number of individuals that will have been infected is $\bar{\rho}N$. When a share r of all infected

turns sick, the long-run number of individuals that were sick at some point, described in (3.9), is

$$\lim_{t\to\infty} N_2^{\text{ever}}(t) = r\bar{\rho}N. \tag{3.12}$$

The long-run number of reported sick individuals therefore only depends on the long-run infection rate $\bar{\rho}$ and the share r of individuals that turn sick *and* are reported. We will exploit this property when we discuss the effects of policy measures below.

3.4 Calibration

3.4.1 Data

Our most important data source is the number of individuals that is reported to be sick. We obtain these outflows from state 1 to 2 from the Robert Koch Institute (RKI, 2020). These data fix the left-hand side of our time path for the number $N_2^{\text{ever}}(t)$ from (3.9) of individuals that are ever reported to be sick. To fix notation, $t = 0$ stands for the 24th February 2020 and T is the day of our final observation.

It seems reasonable to assume that parameters are not constant throughout the entire epidemic. Most centrally, the contact rate a employed in the sickness rate (3.2) will change as a function of governmental regulations. The first major effects consisted in cancelling sports events (14 March) and closing schools (16 March). Other measures followed.[14] Given a median incubation of 5.2 days (Linton et al., 2020, Lauer et al., 2020) and a certain delay between feeling symptoms, contacting a doctor and getting reported at RKI (say, 2-3 days), we employ T = 21 March 2020 as our last observation to fix the contact rate a.

Data to compute the death rate from Corona are also taken from the RKI (2020). We finally employ various sources, to be discussed in what follows, to fix the share r of infected individuals that turn sick and the long-run infection rate $\bar{\rho}$.

3.4.2 Targets and parameters

- Parameters

Some model parameters are straightforward to be fixed. Concerning the recovery rate (3.5), we set n_{rec} equal to an average of 14 days despite strong heterogeneity in the course of disease (Guan et al., 2020). The death rate follows from (3.11). As we want to match the number of reported deaths from COVID-19 by T, the constant death rate for the period from 0 to T can be computed from

$$\lambda_{23}(T) = \frac{N_3^{\text{obs}}(T)}{\int_0^T p_2(s)\, dsN}, \tag{3.13}$$

[14] See https://www.macro.economics.uni-mainz.de/corona-blog/ for more details.

Table 3.2: Parameters for the epidemic without public health measures

n_{rec}	λ_{23}	r	η	a	$\bar{\rho}$	α	β	γ
14	1/500	0.1	0.4	$3.024/10^6$	2/3	0.5751	0.8662	.6459

where $N_3^{\text{obs}}(T)$ is the number of dead individuals at T. Employing this equation yields the value of 1/500 in table 3.2.

A more open parameter for this epidemic but for which there is a lot of information for other epidemics is the long-run share of infected individuals, i.e. the limit $\bar{\rho}$ of (3.1). We set this equal to 0.67 meaning that once the epidemic is over, two thirds of the population will have been infected and one third are still in the original healthy state 1. As there is a lot of uncertainty concerning this value, we consider this widely employed value of 2/3 initially. We will then also consider one tenth of 0.67 as a lower bound further below. This lower bound is motivated by observations from Hubei and South Korea discussed in section 2.[15] Another parameter which is hard to pin down is the share η of healthy individuals in state 4 (i.e. they were or are infected at some point) that can infect other individuals. We set it equal to 0.4. We also undertake robustness analyses with respect to this parameter.

The most interesting parameters are those that allow us to match data reported by RKI. To do so, we minimize the Euclidean distance between the reported data and the predicted values of the model. We target a weighted sum of the squared difference between $N_2^{\text{ever}}(t)$ from (3.9) and observation and the newly-sick $N_2^{\text{new}}(t)$ from (3.10) and observation. More precisely, parameters a, α, β and γ are obtained from

$$\min_{a,\alpha,\beta,\gamma} \Sigma_{t=1}^{T} \left(N_2^{\text{ever}}(t) - N_{2,\text{observed}}^{\text{ever}}(t)\right)^2 + \left(N_2^{\text{new}}(t) - N_{2,\text{observed}}^{\text{new}}(t)\right)^2. \tag{3.14}$$

We impose constraints for α, β, γ to lie between zero and one and for a to be positive. None of the constraints are binding. Table 3.2 presents these and all the other parameter values.

We finally need to quantify r. A generally accepted benchmark says that around 80% of infected individuals do not display any or only weak symptoms.[16] The remaining share of 20% turns sick. One crucial question for equation (3.4), but also for our long-run prediction from (3.12), is therefore the issue of reporting. If a large share of sick individuals show up at the general practitioner who

[15] One might conjecture to employ so-called reproduction numbers to get an idea about $\bar{\rho}$. A reproduction number gives the number of infections in a non-immune population when one infected individual is introduced. Riou and Althaus (2020) estimate this value to lie in the range of 2.0 to 2.5 for Wuhan from December 2019 to January 2020. The WHO (2020) reports it in the same range. These reproduction numbers change over time (as our sickness rate $\lambda_{12}(t)$ changes over time) when the population becomes more and more immune. One can therefore not draw conclusions from these estimates about long-run values like $\bar{\rho}$. We return to this point when discussing the robustness of findings with respect to r after figure 3.7 and the effects of policy measures in figure 3.9.

[16] This percentage is from findings in Wuhan/ China, see Wu and McGoogan (2020).

does not test all individuals that display symptoms that might be caused by CoV-2, then the share r of individuals that are sick *and* reported is much lower than these 20%. We therefore set $r = 0.1$ and undertake robustness analyses further below.

- Model for predictions

After some intermediate steps (see app. C.2), and employing the fact that we can replace probabilities in our system (3.7) by expected numbers of individuals as discussed in footnote 13, we obtain our final theoretical structure,

$$\dot{N}_1(t) = -\frac{a}{r} N_1(t)^{1-\alpha} \left(N_2(t) + \eta N_4(t)\right)^{\beta} N^{\beta-\alpha} \left[\bar{\rho} - \frac{N_2(t) + N_4(t)}{N_1(t) + N_2(t) + N_4(t)}\right]^{\gamma}, \tag{3.15a}$$

$$\dot{N}_2(t) = aN_1(t)^{1-\alpha} \left(N_2(t) + \eta N_4(t)\right)^{\beta} N^{\beta-\alpha} \left[\bar{\rho} - \frac{N_2(t) + N_4(t)}{N_1(t) + N_2(t) + N_4(t)}\right]^{\gamma} - \left(\lambda_{23} + n_{\text{rec}}^{-1}\right) N_2(t), \tag{3.15b}$$

$$\dot{N}_3(t) = \lambda_{23} N_2(t), \tag{3.15c}$$

$$N_4(t) = N - N_1(t) - N_2(t) - N_3(t). \tag{3.15d}$$

This is the system we work with to compute predictions reported in what follows. Initial conditions for our solution are $N_2^{\text{ever}}(0) = N_{\text{observed}}^{\text{ever}}(0) = 16$ for 24 February 2020 (RKI, 2020), $N_3(0) = N_4(0) = 0$ and $N_1(0) = 83.100.000 - N_2^{\text{ever}}(0)$, where 83.1 Million is the population size in Germany (Statistisches Bundesamt, 2020).

3.4.3 The goodness of fit of the calibration

We can judge the quality of our calibration in two ways. First, we can check the fit for observed number of individuals that were ever reported to be sick, $N_2^{\text{ever}}(t)$ from (3.9). Second, we can check the quality of the match of the newly reported sick individuals, i.e. $N_2^{\text{new}}(t)$ from (3.10).

The left part of figure 3.4 shows the number of individuals that were ever reported to be sick. The observations from RKI are depicted as circles. It shows that our parameter choices are sufficiently good to use this model for projections into the future.[17] This impression is also confirmed from the right part of this figure. Obviously, fitting daily values is more demanding due to the greater variance in the data.

The figure also shows where, in the case of no public health measures, Germany would be heading to. Under the assumption that the share of sick *and* reported individuals out of the infected individuals is $r = 0.1$, the final number of reported sick individuals will be (right part of figure) 5.5 million. This level will be reached towards the end of June.[18] Hence, towards the end of the epidemic, at stable parameter values, around 7 in 100 inhabitants will have been sick (and reported).

[17] Appendix C.3 provides an enlarged version of this figure.

[18] This figure intuitively shows that in the medium-run, the long-run sickness share $r\bar{\rho}$ from (3.12) will be predictable from these observed time paths. At this point, there are too few observations and this is not yet feasible.

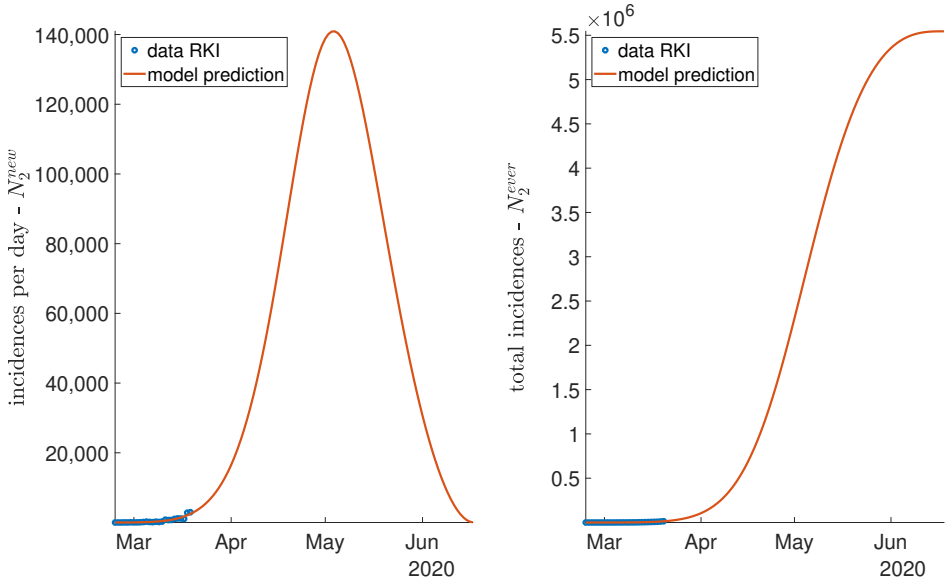


Figure 3.4: New incidences (reported) in the model ($N_2^{\text{new}}(t)$, curve) and in the data (dots) (left figure) and the number $N_2^{\text{ever}}(t)$ of sick in model and data (right figure) for the epidemic without public health measures

Note that this long run value does not only come out of the solution of our calibrated system (3.15). It can also be directly read from the long-run prediction in (3.12), which can be written as $\lim_{t\to\infty} N_2^{\text{ever}}(t)/N = r\bar{\rho}$. When we assume that two thirds of the population will be infected after the epidemic, $\bar{\rho} = 2/3$, and 20% turn sick of which the half is reported such that $r = .1$, we immediately see from (3.12) that the long run share is 6.7%, i.e. around 7 in 100.

3.5 Findings

3.5.1 The number of sick individuals and health care

The most important quantity coming out of our analysis is probably the number $N_2(t)$ of currently sick individuals. This number will decide whether hospital services will be enough for all individuals. The number of sick individuals rises strongly initially. This is the consequence of infections. As recovery sets in only after 2 weeks, the initial period sees this strong increase. Once the number of new sick cases reduces (as the number of individuals in state 1 decreases) and recovery sets in, the number of sick individuals falls again. The peak will be reached mid May. We note that this is a prediction for the case where individual contacts, as captured by the parameter a in (3.2), remain unchanged over time. If the policy measures in place as of 14 March have an effect, we would expect a slowdown in the rise of $N_2(t)$. This picture therefore shows what would have happened if no interventions had taken place. In this case, the number of sick (and reported) individuals on a given

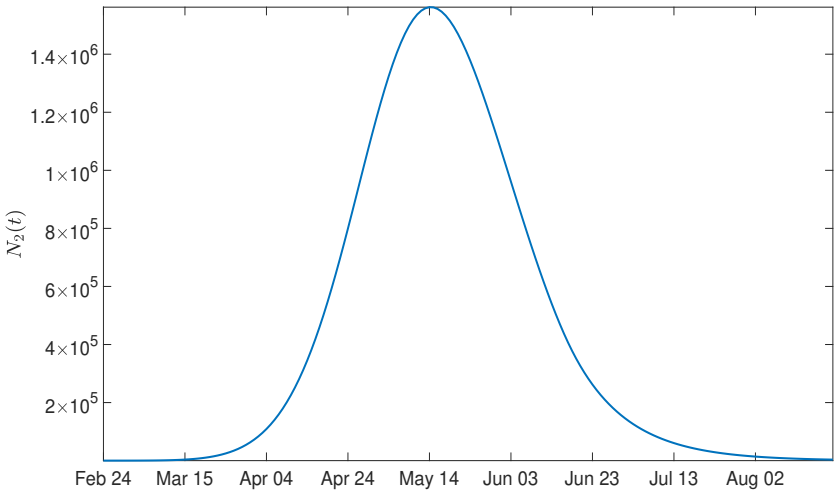


Figure 3.5: The number of sick individuals ($s = 2$) by day starting on 24 February 2020 in the absence of public health measures

day, taking recovery into account, will rise above 1 million end of April.

This figure also provides the maximum number of infected individuals over the epidemic. This allows to understand whether the number of extreme cases is always below the level that can be handled by the German health system. The peak of the number of sick individuals mid April is close to 220,000. According to the Deutsche Krankenhausgesellschaft (2020), there are 28,000 beds available in intensive care units (ICUs) in Germany. A fraction of these are currently available for the treatment of COVID-19 patients. In addition to that, an undefined number of hospital facilities allowing ventilation might be available in due time by reorganization activities. By contrast, a large fraction of COVID-19 patients might be treatable on regular wards. While it is therefore difficult to pin down the precise number of beds available for COVID-19 patients, a peak of more than 1 million on a given day is definitely beyond any level that can be handled by the health care system. The prevention of such a scenario is the key issue of all measures currently taken by federal and state governments. We therefore point to our assessment of theses measures further below.

If one would like to point out one advantage of this scenario without public health measures: the epidemic is fast. As the infection rate is very high, the outflow from the state of being healthy is fast. Hence, a large share of individuals, $1 - r = 90\%$, are infected without showing symptoms (or weak symptoms such that they are not reported) and they are immune (in state 4) quickly. Both figure 3.4 (left) and figure 3.5 show that the epidemic would be over end of June.

3.5.2 Individual risks

One of the concerns of individuals is that they might get infected. How large is this risk in a world without interventions?

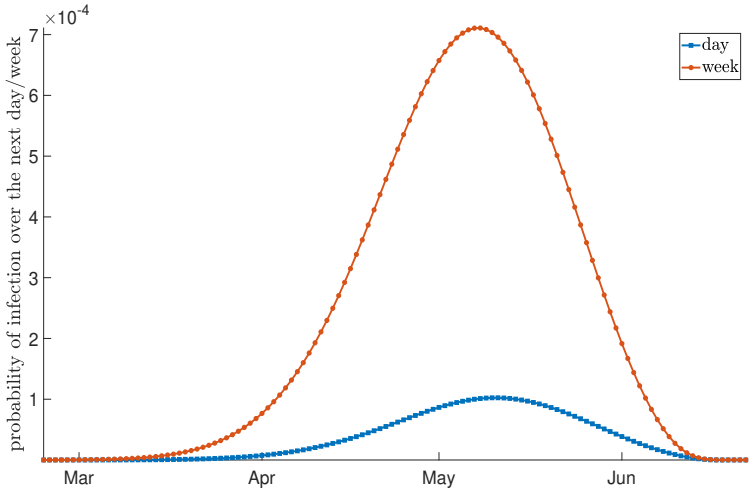


Figure 3.6: The probability per day (dotted) and per week (solid) to get sick in the absence of public health measures.

As this picture shows, the risk to get sick for each individual from (3.8) is very low throughout the entire epidemic. At the peak, the probability to get sick on a given day (red curve) is around 10^{-4}. In words, 1 person in 10,000 inhabitants turns sick per day. The probability to turn sick within a week (blue curve) is of course larger and lies at the peak at around 7 in 10,000 individuals. We emphasize again that this is the number of sick individuals that are actually reported to the RKI.

We compare the daily risk of 1 in 10,000 or the weekly risk of 7 in 10,000 with the long-run risk of 7 in 100 by the end of the epidemic to be or have been sick (as discussed after figure 3.4), it seems that at the individual level, the daily or even weekly risk of the epidemic is not really a major health risk.[19] It is clear, however, that at the level of society, the epidemic is a major issue. This difference between the individual risk and aggregate outcomes is the result of a classic negative externality. Even individuals that take into account the overall risk of 7 in 100 to turn sick might not be concerned. If the risk of infecting others was taken into account, individual behaviour would look differently. This is why governmental interventions have a clear justification from the perspective of market failure and public economics.

3.5.3 Robustness check

- The long-run sickness rate r

Given uncertainty about the share r of individuals that turn sick from an infection, we undertook robustness checks. Instead of our value of 0.1, we also solved our model for an r ten times lower up

[19] This might explain behaviour of some individuals who seem to ignore governmental pleas to reduce social interactions.

Table 3.3: Varying the long-run sickness rate r

r	a	α	β	γ	$\max\{N_2(t)\}$	$t^{\max}$	t_{1000}	t_{100}^{new}
0.01	1.03×10^{-4}	.685	.8289	.6952	182,004	23 April	15 July	19 May
0.025	1.282×10^{-4}	.6707	.8397	.6113	440,135	2 May	4 Aug	24 May
0.05	12.3×10^{-4}	.7505	.8513	.7921	775,478	8 May	25 Aug	25 June
0.1	3.02×10^{-5}	.5751	.8662	.6459	$1,56 \times 10^6$	14 May	7 Sep	17 June
0.2	2.21×10^{-5}	.5481	.8847	.7511	$2,9 \times 10^6$	19 May	27 Sep	16 June

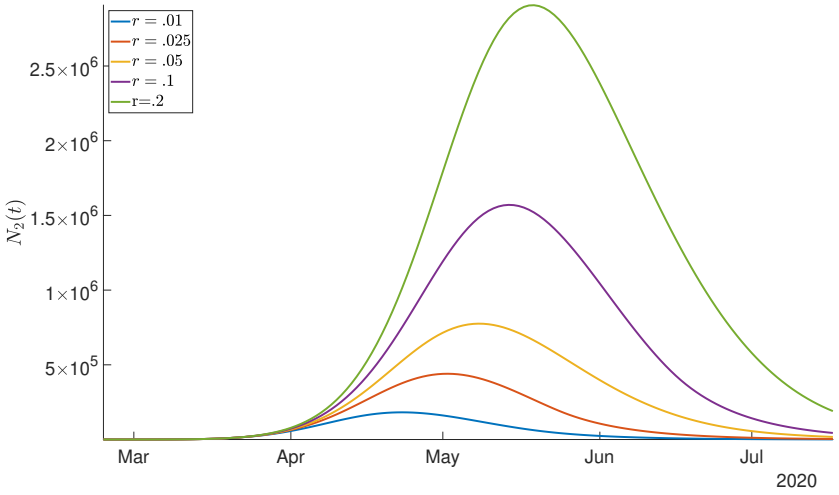


Figure 3.7: The number $N_2(t)$ of sick individuals as a function of assumptions on the probability r to get sick after an infection

to twice as large – as the left column of the next table shows. For each of these new values for r, we recalibrate the values for a and α to match the initial observations of sick individuals.

The model predictions are very reasonable from a qualitative perspective. The larger the individual risk r to get sick, the higher the number $N_2(t)$ of simultaneously sick individuals and the later the peak $t^{\max}$ of the epidemic.

We also find, as was to be expected, that a larger r implies a longer duration of the epidemic. If we define the end of the epidemic by the day t_{100}^{new} where only 100 new incidences are reported per day, we see that the end of the epidemic lies between 19 May and 25 June. Interestingly, the end of the epidemic is not monotonic in r. When the end of the epidemic is defined by the point in time t_{1000} where 1000 sick individuals are left ($N(t_{1000}) = 1000$), the epidemic ends between 15 July and 27 September. Note that the t_{1000} definition implies much later endpoints (of one and the same epidemic) than the t_{100}^{new} definition.

Why are these numbers important from a public health perspective? The data section shows that Hubei and South Korea seem to have managed to keep the long-run ratio $r\bar{\rho}$ from (3.12) at 1 in 1000 (Hubei) or 1 in 5000 (South Korea). If we believe that these values are long-run values, then either the individual probability r to get sick after infection of the long-run infection rate $\bar{\rho}$ must be much lower than the standard values shown in 3.2 or sickness probabilities r must be lower.
As no information is available at this point about the share of individuals that are already immune to CoV-2 in Germany, one might well hope that Germany is heading towards a Hubei long-run equilibrium. As data quality was often argued to be an issue, our long-run ratio $r\bar{\rho}$ with $r = 0.01$ would correspond to 6.7 in 1000 individuals that were sick at some point in the long-run (N_2^{ever}). This would imply that – in the case of no public intervention – the peak in the number of sick individuals would only be around 180 thousand individuals (as the first row of table 3.3 shows). The blue time path of $N_2(t)$ of figure 3.7 would be relevant.

- Infections via individuals without symptoms

We also inquired into the effect of η, i.e. into the share of healthy individuals that can infect other individuals. The value of η in table 3.2 is 0.4 and we now varied it between .25 and .75. While there were qualitative changes (figures are available upon request), there are no quantitative change that would be of interest for any public health questions.

3.6 The effect of policy measures

Given that an undamped spread of CoV-2 in society leads to high reported incidences $N_2(t)$, public health measures are attempting to reduce social contacts. The fewer social contacts, the lower infections per day. These contacts are captured by the parameter a in (3.2). Now assume the number of contacts per individual is restricted in some exogenous manner. Individuals are not allowed to watch soccer matches, go to conferences or move around freely in their region or country. What is the effect of these measures? We also ask whether it makes a difference whether such measures are taken at the beginning of the outbreak of infections or when the measures are delayed for 1 month.

3.6.1 The experiments

We model policy measures by assuming a time path for a. The first major policy measure, 'shut down' for simplicity, took place as of 14 March 2020. Sports events from professional to amateur leagues and schools were closed. This measure is planned to continue until 19 April 2020. As our quantitative analysis starts on 24 February, we employ parameters from table 3.2 for the unrestricted epidemic. We especially keep the value for the contact intensity a of individuals. After 20 days, we reduce a to a lower level a^{low} for the duration of this policy measure. After 37 days, i.e. on 20 April,

we return to the original level, assuming that sports events and schooling resumes. Formally,

$$a(t) = \left\{ \begin{array}{c} a \\ a^{\text{low}} = a/2 \\ a \end{array} \right\} \text{for} \left\{ \begin{array}{c} t < 20 \text{ days} \\ 20 \text{ days} \leq t \leq 57 \text{ days} \\ t > 57 \text{ days} \end{array} \right\} \tag{3.16}$$

captures the shut down initiated on 13 March. For all shut down scenarios we assume that a^{low} is 50% lower compared to a.

To understand whether a later intervention is more effective, we counterfactually assume that the intervention starts 32 days later. Here we set

$$a(t) = \left\{ \begin{array}{c} a \\ a^{\text{low}} \\ a \end{array} \right\} \text{for} \left\{ \begin{array}{c} t < 52 \\ 52 \leq t \leq 89 \\ t > 89 \end{array} \right\}. \tag{3.17}$$

We call this the 'delayed shut down'.

Finally, we investigate into the effect of extending the shut down. We extend the period with lower social contact rate a^{low} to last for 74 days, i.e.

$$a(t) = \left\{ \begin{array}{c} a \\ a^{\text{low}} \\ a \end{array} \right\} \text{for} \left\{ \begin{array}{c} t < 20 \\ 20 \leq t \leq 94 \\ t > 94 \end{array} \right\}. \tag{3.18}$$

This is our 'extended shut down'.

3.6.2 The quantitative effects of policy measures

- Baseline analysis

For this baseline analysis, we do not recalibrate the model as a in our scenarios changes after T, the day providing the last observation used for calibration. Parameter values are therefore given by the values in table 3.2. The effect of various forms of shut downs are captured by a drop of the contact rate a by 50% to a^{low}, as shown in (3.16).

The results are shown in the next figure. Our main variable for judging policy measures is the number $N_2(t)$ of individuals that are simultaneously sick.

First, the policy is effective. Reducing contacts reduces the speed of infections and new cases of sickness – but also the transition from healthy in state 1 to healthy (with infection) in state 4. The case without shut down is illustrated by the red curve in this figure. It is identical to the curve in figure 3.5. The shut down from (3.16) which reflects the planned closing of schools in Germany until mid April is drawn as the green curve. It reduces the rise of the number of sick individuals until mid April. Afterwards, the number of sick individuals will rise more quickly again. The peak is reached in June and this peak is lower than the one without the shut down. In this respect the shut down is a

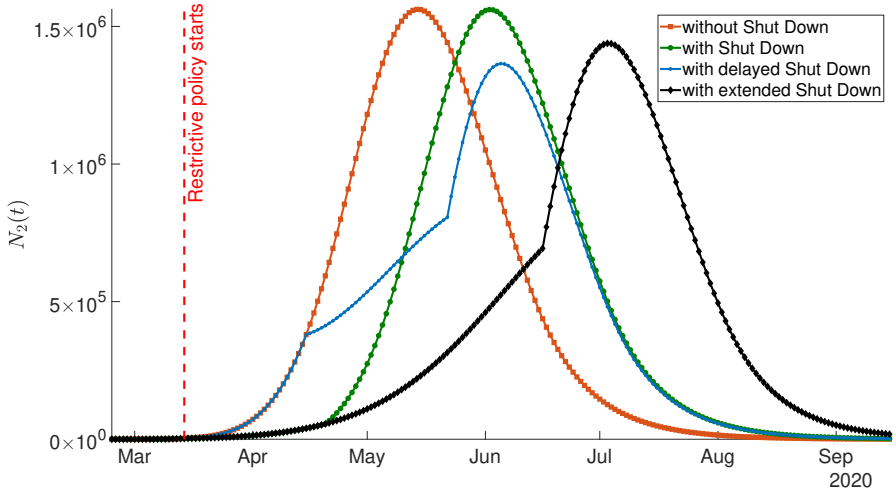


Figure 3.8: The effect of a shut down

good policy measure. When the shut down is extended (the black curve), the peak is further shifted to the right and also smaller.[20]

Quantitatively speaking, the shut down prolongs the epidemic in terms of its end as defined by t_{100}^{new} also employed in table 3.3 (100 or fewer new incidences are reported per day). For the shut down currently in place and for the delayed shut down, the end is beginning of July. For the longer shut down, the end is beginning of August. Concerning the peak demands on the health system, the shut down as currently in place, and also an extended shut down, hardly have any effect. There is a certain delay – which helps the health sector to prepare for higher numbers in the future. But the level of the peak, as the figure shows, hardly changes. This looks like a bleak outlook.

To be fair, we should stress at this point that the assumption that a shut down reduces the contact rate and thereby the infection rate by 50% is open to empirical investigation. At this point, this is a model assumption for which empirical evidence is still lacking. Hartl et al. (2020) show that statistically convincing evidence will be available at the earliest beginning of April. It seems fair to say, however, that it can hardly be expected at this point that current measures reduce the infection rate by 50%. Hence, the above evaluation of the shut down might even be too optimistic.[21]

- Is a delayed shut down a good idea?

Interestingly, a delayed shut down (blue curve) is almost as good as an extended shut down and in

[20] If we had a sequence of shut downs with breaks in between, it would qualitatively look like a combination of the green and the blue curve. The green curve represents a shut down that ends end of April where the green curve starts rising quickly again. Then when a new shut down takes place as of mid May, the number of sick individuals would follow the blue curve. When the "blue shut down" ends end of May, the number of sick individuals starts rising again as shown by the blue curve.

[21] All of our scenarios could look much more optimistic if we had a vaccine. To the best of our knowledge, this will not be available over the next 6 months.

some respect is even better: A delayed shut down reduces the peak in the number of sick individuals. Hospitals are therefore less crowded with a delayed shut down as opposed to an immediate shutdown. This implies, more generally speaking, that there is an optimal timing for a shut down.

Why is there an optimal point for a shut down? Imagine the shut down takes place early (our curve). Then the number of sick individuals does not increase that quickly. The downside is that, once the shut down is over, there are a lot of healthy individuals left that can turn sick. The number of new incidences jumps up and the peak is reached quickly. When the shut down takes place later (the blue curve), the number of sick individuals is larger at the moment the shut down starts. Once the shut down is over, the number of healthy individuals that are left to turn sick is not as large as in the case of an early shut down. As a consequence, the peak is less high.

In a perfect world with complete knowledge about the number of sick individuals, the long run infection rate $\bar{\rho}$, the individual risk r to turn sick from an infection and other parameters employed in this model, policy makers should therefore choose the optimal point in time for an intervention. In the world we live in, this information is not available, neither from a medical, nor from a statistical and data processing perspective. The finding nevertheless makes the strong point that policy interventions should be conditioned on the current situation in a country or region. Ideas that policy measures must be the same for all regions in Germany at all points in time are strongly contradicted.[22]

- The policy effects with lower $\bar{\rho}$

Let us now return to the uncertainty about the value of the long-run share $\bar{\rho}$ of infected individuals. We varied r in table 3.3 such that we could also study a "data-quality adjusted Hubei-scenario" where $r\bar{\rho}$ implies that there are 6.7 in 1000 (ever) sick individuals. This was discussed after figure 3.7. We now study changes in $\bar{\rho}$ at unchanged $r = 0.1$ such that we end up at $r\bar{\rho} = 6.7/1000$. We therefore reduce $\bar{\rho}$ by a factor of 10. We do recalibrate our model for this scenario as $\bar{\rho}$ affects the calibration outcome. The fit is similarly good to the one discussed after figure 3.4. Given the new parameter values, we change the contact rate according to the shut down (3.16) and the extended shut down (3.18).

Let us start with the long-run effects. As we know from (3.12), the long-run effects with $\bar{\rho}$ reduced by a factor of 10 are 10 times lower. This is visible in the right part of figure 3.9 displaying $N_2^{\text{ever}}(t)$. Once the epidemic is over, there will be 550 thousand individuals that were sick at some point, one tenth of the level in figure 3.4. This is again independent of policy measures as the latter do not affect $r\bar{\rho}$.

Concerning short-run effect, we see that in the case of no intervention (or interventions that do not have an effect), the peak comes earlier for $N_2(t)$. The red curve in figure 3.9 peaks in mid April in contrast to early to mid May for the higher long-run infection rate of figure 3.8. The effect of the

[22] We understand that from a policy perspective it is simpler to "sell a measure" that is the same for everybody than measures that differ across individuals. Yet, treating unequals as equals has rarely been a convincing idea.

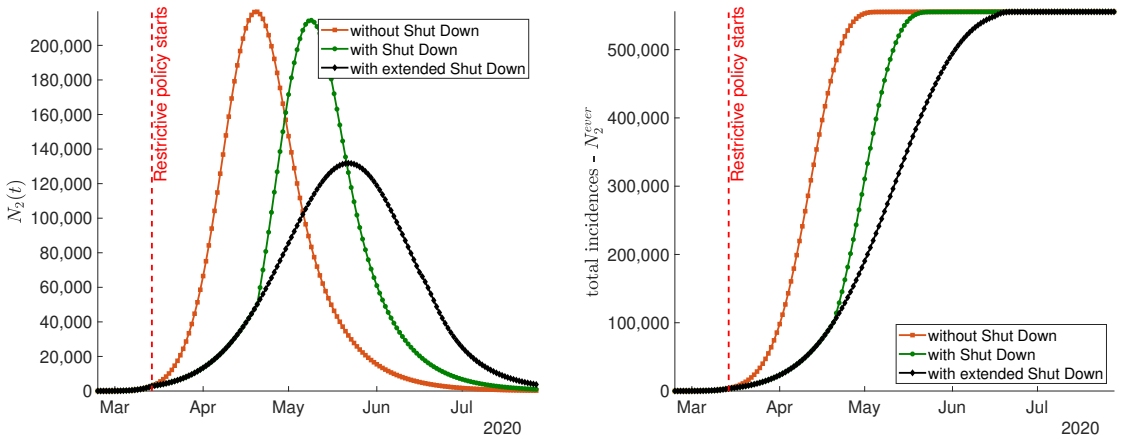


Figure 3.9: The effect of changes in the long-run infection rate $\bar{\rho}$

shut down currently in place looks qualitatively similar to the effect when $\bar{\rho}$ is higher. By contrast, the extended shut down in figure 3.9 does not display a strong rise in growth rates of $N_2(t)$ once it is over. It is directly followed by the decline in the number of sick individuals.

Most importantly, however, theses two figures display a strong quantitative difference. If the long run infection rate in Germany is in the "data-quality adjusted Hubei range", the peaks are at a much lower level. In the case without policy measures and in the case where the current shut down (3.16) is the only policy measure, the peaks of the simultaneously sick number of individuals is just above 200 thousand. This strongly contrasts to earlier values above 1 million. Another difference consists in the quantitative difference here between the extended shut down and the shut down. In the scenario with a lower long-run infection rate, an extended shut down implies that the peak of $N_2(t)$ will be much lower than the peak of the normal shut down. This was not the case in figure 3.8. This points to the fact that shut downs need to be sufficiently strong relative to the overall potential number of sick individuals.

When the long-run infection rate is in this Hubei-range, then the epidemic is also over more quickly than in the case of an infection rate of two thirds. In any case, the epidemic will last until June, in the case of extended shut downs (or other measures) it can even last until mid July.

3.7 Conclusion

We have developed a model that allows to study the spread of an infection in a society. We have solved this model and calibrated it to Germany. We employ the observed number of reported infections to match the initial increase of the number of sick individuals. In addition, we employ parameter values from the medical literature to quantify e.g. long-run infection rates or individual risks to turn sick after an infection. Given uncertainty about the precision of these parameter values, we undertook

robustness checks with respect to this long-run measure.

When we studied the effect of policy measures, we found that any intervention that reduces the contact rate of individuals reduces the strength of the epidemic. Both the peak of the number of simultaneously sick individuals as well as the individual sickness risk go down. This decrease is stronger, the longer the policy measure lasts.

We also find, however, that there is an optimal point in time when an intervention, effective for a fixed length of time (say, 3 weeks), should start. When the intervention starts too early, there are "too many" healthy individuals left at the end of the intervention. The rise in the number of sick individuals would be accelerated. When the intervention starts later, the number of sick individuals is larger initially. Once the intervention is over, there are, as a consequence, not so many healthy individuals left and the peak is smaller. While implementing optimal schemes for real world societies probably requires too much information, this finding points towards implementing policy measures which are contingent on the current situation (number of sick individuals per capita) of a region or country.

In the long-run, for given parameters, the number of sick individuals is not affected by policy measures when the latter only reduce contact rates of individuals. All findings are subject to large uncertainty given little knowledge about long-run infection rates following a CoV-2 epidemic and individual sickness rates. It seems to be certain, however, that the epidemic will last at least until July, given the public health measures in place. Even in the case of a good scenario with extremely low long-run infection rates, the peak of the number of simultaneously sick individuals will lie above 200 thousand individuals.

A lot of extensions of this framework are worth being undertaken. First, optimal behaviour of individuals could be taken into account. The activity parameter a is like the outcome of decisions of individuals on how much to work and how much to enjoy social interactions. One can imagine a trade-off between reducing the risk of getting infected by lowering a and working fewer hours and thereby experiencing lower labour income. The same trade-off is present at the aggregate level to be tackled by a government. Such a trade-off might also include behavioural features like anxiety as in Caplin and Leahy (2001), potentially in a simplified way as presented in Wälde and Moors (2017). These extensions will help to better understand the challenges posed by this pandemic.

A more detailed modelling of the long-run infection rate would also be useful. The path of social contacts over time should have an influence on the share of infected individuals in the long-run. A more detailed matching mechanism than the one employed here – borrowed in spirit from search and matching models – should be developed.

Finally, the calibration process employed here can be extended to full structural estimation. Further, the effects of the policy measures in place since 14 March can be estimated or calibrated. The calibration in this paper employs data up to the point where measures of 14 March should start having an effect. As of end of March, the intervention regime discussed above and the corresponding

(reduced) contact rate could be quantified. This would provide a good idea of the effects of these policy measures.

C.1 Data sources

- Data sources and calculations

The data used in this note is taken from two sources: (i) the World Health Organisation's Situation reports on the progress of COVID-19, and (ii) the dataset put together by the nCoV-2019 Data Working Group (Dong et al. (2020)). In particular, wherever available the data from the nCoV-2019 Group is used instead of the data from the WHO, in an effort to remain as consistent as possible by using only one source at a time. Data from the nCoV-2019 Group is used as it provides a centralised and standardised repository of multiple official sources.
The figures showing the number of new confirmed cases at a date t were produced by computing the difference between the cumulative total number of confirmed cases at t minus the cumulative total number of confirmed cases at $t-1$, i.e.

$$\text{new cases}_t = \text{total cases}_t - \text{total cases}_{t-1}.$$

As detailed in the Table 3.1, the infection rate is computed by dividing the cumulative total number of confirmed cases as of the time of writing by the population of the corresponding country in 2019. Meanwhile, the risk factor is simply the inverse of that fraction and is given by the total population divided by the cumulative total number of confirmed cases, i.e.

$$\text{Infection Rate in } \% = \frac{\text{total cases}}{\text{population}_{2019}} \times 100 \Leftrightarrow \text{Risk of infection} = \frac{\text{population}_{2019}}{\text{total cases}}.$$

C.2 Intermediate steps for (3.15)

Recall (3.2) and write it as

$$\begin{aligned}\lambda_{12}(t) &= aN_1(t)^{-\alpha}\left(N_2(t)+\eta N_4(t)\right)^{\beta}\left[\bar{\rho}-\frac{N_2(t)+N_4(t)}{N_1(t)+N_2(t)+N_4(t)}\right]^{\gamma}\\ &= a\left(N_1(t)\frac{N}{N}\right)^{-\alpha}\left(N_2(t)\frac{N}{N}+\eta N_4(t)\frac{N}{N}\right)^{\beta}\left[\bar{\rho}-\frac{p_2(t)+p_4(t)}{p_1(t)+p_2(t)+p_4(t)}\right]^{\gamma}\end{aligned}$$

where the fraction in squared brackets was multiplied by N/N as well and we employed $p_s(t) = N_s(t)/N$. Applying this to term one and two as well,

$$\lambda_{12}(t) = ap_1(t)^{-\alpha}\left(p_2(t)+\eta p_4(t)\right)^{\beta}N^{\beta-\alpha}\left[\bar{\rho}-\frac{p_2(t)+p_4(t)}{p_1(t)+p_2(t)+p_4(t)}\right]^{\gamma}.$$

If we want homogeneity of degree 0, we would set $\alpha = \beta$. Employing the transition rate λ_{14} from (3.4) and taking λ_{24} from (3.5) into account, our model in (3.7) becomes

$$\dot{p}_1(t) = -\left(\lambda_{12}(t)+\frac{1-r}{r}\lambda_{12}(t)\right)p_1(t)$$

$$= -\frac{\lambda_{12}(t)}{r} p_1(t) \tag{3.19}$$

$$= -\frac{a}{r} p_1(t)^{-\alpha} \left(p_2(t) + \eta p_4(t)\right)^{\beta} N^{\beta-\alpha} \left[\bar{\rho} - \frac{p_2(t) + p_4(t)}{p_1(t) + p_2(t) + p_4(t)}\right]^{\gamma} p_1(t) \tag{3.20}$$

$$\dot{p}_2(t) = \lambda_{12}(t)\, p_1(t) - \left(\lambda_{23} + n_{\text{rec}}^{-1}\right) p_2(t), \tag{3.21}$$

$$= a p_1(t)^{-\alpha} \left(p_2(t) + \eta p_4(t)\right)^{\beta} N^{\beta-\alpha} \left[\bar{\rho} - \frac{p_2(t) + p_4(t)}{p_1(t) + p_2(t) + p_4(t)}\right]^{\gamma} p_1(t) - \left(\lambda_{23} + n_{\text{rec}}^{-1}\right) p_2(t) \tag{3.22}$$

$$\dot{p}_3(t) = \lambda_{23} p_2(t) \tag{3.23}$$

Simplifying, we get (3.15) in the main text.

C.3 Goodness of fit

We provide a graphical illustration of how well our parameter choice via the minimization procedure in (3.14) fits the data.

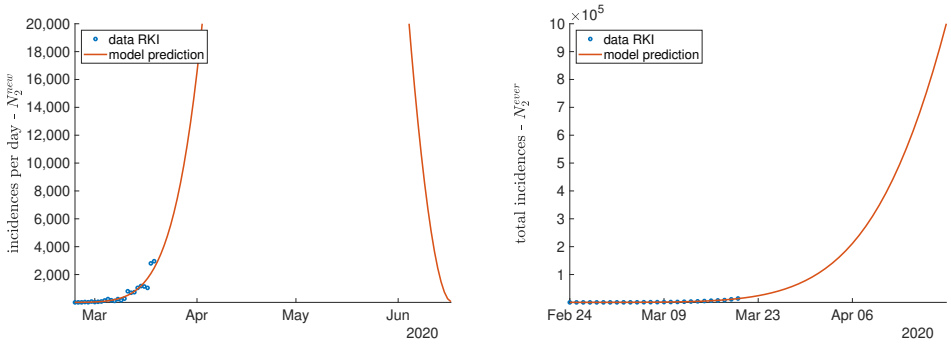


Figure 3.10: Enlarged version of figure 3.4 with data and model fit

Bibliography Chapter 3

Achdou, Y., Han, J., Lasry, J., Lions, P., and Moll, B. (2020). Income and wealth distribution in macroeconomics: A continuous-time approach. *mimeo Princeton University*.

Akbarpour, M. and Jackson, M. (2018). Diffusion in networks and the virtue of burstiness. *PNAS*, 115(30):1–9.

an der Heiden, M. and Buchholz, U. (2020). Modellierung von beispielszenarien der sars-cov-2-epidemie 2020 in deutschland. *DOI 10.25646/6571.2*, pages 1–11.

Anderson, R., Anderson, B., and May, R. (1992). *Infectious diseases of humans: dynamics and control*. Oxford university press.

Atkeson, A. G. (2020). What will be the economic impact of covid-19 in the us? rough estimates of disease scenarios. *NBER working paper 26867*.

Balcan, D., Goncalves, B., Hu, H., Ramasco, J., Colizza, V., and Vespignani, A. (2010). Modeling the spatial spread of infectious diseases: The global epidemic and mobility computational model. *Journal of Computational Science*, 1(3):132–145.

Barro, R., Ursua, J., and Wenig, J. (2020). The coronavirus and the great influenza epidemic lessons from the 'spanish flu' for the coronavirus's potential effects on mortality and economic activity. *CESifo Working Paper 8166, https://www.cesifo.org/en/publikationen/2020/working-paper/coronavirus-and-great-influenza-epidemic-lessons-spanish-flu*.

Bayer, C., Rendall, A. D., and Wälde, K. (2019). The invariant distribution of wealth and employment status in a small open economy with precautionary savings. *Journal of Mathematical Economics*, 85:17–37.

Bayer, C. and Wälde, K. (2010). Matching and saving in continuous time: Proofs. *CESifo Working Paper 3026-A*.

Bundesamt, S. (2020). Bevölkerungsstand. *https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Bevoelkerungsstand*.

Caplin, A. and Leahy, J. (2001). Psychological expected utility theory and anticipatory feelings. *Quarterly Journal of Economics*, 116(1):55–79.

Chinazzi, M., Davis, J., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., Pastore y Piontti, A., Mu, K., Rossi, L., Sun, K., Viboud, C., Xiong, X., Yu, H., Yu, H., Halloran, E., Longini Jr., I., and Vespignani, A. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak. *Science*, science.aba9757/DC1:1–17.

Deutsche Gesellschaft für Epidemiologie (2020). Stellungnahme der deutschen gesellschaft für epidemiologie (dgepi) zur verbreitung des neuen coronavirus (sars-cov-2). *Deutsche Gesellschaft für Epidemiologie*, pages 1–4.

Diamond, P. A. (1982). Aggregate demand management in search equilibrium. *Journal of Political Economy*, 90:881–94.

Dong, E., Du, H., and Gardner, L. (2020). An interactive web-based dashboard to track covid-19 in real-time. *The Lancet Infectious Diseases, https://doi.org/10.1016/S1473-3099(20)30120-1 ¡https://doi.org/10.1016/S1473-3099(20)30120-1¿*, pages 1–2.

Dorn, F., Fuest, C., Göttert, M., Krolage, C., Lautenbacher, S., Link, S., Peichl, A., Reif, M., Sauer, S., Stöckli, M., Wohlrabe, K., and Wollmershäuser, T. (2020). Die volkswirtschaftlichen kosten des corona-shutdown für deutschland: Eine szenarienrechnung. *ifo Schnelldienst*, 73(4).

Eichenbaum, M., Rebelo, S., and Trabandt, M. (2020). The macroeconomics of epidemics. *NBER Working Paper Series*, Working Paper 26882.

Ferguson, N., Laydon, D., Nedjati-Gilani, G., Imai, N., Ainslie, K., Baguelin, M., Bhatia, S., Boonyasiri, A., Cucunuba, Z., Cuomo-Dannenburg, G., Dighe, A., Dorigatti, I., Fu, H., Gaythorpe, K., Green, W., Hamlet, A., Hinsley, W., Okell, L., van Elsland, S., Thompson, H., Verity, R., Volz, E., Wang, H., Wang, Y., Walker, P. G., Walters, C., Winskill, P., Whittaker, C., Donnelly, C., Riley, S., and Ghani, A. (2020). Impact of non-pharmaceutical interventions (npis) to reduce covid19 mortality and healthcare demand. *Imperial College COVID-19 Response Team*, pages 1–20.

Guan, W., Ni, Z., Hu, Y., Liang, W., Ou, C., He, J., Liu, L., Shan, H., Lei, C., Hui, d., Du, B., Li, L., Zeng, G., Yuen, K.-Y., Chen, R., Tang, C., Wang, T., Chen, P., Xiang, J., Li, S., Wang, J.-l., Liang, Z., Peng, Y., Wei, L., Hu, Y.-h., Peng, P., Wang, J.-m., Liu, J., Chen, Z., Li, Z., Zheng, Z., Qiu, S., Luo, J., Ye, C., Zhu, S., and Zhong, N. (2020). Clinical characteristics of coronavirus disease 2019 in china. *The New England Journal of Medicine*, doi: https://doi.org/10.1101/2020.02.06.20020974:1–12.

Hartl, T., Wälde, K., and Weber, E. (2020). Bundesliga, schulen, kinos - wann wissen wir, ob die maßnahmen gegen corona etwas gebracht haben? *https://www.macro.economics.uni-mainz.de/corona-blog/, Einschätzung vom 23. März.*

Hellewell, J., Abbott, S., Gimma, A., Bosse, N., Jarvis, C., Russell, T., Munday, J., Kucharski, A., Edmunds, W., Funk, S., and Eggo, R. (2020). Feasibility of controlling covid-19 outbreaks by isolation of cases and contacts. *The Lancet Global Health*, 8:e488–e496.

Hethcote, H. (2000). The mathematics of infectious diseases. *SIAM Review*, 42(4):559–653.

Kaplan, G., Moll, B., and Violante, G. L. (2018). Monetary policy according to hank. *American Economic Review*, 108:697–743.

Khieu, H. and Wälde, K. (2020). Capital income risk and the dynamics of the wealth distribution. *mimeo Johannes Gutenberg University Mainz.*

Krankenhausgesellschaft, D. (2020). Coronavirus: Fakten und infos. *https://www.dkgev.de/dkg/coronavirus-fakten-und-infos/.*

Lachmann, A. (2020). Correcting under-reported covid-19 case numbers. *medRxiv preprint, https://doi.org/10.1101/2020.03.14.20036178*, pages 1–5.

Lauer, S., Grantz, K., Bi, Q., Jones, F., Zheng, Q., Meredith, H., Azman, A., Reich, N., and Lessler, J. (2020). The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: Estimation and application. *Annals of Internal Medicine*, doi:10.7326/M20-0504:1–7.

Linton, N., Kobayashi, T., Yang, Y., Hayashi, K., Akhmetzhanov, A. R., Jung, S.-m., Yuan, B., Kinoshita, R., and Nishiura, H. (2020). Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: A statistical analysis of publicly available case data. *Journal of Clinical Medicine*, 9(2)(538):1–9.

Mortensen, D. T. (1982). Property rights and efficiency in mating, racing, and related games. *American Economic Review*, 72:968–79.

Nishiura, H., Kobayashi, T., Suzuki, A., Jung, S.-M., Hayashi, K., Kinoshita, R., Yang, Y., Yuan, B., Akhmetzhanov, A., and Linton, N. (2020). Estimation of the asymptomatic ratio of novel coronavirus infections (covid-19). *International Journal of Infectious Diseases*, Journal Pre-proof, https://doi.org/10.1016/j.ijid.2020.03.020:1–7.

Organization, W. H. (2020). Coronavirus disease 2019 (covid-19) situation report - 46. *https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200306-sitrep-46-covid-19.pdf?sfvrsn=96b04adf_4*.

Pissarides, C. A. (1985). Short-run equilibrium dynamics of unemployment vacancies, and real wages. *American Economic Review*, 75:676–90.

Riou, J. and Althaus, C. (2020). Pattern of early human-to-human transmission of wuhan 2019 novel coronavirus (2019-ncov), december 2019 to january 2020. *Eurosurveillance*, 25(4).

Robert Koch Institut (RKI) (2020). Covid-19: Fallzahlen in deutschland und weltweit. *https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Fallzahlen.html*.

Ross, S. M. (1993). *Introduction to Probability Models, 5th edition*. Academic Press, San Diego.

Simon, P. (2020). Deutschland, welches horrorszenario willst du? man sollte die katastrophe vom ende her berechnen! *mimeo University Mainz*, pages 1–3.

Wilder-Smith, A., Chiew, C., and Lee, V. (2020). Can we contain the covid-19 outbreak with the same measures as for sars? *Lancet Infect Dis*.

Wu, Z. and McGoogan, J. M. (2020). Characteristics of and important lessons from the coronavirus disease 2019 (covid-19) outbreak in china: summary of a report of 72-314 cases from the chinese center for disease control and prevention. *JAMA, DOI:10.1001/jama.2020.2648*.

Wälde, K. and Moors, A. (2017). Current emotion research in economics. *Emotion Review*, 9:271–278.

Ying, Y., Mo, P., Xiao, Y., Zhao, O., Zhang, Y., and Wang, F. (2020). Post-discharge surveillance and positive virus detection in two medical staff recovered from coronavirus disease 2019 (covid-19), china, january to february 2020. *Euro Surveill., https://doi.org/10.2807/1560-7917*, 25(10):1–4.

Appendices

Appendix D

General Appendix

D.1 Abstract

Chapter 1

The first essay of this dissertation, *A general equilibrium model of earnings, income, and wealth heterogeneity in continuous time*, a joint work with Frédéric Enrico Haller and Marc Puche Niubó, confronts the challenge of jointly explaining the empirically observed distributions of earnings, income, and wealth within a unified general equilibrium framework. Prior studies, particularly those based on standard Bewley-Huggett-Aiyagari models, often struggle to replicate the pronounced concentration of wealth at the upper tail without resorting to ad-hoc or empirically questionable mechanisms like "superstar" income states or highly persistent shocks. Simultaneously matching the distinct distributional shapes of earnings and income further complicates the task. This essay proposes an extension of the standard heterogeneous-agent framework, cast entirely in continuous time, specifically designed to address these shortcomings and provide a more integrated understanding of the sources and interactions of inequality along these three key dimensions, as called for by empirical observations, e.g., in Quadrini and Rıos-Rull (1997).

Methodologically, the essay leverages the analytical tractability and computational advantages inherent in continuous-time modeling. It builds substantially upon the numerical framework developed by Achdou et al. (2022) for solving single-state continuous-time models using finite-difference methods (specifically the upwind scheme for stability). The key innovation here is the extension and adaptation of this numerical apparatus to handle a multi-dimensional continuous state space, incorporating not only household wealth but also continuous stochastic processes for both individual labor earnings and idiosyncratic returns on capital. This dual source of continuous risk, drawing inspiration from the work of Benhabib et al. (2011) and Benhabib et al. (2015) on wealth tails through return heterogeneity, is crucial for generating the necessary dynamics. The entire system is solved within a

general equilibrium context where aggregate factor prices (the mean wage level w and the risk-free rate component θ) are determined endogenously by market clearing conditions, requiring an iterative solution approach coupling the HJB and Fokker-Planck equations with the aggregate capital derived from the stationary distribution.

Upon calibration to match key moments of U.S. microdata, the model demonstrates considerable empirical success. It successfully replicates the observed Gini coefficients for earnings, income, and wealth and captures significant portions of their distributional shapes. Crucially, there is no need for "superstar" states. The results underscore the distinct roles of the two stochastic drivers: the continuous earnings process is essential for matching the lower parts of the wealth distribution and ensuring numerical stability, while the heterogeneous interest rate process is paramount for generating the thick upper tail of the wealth distribution – creating a "Goldilocks zone" where high wealth agents with high returns maintain substantial saving incentives. A policy experiment investigating an increase in unemployment benefits yields the counterintuitive result of increased steady-state wealth inequality, driven by a reduction in precautionary savings among lower-income individuals. The primary contribution lies in providing both a novel theoretical framework and a viable computational methodology for jointly analyzing earnings, income, and wealth inequality in a fully specified continuous-time general equilibrium model.

Chapter 2

The second essay, entitled *Sequence-space Jacobians in continuous time*, addresses a critical computational bottleneck hindering the analysis of dynamic heterogeneous-agent models: the calculation of transition paths and impulse responses following aggregate shocks. While steady-state analysis is relatively standard, computing how the distribution of agents and aggregate variables evolves over time after a shock is often computationally prohibitive, especially in continuous time, where traditional methods typically involve iterating on infinite-dimensional objects (like the value function or distribution) over time. This chapter introduces a powerful and highly efficient alternative by adapting the sequence-space Jacobian (SSJ) methodology, which was pioneered by Auclert et al. (2021) for discrete-time models, to the continuous-time setting. The SSJ approach linearizes the model's response around the steady state with respect to sequences of aggregate inputs, providing sufficient statistics (Jacobians) that map shocks to outcomes without simulating the full non-linear model forward.

The core methodological advance presented here is the demonstration that the infinitesimal generator of the relevant Hamilton-Jacobi-Bellman (HJB) or Fokker-Planck equation can be directly utilized to efficiently compute the key expectation vectors needed for constructing the sequence-space Jacobians in continuous time. Doing so avoids the need for explicit time iteration or the simulation of individual agents' responses, which underlies the computational burden of alternative continuous-time methods

(e.g., Achdou et al. (2022)). By operating in the much lower-dimensional sequence space (mapping sequences of aggregate shocks to sequences of aggregate outcomes) rather than the high-dimensional agent state space, the method's computational cost becomes largely independent of the state space's size. The essay demonstrates the application to both a neoclassical Krusell-Smith style model and a one-asset HANK model, showing its compatibility with different types of idiosyncratic shock processes (Poisson and diffusion). Furthermore, it utilizes a directed acyclical graph (DAG) representation to systematically compute general equilibrium Jacobians from their partial equilibrium counterparts.

The principal finding is a dramatic increase in computational speed. For the examples studied, calculating transition dynamics using the continuous-time SSJ method is shown to be orders of magnitude faster (between 30 and 1,500 times, depending on the specific comparison) than conventional continuous-time iterative approaches, once the steady state has been computed. This efficiency gain transforms the analysis of dynamic responses in continuous-time heterogeneous-agent models from a computationally intensive task to a near-instantaneous one (often fractions of a second on standard hardware like a MacBook). This advance opens the door to applications previously considered infeasible, such as likelihood-based estimation or the analysis of models with very large state spaces or complex dynamics. The contribution lies in bridging the gap between the powerful SSJ methodology and the analytical advantages of continuous-time modeling, providing a readily applicable and highly efficient toolkit for macroeconomists working with heterogeneous-agent models.

Chapter 3

The third essay, *Projecting the spread of COVID-19 for Germany*, a joint work with Jean Roch Donsimoni, Bodo Plachter, and Klaus Wälde, illustrates the broad applicability of continuous-time stochastic modeling techniques. Typically employed in economics, they are adapted here to address a pressing real-world problem: the forecasting and policy analysis of the COVID-19 pandemic in Germany during its early stages. The chapter adapts a standard epidemiological SEIR (Susceptible, Exposed, Infectious, and Recovered) model to propose a richer framework based on a continuous-time Markov chain where individuals transition between four distinct health states: healthy/susceptible, sick (reported cases), recovered/immune or asymptomatic, and deceased. A key feature, drawing analogy to search and matching models in economics, is the endogeneity of the infection rate, which depends dynamically on the proportion of infectious individuals (both symptomatic and asymptomatic) in the population.

Methodologically, the model is formalized as a system of ordinary differential equations (ODEs) governing the expected number of individuals in each state, derived from the underlying continuous-time Markov process. Calibration plays a crucial role: parameters governing recovery and mortality are informed by early epidemiological estimates, while the parameters driving the infection dynamics (contact rate a and its sensitivity parameters) are estimated by fitting the model's output for cu-

mulative reported cases (N_2^{ever}) and daily new cases (N_2^{new}) to the initial time series data from Germany's Robert Koch Institute (RKI) up to March 21, 2020. The analysis explicitly incorporates and investigates the significant uncertainty surrounding crucial parameters like the long-run total infection rate and the reported number of infected individuals who become sick, performing robustness checks across plausible ranges informed by data from Hubei and South Korea and early estimates for Germany. Policy interventions, such as lockdowns, are modeled counterfactually as temporary reductions in the contact rate parameter.

The calibrated model successfully tracks the initial exponential rise in reported cases in Germany. The projections and counterfactual analyses yield several insights: (1) Pandemic outcomes are highly sensitive to the assumed parameter values, underscoring the need for caution and robustness checks in forecasting. (2) Public health interventions (lockdowns) can effectively flatten the curve and reduce peak loads on the healthcare system, but their impact on the total number of eventual infections is limited within the model's structure unless they permanently alter the long-run infection rate $\overline{\rho}$. (3) The timing of interventions is critical, with the analysis suggesting a non-monotonic relationship where overly early or delayed interventions might be less effective at reducing the peak than optimally timed ones. The chapter's contribution lies in demonstrating the value of applying rigorous continuous-time stochastic modeling from economics to epidemiology. This approach provides a transparent, flexible framework for pandemic analysis that explicitly handles endogenous infection dynamics and parameter uncertainty, yielding policy-relevant insights into intervention strategies.

D.2 Zusammenfassung

Die Zusammenfassung beschreibt die Inhalte sowie die zentralen Ergebnisse der vorliegenden Dissertation.

Kapitel 1

Das erste Essay dieser Dissertation mit dem Titel *A General Quilibirum Model of Earnings, Income, and Wealth* (dt. Ein allgemeines Gleichgewichtsmodell für Arbeitseinkommen, Gesamteinkommen und Vermögen), in gemeinsamer Arbeit mit Frédéric Enrico Haller und Marc Puche Niubó, widmet sich der Herausforderung, die empirisch beobachteten Verteilungen von Arbeitseinkommen, Gesamteinkommen und Vermögen innerhalb eines einheitlichen allgemeinen Gleichgewichtsmodells gleichzeitig zu erklären. Die existierende Literatur, insbesondere die klassischen Bewley-Huggett-Aiyagari-Modelle, hat oft Schwierigkeiten, die ausgeprägte Vermögenskonzentration am oberen Rand der Verteilung abzubilden, ohne auf ad-hoc-Annahmen, Mechanismen wie "Superstar"-Einkommenszustände oder extrem persistente Schocks, zurückzugreifen. Zugleich erschwert die simultane Ab-

bildung der unterschiedlichen Formen der Einkommens- und Vermögensverteilungen die Modellierung zusätzlich. Dieses Essay schlägt eine Erweiterung des heterogenen Agenten-Modells vor, das vollständig in kontinuierlicher Zeit formuliert ist, um genau diese Probleme zu adressieren und ein integriertes Verständnis der Ursachen und Wechselwirkungen von Ungleichheit entlang dieser drei Dimensionen zu bieten, wie empirisch beobachtet (z. B. in Quadrini and Rıos-Rull (1997)).

Methodisch nutzt das Essay die analytische Handhabbarkeit und rechnerischen Vorteile von Modellen in kontinuierlicher Zeit. Dabei baut es wesentlich auf der numerischen Methodik von Achdou et al. (2022) auf, die auf finiten Differenzen (insbesondere dem Upwind-Verfahren für Stabilität) basiert. Die wesentliche Innovation liegt hier in der Erweiterung dieser numerischen Verfahren auf einen multidimensionalen Zustandsraum, der neben dem Haushaltsvermögen auch kontinuierliche stochastische Prozesse für individuelle Arbeitseinkommen und idiosynkratische Kapitalrenditen umfasst. Diese zwei simultanen Quellen kontinuierlichen Risikos, inspiriert durch Arbeiten von Benhabib et al. (2011) und Benhabib et al. (2015) zur Rolle heterogener Renditen, sind entscheidend für die Generierung der erforderlichen Verteilungsdynamiken. Das gesamte System wird im Rahmen eines allgemeinen Gleichgewichts gelöst, bei dem die aggregierten Faktorpreise (durchschnittliches Lohnniveau w und risikofreier Zinssatz θ) endogen durch Marktgleichgewichtsbedingungen bestimmt werden. Dies erfordert eine iterative Lösung, welche die HJB- und Fokker-Planck-Gleichungen mit dem aggregierten Kapital aus der stationären Verteilung koppelt.

Die Kalibrierung an zentrale Momente von US-Mikrodaten zeigt die empirische Flexibilität des Modells. Es repliziert erfolgreich die beobachteten Gini-Koeffizienten für Arbeitseinkommen, Gesamteinkommen und Vermögen sowie zentrale Teile ihrer Verteilungsformen, ohne dabei "Superstar"-Zustände nutzen zu müssen. Die Ergebnisse verdeutlichen die unterschiedlichen Rollen der beiden stochastischen Prozesse: Der kontinuierliche Einkommensprozess ist essenziell für die unteren Bereiche der Vermögensverteilung und gewährleistet numerische Stabilität, während der Prozess heterogener Zinssätze entscheidend für die Bildung des ausgeprägten oberen Verteilungsrands des Vermögens ist – es entsteht eine "Goldlöckchen-Zone", in der vermögende Haushalte durch hohe Renditen starke Sparanreize erhalten. Ein politisches Experiment, das die Wirkung erhöhter Arbeitslosenunterstützung untersucht, führt zu dem kontraintuitiven Ergebnis einer erhöhten langfristigen Vermögensungleichheit aufgrund einer Verringerung der Anreize für Vorsorgeersparnisse unter einkommensschwachen Haushalten. Der primäre Beitrag liegt in der Entwicklung eines neuartigen theoretischen Rahmens und einer praktikablen numerischen Methode zur gemeinsamen Analyse von Einkommens- und Vermögensungleichheit in einem vollständig spezifizierten allgemeinen Gleichgewichtsmodell in kontinuierlicher Zeit.

Kapitel 2

Das zweite Essay mit dem Titel *Sequence-Space Jacobians in Continuous Time* (dt. Zeitstetige Jacobi-

Matrizen im Folgenraum) befasst sich mit einem kritischen numerischen Engpass, der die Analyse dynamischer Modelle mit heterogenen Agenten erschwert: die Berechnung von Übergangspfaden und Impulsantworten nach aggregierten Schocks. Während die Analyse des stationären Gleichgewichts relativ standardisiert ist, ist die Berechnung, wie sich die Verteilung der Agenten und aggregierte Variablen nach einem Schock über die Zeit entwickeln, rechnerisch oft teuer oder gar nicht numerisch (in angemessener Zeit) durchführbar, insbesondere in stetiger Zeit, wo traditionelle Methoden typischerweise die Iteration über unendlichdimensionale Objekte (wie die Wertfunktion oder die Verteilung) über die Zeit erfordern. Dieses Kapitel führt eine hocheffiziente Alternative ein, indem es die Methodik der Jacobi-Matritzen im Folgenraum (SSJ) adaptiert, die von Auclert et al. (2021) für zeitdiskrete Modelle entwickelt wurde, und auf den stetigen Zeitrahmen überträgt. Der SSJ-Ansatz linearisiert die Reaktion des Modells in Bezug auf Folgen aggregierter Inputs um das stationäre Gleichgewicht herum und liefert hinreichende Statistiken (Jacobians), die Schocks auf Ergebnisse abbilden, ohne das vollständige nichtlineare Modell simulieren zu müssen.

Der zentrale methodische Fortschritt, der hier vorgestellt wird, ist der Beweis, dass der Infinitesimal-Generator der relevanten Hamilton-Jacobi-Bellman (HJB)- oder Fokker-Planck-Gleichung in stetiger Zeit direkt genutzt werden kann, um die zentralen Erwartungsvektoren effizient zu berechnen, die für die Konstruktion der Jacobians benötigt werden. Dies vermeidet die Notwendigkeit expliziter Zeititeration oder der Simulation individueller Agentenantworten, was den Rechenaufwand alternativer stetiger Zeitmethoden (z. B. Achdou et al. (2022)) verursacht. Durch das im Essay im wesentlich niedrigerdimensionalen Sequenzraum (der Folgen aggregierter Schocks auf Folgen aggregierter Ergebnisse abbildet) anstatt im hochdimensionalen Zustandsraum der Agenten gearbeitet wird, werden die Berechnungskosten der Methode weitgehend unabhängig von der Größe des Zustandsraums. Das Essay demonstriert die Anwendung sowohl auf ein neoklassisches Modell im Stil von Krusell-Smith als auch auf ein "one-asset-HANK-Modell" und zeigt die Kompatibilität mit verschiedenen Arten von idiosynkratischen Schockprozessen (Poisson und Diffusion). Darüber hinaus wird eine Darstellung mittels gerichteter azyklischer Graphen (DAG) verwendet, um systematisch allgemeine Gleichgewichts-Jacobians aus ihren Partialgleichgewichts-Gegenstücken zu berechnen.

Das Hauptergebnis ist eine signifikante Steigerung der Rechengeschwindigkeit. Für die untersuchten Beispiele wird gezeigt, dass die Berechnung von Übergangsdynamiken mit der SSJ-Methode in stetiger Zeit um mehrere Größenordnungen schneller ist (zwischen 30- und 1500-mal, je nach spezifischem Vergleich) als herkömmliche iterative Ansätze in stetiger Zeit, sobald der stationäre Zustand berechnet wurde. Diese Effizienzsteigerung verwandelt die Analyse dynamischer Reaktionen in stetigen Zeitmodellen mit heterogenen Agenten von einer rechenintensiven Aufgabe in eine nahezu instantane (oft Bruchteile einer Sekunde auf Standardhardware wie einem MacBook). Dies öffnet die Tür für Anwendungen, die bisher als undurchführbar galten, wie beispielsweise likelihood-basierte Schätzung oder die Analyse von Modellen mit sehr großen Zustandsräumen oder komplexen Dynamiken. Der Beitrag liegt darin, die Lücke zwischen der leistungsstarken SSJ-Methodik und den analytischen

Vorteilen der stetigen Zeitmodellierung zu schließen und ein sofort anwendbares und hocheffizientes Instrumentarium für Makroökonomen bereitzustellen, die mit heterogenen Agentenmodellen arbeiten.

Kapitel 3

Das dritte Essay mit dem Titel *Projecting the Spread of COVID-19 for Germany* (dt. "Projektion der Ausbreitung von COVID-19 für Deutschland"), eine Gemeinschaftsarbeit mit Jean Roch Donsimoni, Bodo Plachter und Klaus Wälde, veranschaulicht die breite Anwendbarkeit von stetigen stochastischen Modellierungstechniken, die typischerweise in der Ökonomie eingesetzt werden, indem sie zur Bewältigung eines der drängsten realen Probleme angepasst werden: die Prognose und Politikanalyse der COVID-19-Pandemie in Deutschland in ihrer frühen Phase. Das Essay adaptiert ein epidemiologisches Standard-SEIR-Modell, um einen erweiterten Modellrahmen vorzuschlagen, der auf einer stetigen Markov-Kette basiert, in dem Individuen zwischen vier verschiedenen Gesundheitszuständen wechseln: gesund/anfällig, krank (gemeldete Fälle), genesen/immun oder asymptomatisch und verstorben. Ein Schlüsselmerkmal, in Analogie zu Such- und Matching-Modellen in der Ökonomie, ist die Endogenität der Infektionsrate, die dynamisch vom Anteil infektiöser Personen (sowohl symptomatischer als auch asymptomatischer) in der Bevölkerung abhängt.

Methodisch wird das Modell als ein System gewöhnlicher Differentialgleichungen (ODEs) formalisiert, die die erwartete Anzahl von Individuen in jedem Zustand abbilden und welche aus dem zugrunde liegenden stetigen Markov-Prozess abgeleitet werden. Die Kalibrierung spielt hierbei eine entscheidende Rolle: Parameter, die Genesung und Mortalität steuern, basieren auf frühen epidemiologischen Schätzungen, während die Parameter, die die Infektionsdynamik steuern (Kontaktrate a und ihre Sensitivitätsparameter), geschätzt werden, indem der Modellausstoß für kumulierte gemeldete Fälle (N_2^{ever}) und tägliche neue Fälle (N_2^{new}) an die anfänglichen Zeitreihendaten des deutschen Robert Koch-Instituts (RKI) bis zum 21. März 2020 angepasst wird. Die Analyse berücksichtigt und untersucht explizit die erhebliche Unsicherheit bezüglich entscheidender Parameter wie der langfristigen Gesamtinfektionsrate und dem Anteil der infizierten Personen, die erkranken und gemeldet werden, und führt Robustheitsprüfungen über plausible Bereiche durch, die durch Daten aus Hubei und Südkorea sowie frühe Schätzungen für Deutschland gestützt werden. Politische Interventionen wie Lockdowns werden kontrafaktisch als vorübergehende Reduzierungen des Kontaktratenparameters modelliert.

Das kalibrierte Modell bildet den anfänglichen exponentiellen Anstieg der gemeldeten Fälle in Deutschland erfolgreich nach. Die Projektionen und kontrafaktischen Analysen liefern mehrere Erkenntnisse: (1) Pandemieergebnisse reagieren sehr sensibel auf die angenommenen Parameterwerte, was die Notwendigkeit von Vorsicht und Robustheitsprüfungen bei Prognosen unterstreicht. (2) Maßnahmen des öffentlichen Gesundheitswesens (Lockdowns) können die Kurve effektiv abflachen und Spitzenbelastungen des Gesundheitssystems reduzieren, ihr Einfluss auf die Gesamtzahl der letztendlichen

Infektionen ist jedoch innerhalb der Modellstruktur begrenzt, es sei denn, sie verändern dauerhaft die langfristige Infektionsrate ($\overline{\rho}$). (3) Der Zeitpunkt der Interventionen ist entscheidend, wobei die Analyse auf eine nicht-monotone Beziehung hindeutet, bei der zu frühe oder verzögerte Interventionen möglicherweise weniger wirksam bei der Reduzierung des Peaks sind als optimal getimte. Der Beitrag des Essays liegt darin, den Wert der Anwendung rigoroser stetiger stochastischer Modellierungsmethoden aus der Ökonomie auf die Epidemiologie aufzuzeigen, indem es einen transparenten, flexiblen Rahmen für die Pandemieanalyse bereitstellt, der endogene Infektionsdynamiken und Parameterunsicherheit explizit berücksichtigt, und politikrelevante Einblicke in Interventionsstrategien liefert.

Eidesstattliche Versicherung

Ich, René Marco Glawion, versichere an Eides statt, dass ich die Dissertation mit dem Titel:

"Continuous Time Heterogeneous Agent Models"

selbst und bei einer Zusammenarbeit mit anderen Wissenschaftlerinnen oder Wissenschaftlern gemäß den beigefügten Darlegungen nach § 6 Abs. 3 der Promotionsordnung der Fakultät Wirtschafts- und Sozialwissenschaften vom 18. Januar 2017 verfasst habe. Andere als die angegebenen Hilfsmittel habe ich nicht benutzt.

Ort/Datum Unterschrift

René Marco Glawion

Erklärung

Hiermit erkläre ich, René Marco Glawion, dass ich keine kommerzielle Promotionsberatung in Anspruch genommen habe. Die Arbeit wurde nicht schon einmal in einem früheren Promotionsverfahren angenommen oder als ungenügend beurteilt.

Ort/Datum Unterschrift

Selbstdeklaration bei kumulativen Promotionen

Die Einschätzung des geleisteten Anteils erfolgt mittels Punkteinschätzung von 1–100%.

Für mindestens einen der vorliegenden Artikel liegt die Eigenleistung bei 100%.

Koautoren bei Artikel 1: *Dr. Frédéric Enrico Haller*, *Dr. Marc Puche Niubó*
Koautoren bei Artikel 3: *Dr. Jean Roch Donsimoni*, *Prof. Dr. Bodo Plachter*, *Prof. Dr. Klaus Wälde*

Für den ersten Artikel liegt die Eigenleistung für	
das Konzept / die Planung bei	80%
die Durchführung bei	70%
der Manuskripterstellung bei	80%
Für den zweiten Artikel liegt die Eigenleistung für	
das Konzept / die Planung bei	100%
die Durchführung bei	100%
der Manuskripterstellung bei	100%
Für den dritten Artikel liegt die Eigenleistung für	
das Konzept / die Planung bei	70%
die Durchführung bei	80%
der Manuskripterstellung bei	40%

Die vorliegende Einschätzung in Prozent über die von mir erbrachte Eigenleistung wurde mit den am Artikel beteiligten Koautoren einvernehmlich abgestimmt.

Ort/Datum Unterschrift

René Marco Glawion

Liste der aus der Dissertation hervorgegangenen Veröffentlichungen

Kapitel 3 der in dieser Dissertation enthaltenen Essays is in einem Journal mit Peer-Review Verfahren veröffentlicht:

Donsimoni, J. R., Glawion, R., Plachter, B., Wälde, K. (2020). Projecting the spread of COVID-19 for Germany. *German Economic Review*, 21(2), 181-216.