

Dissertation

Optimal prevention and adaptation under risk in health and environmental economics

Ausgeführt zum Zwecke der Erlangung des akademischen Grades eines Doktors der technischen Wissenschaften, eingereicht an der TU Wien, Fakultät für Mathematik und Geoinformation, von

Dipl.-Ing. Michael Freiberger

Mat.Nr.: 01025675

unter der Leitung von

Univ.Prof. Dipl.-Ing. Dr.techn. Alexia Fürnkranz-Prskawetz
Institut für Stochastik und Wirtschaftsmathematik

Wien, Mai 2022

Dissertation

Optimal prevention and adaptation under risk in health and environmental economics

carried out for the purpose of obtaining the degree of Doctor technicae (Dr. techn.),
submitted at TU Wien, Faculty of Mathematics and Geoinformation, by

Dipl.-Ing. Michael Freiberger

Mat.Nr.: 01025675

under supervision of

Univ.Prof. Dipl.-Ing. Dr.techn. Alexia Fürnkranz-Prskawetz
Institute of Statistics and Mathematical Methods in Economics

Vienna, May 2022

REVIEWERS

Univ.Prof. Dipl.-Ing. Dr.techn. Alexia Fürnkranz-Prskawetz
Faculty of Mathematics and Geoinformation,
Institute of Statistics and Mathematical Methods in Economics
Wiedner Hauptstraße 8-10/105b, 1040 Vienna, Austria
afp@econ.tuwien.ac.at

Univ.Prof. Dipl.-Math. Dr. Timo Trimborn
Department of Economics and Business Economics,
Fuglesangs Allé 4, 8210 Aarhus V, Denmark
tt@econ.au.dk

Univ.Prof.i.R. Dipl.-Ing. Dr.techn. Vladimir Veliov
Faculty of Mathematics and Geoinformation,
Institute of Statistics and Mathematical Methods in Economics
Wiedner Hauptstraße 8-10/105b, 1040 Vienna, Austria
vladimir.veliov@tuwien.ac.at

Affidavit

I declare in lieu of oath, that I wrote this thesis and performed the associated research myself, using only literature cited in this volume. If text passages from sources are used literally, they are marked as such.

I confirm that this work is original and has not been submitted elsewhere for any examination, nor is it currently under consideration for a thesis elsewhere.

I acknowledge that the submitted work will be checked electronically-technically using suitable and state-of-the-art means (plagiarism detection software). On the one hand, this ensures that the submitted work was prepared according to the high-quality standards within the applicable rules to ensure good scientific practice "Code of Conduct" at the TU Wien. On the other hand, a comparison with other student theses avoids violations of my personal copyright.

Michael Freiberger
Vienna, May 2022

Abstract

Risk and uncertainty significantly shape human lives in many different ways, at different levels and over different time scales. Examples can be found in the savings decisions at the individual level, which are driven by short-term fluctuations in labour income. More long-term considerations are relevant for the behaviour with respect to health. Unhealthy behaviour in the present can have long-term impacts on health or even affect ones mortality risk in the future. Regarding risks affecting a multitude of people simultaneously (instead of independently), we can consider natural disasters or other environmental risks. These types of risks are not only relevant for policy makers on the macro level, but for decision making at the individual level as well. On the other hand, as the Covid-19 pandemic has shown, individual risks like a virus infection can also have aggregate effects, which make macro level intervention necessary.

This thesis discusses human behaviour with respect to risk and uncertainty in these aforementioned settings. The focus of each of the three models presented is on a different type of risk ranging from health risk on the individual and macro level to environmental risk. Different frameworks are used to account for the characteristics of each type of risk.

The first paper introduces large health shocks into individual life-cycle models. Compared to smaller shocks, which mostly have transitory or temporary effects on a person's life, large shocks (such as a cancer diagnosis or a cardio-vascular disease) have the potential to put the individuals life course on a completely different trajectory. Although large health shocks cannot be completely eliminated through any type of behaviour, the individual can affect the hazard rate of the shock using general and disease specific health investments. On the other hand one can also undertake precautionary measures for the time after a potential shock to lessen its impact, such as keeping precautionary savings to finance future health care.

Using this framework we extend the definitions of the classical statistical value of life with respect to different aspects of the individual's life and apply these new valuations to characterise the optimal health investments over the life-cycle. Moreover we derive a decomposition for each valuation with respect to its different contributing channels. We also calibrate the model for the scenario of a potential cancer diagnosis for individuals in the US and are able to replicate the health decisions and outcomes qualitatively and quantitatively. The results of the calibration exercise allow us to numerically assess the impact of a cancer diagnosis on general consumption, and health investments and outcomes compared to a cancer-free person. Moreover, we identify the role of the timing of the shock, i.e. the age at which the individual is diagnosed with cancer, with respect to the optimal decision making. Using a decomposition of Euler-type equations for consumption and health investments we further establish several different drivers for the optimal behaviour resulting from the cancer diagnosis before and after it has occurred.

The second paper considers the optimal health policy on the macro level facing the pandemic spread of an infections disease. The work extends a SIR-compartment model with respect to several aspects. Distinction between light and heavy courses of the disease allows for a proper differentiation between individuals contributing to the spread of infections and people putting burden on the health capacities of an economy. Additional to the commonly used interventions in the form of lockdowns to reduce the transmission rates we introduce testing as a potential strategy for the policy maker. Testing allows to identify infectious individuals and remove them from the transmitting population. Furthermore we allow for heterogeneity within the population analysed by introducing a network structure. This enables

us to consider the optimal containment of the pandemic across a regional network. At the same time a network is also able to capture the interactions of different social groups within the same region. In both cases (or even a combination of them) we are able to assign each group heterogeneous characteristics from both an economic and epidemiological point of view. Using this framework we can identify group specific optimal intervention strategies, which reduce the aggregated burden of the pandemic compared to interventions applied uniformly.

Solving the model, we derive analytical characterisations of the optimal policy and identify a trade-off between investments in testing and lockdowns. Furthermore we illustrate the capabilities of the model numerically for a Covid-like disease (similar epidemiological characteristics) in a three-region setting with an initial hotspot of infections in one of the regions. We show that region-specific measures are set to simultaneously avoid overburdening the hospital capacities and homogenise the pandemic situation across the different regions. Moreover we find that testing allows for both reductions in the total costs of the pandemic as well relaxing of lockdown measures. However, the impact of testing is substantially higher, if tests are allocated more specifically to individuals with a high probability of being infected instead of broadly across the population. This suggests high potential benefits from properly conducted contact tracing.

In the third paper we return to the decision making process at the individual level. Motivated by the empirical literature assessing the correlation of education and different aspects of household behaviour with respect to disaster risk, we introduce a dynamic household model to explain these correlations systematically. Within the model we incorporate the four main channels through which education shapes the individual decision making found in the literature: income, awareness, time preference and access to prevention measures.

In our framework households face the risk of losing substantial amounts of their physical wealth (e.g. housing or other valuable items), in case they are hit by a natural disaster. These losses can be replaced through financial efforts, which can require reductions in consumption and/or financial saving and hence need to be carefully chosen. On the other hand, these ex-post strategies can be substituted by ex-ante interventions which either reduce the share of physical assets destroyed during a disaster (e.g. flood protection measures) or decrease the probability that the household is affected in the first place (e.g. relocation to different settlement).

Using empirical data from Thailand and Vietnam we parameterize and calibrate the model for the numerical solution. Utilizing Monte-Carlo-Simulations we generate a synthetic population, which replicates the socio-economic structure of the empirical data. We find, that our model is able to replicate the empirical household outcomes qualitatively well. Graphical illustrations and regression analysis are used to assess the impact of the educational distribution on the equilibrium outcomes of households with respect to assets and disaster risk. We also investigate how the educational level affects the decision making of households under similar conditions. Furthermore we decompose the total effects of education on risk related variables into their direct and indirect impact channels. After controlling for income, awareness, prevention access and time preference, the education level (and its impact on the households expected income developments) still remains the most influential factor among all household characteristics considered with respect to risk behaviour.

To conclude, this thesis illustrates three examples of human behaviour with respect to different types of risk in the contexts of individual health, public health and environmental disasters. Although this also requires different methodological approaches, the key aspect of finding trade-offs between preventive efforts in anticipation of a potential shock and decisions to adjust/react after the shock are present in all of them. The final outcomes naturally depend on the specific setting and characteristics of the decision maker. In conclusion, this thesis provides strong insights into the driving forces behind the optimal decisions at the individual and macro level when facing risk and uncertainty.

Kurzfassung

Risiken und Ungewissheiten prägen das menschliche Leben auf vielfältige Weise, auf verschiedenen Ebenen und über unterschiedliche Zeiträume hinweg. Beispiele dafür finden sich in den Sparentscheidungen auf individueller Ebene, die von kurzfristigen Schwankungen des Arbeitseinkommens getrieben werden. Für das Gesundheitsverhalten andererseits sind längerfristige Überlegungen entscheidend. Ungesundes Verhalten in der Gegenwart kann sich langfristig auf die Gesundheit auswirken oder sogar das zukünftige Mortalitätsrisiko beeinflussen. Naturkatastrophen oder andere Umweltrisiken sind andererseits Ereignisse, die Menschen nicht auf individueller Ebene sondern gleichzeitig und kollektiv betreffen können. Diese Arten von Risiken sind nicht nur für politische Entscheidungsträger auf der Makroebene relevant, sondern auch für die Entscheidungsfindung auf individueller Ebene. Wie die Covid-19-Pandemie wiederum gezeigt hat können individuelle Risiken (wie eine Virusinfektion) aufgrund ihrer aggregierten Auswirkungen ein Eingreifen auf der Makroebene erforderlich machen.

Diese Dissertation behandelt menschliches Verhalten in Bezug auf Risiko und Unsicherheit in den oben beschriebenen Situationen. Der Fokus der drei vorgestellten Modelle liegt jeweils auf einem anderen Risikotypus, i.e. vom Gesundheitsrisiko auf individueller und Makroebene bis zum Umweltrisiko. Unterschiedliche Modellierungsansätze ermöglichen es die speziellen Merkmale jeder Risikoart zu berücksichtigen.

Das erste Modelle behandelt die Erweiterung klassischer Lebenszyklusmodelle um signifikante Gesundheitsschocks. Im Vergleich zu kleineren Schocks, die meist nur temporäre Auswirkungen auf das Leben einer Person haben, verfügen ausgeprägte Schocks (wie z.B. eine Krebsdiagnose oder eine Herz-Kreislauf-Erkrankung) über ein Potenzial, den Lebensverlauf auf eine völlig andere Bahn zu lenken. Obwohl ausgeprägte Gesundheitsschocks durch keinerlei Maßnahmen komplett eliminiert werden können, kann durch krankheitsspezifische Gesundheitsinvestitionen die Auftretenswahrscheinlichkeit auf individueller Ebene beeinflusst werden. Andererseits können alternativ auch Vorsorgemaßnahmen getroffen werden für die Zeit nach einem möglichen Schock, um dessen Auswirkungen abzufedern (z.B. Akkumulierung zusätzlicher Ersparnisse um die zusätzlichen Gesundheitsausgaben zu finanzieren).

Im Rahmen dieses Modells erweitern wir die Definition des klassischen Value-of-Life auf weitere Aspekte des persönlichen Lebens (insbesondere im Bezug auf Gesundheit) und verwenden diese neuen Bewertungen, um die optimalen Gesundheitsinvestitionen über den Lebenszyklus zu charakterisieren. Darüber hinaus präsentieren wir eine Zerlegung für jede dieser Bewertungen in Bezug auf ihre verschiedenen Einflussfaktoren. Anschließend kalibrieren wir das Modell für das Szenario einer potenziellen Krebsdiagnose für Personen in den USA und sind in der Lage, die individuellen Gesundheitsentscheidungen und -entwicklungen qualitativ und quantitativ zu replizieren. Außerdem ermöglichen die Ergebnisse der Kalibrierung die Auswirkungen einer Krebsdiagnose auf den allgemeinen Konsum sowie die Gesundheitsinvestitionen und -entwicklungen im Vergleich zu einer nicht diagnostizierten Person numerisch zu evaluieren. Weiters können wir die Rolle des Zeitpunkts des Schocks, d. h. des Alters, in dem eine Person mit Krebs diagnostiziert wird, im Hinblick auf die optimale Entscheidungsfindung analysieren. Unter Verwendung von Zerlegungen der Euler-Gleichungen für Konsum und Gesundheitsinvestitionen ermitteln wir identifizieren wir mehrere unterschiedliche Treiber für das optimale Verhalten vor und nach einer Krebsdiagnose.

Der zweite Teil dieser Dissertation betrachtet die optimale Gesundheitspolitik auf der Makroebene angesichts der pandemischen Ausbreitung einer Infektionskrankheit. Diese Arbeit erweitert ein SIR-

Compartment-Modell in Bezug auf mehreren Aspekte. Die Unterscheidung zwischen leichten und schweren Krankheitsverläufen ermöglicht eine entsprechende Unterscheidung zwischen Personen, die zur Ausbreitung von Infektionen beitragen, und Personen, die die Kapazitäten des Gesundheitssystems belasten. Zusätzlich zu den häufig verwendeten Interventionen in Form von Lockdowns zur Reduzierung der Übertragungsraten führen wir Tests als mögliche Strategie für die politischen Entscheidungsträger ein. Durch Tests können infektiöse Personen identifiziert und vom Infektionsgeschehen getrennt werden. Um darüber hinaus die Heterogenität innerhalb der analysierten Population zu berücksichtigen führen wir eine Netzwerkstruktur ein. Dies ermöglicht es uns, die optimale Eindämmung der Pandemie über ein regionales Netzwerk hinweg zu betrachten. Gleichzeitig ist eine Netzwerkstruktur auch in der Lage, die Interaktionen verschiedener sozialer Gruppen innerhalb derselben Region zu replizieren. In beiden Fällen (oder auch in Kombination) können wir dadurch jeder Gruppe sowohl aus ökonomischer als auch aus epidemiologischer Sicht heterogene Merkmale zuordnen. Mit Hilfe dieses Frameworks können wir gruppenspezifische optimale Interventionsstrategien identifizieren, die die Gesamtbelastung der Pandemie im Vergleich zu uniform angewandten Interventionen reduzieren.

Die analytische Lösung des Modells erlaubt uns die optimale Strategie zu charakterisieren und einen Trade-off zwischen Investitionen in Tests und Lockdowns zu identifizieren. Darüber hinaus veranschaulichen wir die potentiellen Möglichkeiten des Modells numerisch für eine Infektionskrankheit mit Covid-ähnlichen epidemiologische Merkmalen in einem Drei-Regionen-Setting mit einem anfänglichen Infektionsherd in einer der Regionen. Wir zeigen, dass die optimalen regionalspezifische Maßnahmen beinhalten gleichzeitig eine Überlastung der Krankenhauskapazitäten zu vermeiden und die Pandemiesituation über die verschiedenen Regionen hinweg zu homogenisieren. Weiters stellen wir fest, dass Tests sowohl eine Reduzierung der Gesamtkosten der Pandemie als auch eine Reduzierung der Lockdown-Maßnahmen ermöglichen. Die Effekt von Tests ist jedoch wesentlich höher, wenn diese für die gezieltere Testung von Personen mit hoher Wahrscheinlichkeit einer Ansteckung verwendet werden, anstatt uniform über die Bevölkerung verteilt zu werden. Dies deutet auf hohe potenzielle Vorteile einer gewissenhaft durchgeführten Kontaktverfolgung hin.

Für den dritten Teil des Dissertation kehren wir zum Entscheidungsfindungsprozess auf individueller Ebene zurück. Motiviert durch die empirische Literatur, die den Zusammenhang zwischen Bildung und verschiedenen Aspekten des Haushaltsverhaltens in Bezug auf Risiken von Naturkatastrophen untersucht, entwickeln wir ein dynamisches Haushaltsmodell, das versucht diese Zusammenhänge systematisch zu erklären. In diesem Modell integrieren wir die vier Hauptkanäle aus der empirischen Literatur, durch die Bildung die individuelle Entscheidungsfindung beeinflusst: Einkommen, Risikobewusstsein, Zeitpräferenz und Zugang zu Präventionsmaßnahmen.

In unserem Framework sind Haushalte dem Risiko ausgesetzt, beträchtliche Mengen ihres materiellen Vermögens (z.B. Wohnraum oder andere Wertgegenstände) zu verlieren, falls sie von einer Naturkatastrophe getroffen werden. Diese Verluste können mit Hilfe finanzieller Anstrengungen ersetzt werden, die jedoch im Gegenzug Reduzierung des Konsums und/oder finanzielle Einsparungen erfordern können und daher sorgfältig behandelt werden müssen. Andererseits können diese Ex-Post-Strategien potentiell durch Ex-Ante-Interventionen ersetzt werden, die entweder den Anteil der bei einer Katastrophe zerstörten Sachwerte (z.B. Hochwasserschutzmaßnahmen) verringern oder die Wahrscheinlichkeit verringern, dass der Haushalt überhaupt von einer Katastrophe betroffen ist (z.B. Umzug an einen anderen Siedlungsort).

Für die numerische Lösung parametrisieren und kalibrieren wir das Modell anhand empirischer Daten aus Thailand und Vietnam. Mithilfe von Monte-Carlo-Simulationen generieren wir eine synthetische Population, die die sozioökonomische Struktur der empirischen Daten repliziert. Die Ergebnisse zeigen, dass unser Modell in der Lage ist, die empirischen Haushaltsdaten qualitativ zu replizieren. Mit Hilfe von geeigneten grafischen Darstellungen und Regressionsanalysen illustrieren wir den Einfluss von Bildung auf die Gleichgewichtsverteilungen von Wohlstand und Katastrophenrisiko. Außerdem untersuchen wir, wie sich das Bildungsniveau auf die Entscheidungsfindung von Haushalten unter ähnlichen Voraussetzungen auswirkt. Darüber hinaus zerlegen wir die Gesamteffekte von Bildung auf risikorelevante Variablen in ihre direkten und indirekten Wirkungskanäle. Selbst unter Berücksichtigung von Einkommen,

Risikobewusstsein, Präventionszugang und Zeitpräferenz bleibt das Bildungsniveau (und sein Einfluss auf die erwartete Einkommensentwicklung der Haushalte) immer noch der einflussreichste Faktor unter allen betrachteten Haushaltsmerkmalen in Bezug auf das Risikoverhalten.

Zusammenfassend veranschaulicht diese Arbeit drei Beispiele menschlichen Verhaltens in Bezug auf unterschiedliche Typen von Risiken im Zusammenhang mit individueller Gesundheit, öffentlicher Gesundheit und Umweltkatastrophen. Obwohl wir für die Analyse unterschiedliche methodische Ansätze verwenden, ist der zentrale Aspekt Kompromisse zwischen präventiven Maßnahmen in Erwartung eines potenziellen Schocks und adaptiven Reaktionen nach dem Schock zu finden, in allen Modellen vorhanden. Die detaillierten Ergebnisse hängen dabei klarerweise von den spezifischen Rahmenbedingungen und den Eigenschaften des Entscheidungsträgers ab. In seiner Gesamtheit bietet diese Dissertation tiefgehende Einblicke in die treibenden Kräfte hinter den optimalen Entscheidungen unter Einfluss von Risiko und Unsicherheit auf individueller und Makroebene.

Acknowledgements

This dissertation is the result of a long process and would not have been possible without the incredible support of many colleagues and friends along the way.

First and foremost I want to thank my supervisor Alexia Fürnkranz-Prskawetz, who supported me in my ambitions to join the scientific research community already during my early years at university. Not only have I been able to benefit from her knowledge and expertise in a wide range of topics related to economics, she also opened many doors to great opportunities for me all throughout my career.

I am also very grateful for the trust my current colleagues Stefan Wrzaczek and Michael Kuhn put in me as they gave me the opportunity to join the FWF-project “Life-cycle behaviour in the face of large shocks to health” (P 30665-G27) at the Vienna Institute of Demography in 2018. Despite my limited experience they allowed me to play significant role in the successful completion of the project. I am also more than happy and grateful that Michael Kuhn gave me the chance to continue working with both them at the International Institute for Applied Systems Analysis.

Next I would like to thank Franz Hof for initially igniting my passion for economic theory in my second year at University by his great teaching skills and I appreciate that we are still keeping contact after all those years.

Especially deserving to be mentioned here is my cross-institutional office-colleague and friend Bernhard Rengs. Although we never directly collaborated on a project, his support and expertise on the technical and computational aspects of modern research saved me countless days/weeks of troubleshooting and I have been able to pick up substantial knowledge on topics I would not have received anywhere else. I also always deeply enjoyed our talks and discussion on non-scientific topics ranging from pop-culture to politics. I honestly hope our academic paths cross again in future.

Furthermore, I would like to thank all the other colleagues who I met along my way and who provided me with new thoughts and ideas during great conversation. At the TU I had inspiring interactions at the Institute of Statistics and Mathematical Methods in Economics with Johanna Games, Martin Kerndler, Emanuel Gasteiger, Timo Trimborn, Nawid Siassi, Dieter Grass, Andrea Seidl, Anna Dugan and Stefani Rivic, but also great interdisciplinary collaborations with Helmut Rechberger and Oliver Cencic.

At the VID I was lucky enough to meet and work with Roman Hoffmann, Miguel Sanchez-Romero, Bernhard Hammer, Sonja Spitzer, Klaus Prettnner and Gustav Feichtinger during different projects. At IIASA I found great support from my colleague Yuliya Kulikova and profited from a fruitful international collaboration with Alessandra Buratto and Maddalena Muttoni.

Before expressing my appreciation for all the people accompanying me along the way outside of academia, I need to acknowledge the three different institutions which provided me with the financial stability and support to conduct the research in- and outside of this thesis. The Vienna Doctoral Programme on Water Resource Systems (DK-plus W1219-N22) based at the TU Wien with financial support from the Austrian Science Funds (FWF) allowed me to start my research on interdisciplinary topics, which I am very grateful for and want especially thank the chair of the programme Günter Blöschl. Next I received support from the FWF through the project “Life-cycle behaviour in the face of large shocks to health” (P 30665-G27) under the leadership of Stefan Wrzaczek. Lastly I want to thank the International Institute for Applied Systems Analysis and my current program director Michael Kuhn for their support during the last year.

Finally I want to describe my appreciation for my friends and family, who supported me through the

years. First and foremost I want to thank Eva for her support and keeping up with my quirks over all the years. Everything I accomplished in my career would not have been possible without her. I am also grateful for the support I received from my family Andrea, Laurie, Claudia and Alexander. Furthermore I also have to thank Denis for being my “companion in misfortune” during the occasionally bothersome times of writing a dissertation. Lastly I want to express a collective thank you to all my other friends, who were there for me over the last years: Thomas, Ines, Clemens, Alex, Tudor, Dagmar, Simon, Raoul, Martin, Stefan, Joscha and all the other Honeybees.

Contents

1	Introduction and Motivation	19
2	Integrating large health shocks	23
2.1	Introduction	23
2.2	Model	26
2.3	Analytical results	30
2.3.1	Valuations of Health	30
2.3.2	First order optimality conditions	34
2.4	Numerical analysis: an application to cancer	36
2.4.1	Data sources	36
2.4.2	Remarks on the annuity market	38
2.4.3	Solution strategy	38
2.4.4	Functional specifications	38
2.5	Numerical results	41
2.5.1	Expenditure and survival patterns: Model vs. data	42
2.5.2	Consumption profiles	43
2.5.3	Health expenditure and state profiles	47
2.5.4	Euler equations for health care investments	51
2.6	Conclusions	58
2.7	Appendix	60
2.7.1	The model as a vintage optimal control model	60
2.7.2	Necessary optimality conditions	62
2.7.3	Proof of proposition 1 on the valuations of health	63
2.7.4	Proof of proposition 2 on the first order optimality conditions	67
2.7.5	Derivation of the Euler equations	67
2.7.6	Estimation and calibration strategies	70
3	Allocating tests and lockdowns within a network	73
3.1	Introduction	73
3.2	Literature review	76
3.3	SIR-type Model	77
3.3.1	Dynamical system	79
3.3.2	Costs and objective function	81
3.3.3	Complete problem	83
3.4	Properties of the optimal solution	84
3.4.1	Optimal allocation	84
3.4.2	Properties	85
3.5	Numerical Analysis	86
3.5.1	Calibration	87
3.5.2	Results	90
3.6	Conclusions	101

3.7	Appendix	102
3.7.1	Optimality conditions	102
3.7.2	Interpretation of shadow prices	104
3.7.3	Summary tables for numerical results	106
4	Modelling Disaster Risks: A Dynamic Household Model	109
4.1	Introduction	109
4.2	Education and Disaster Risk Reduction: The Empirical Evidence	111
4.3	Conceptual framework	112
4.4	The household model	113
4.4.1	Impact of household characteristics	115
4.4.2	Problem summary	117
4.4.3	Analytical results	117
4.4.4	Bellman formulation	119
4.4.5	Numerical solution strategy	120
4.5	Parametrisation and calibration	120
4.5.1	Parameter estimates	121
4.5.2	Empirical distributions	123
4.6	Numerical results	124
4.6.1	Equilibrium distributions	125
4.6.2	Impact of education and other household characteristics	127
4.6.3	Optimal policy functions	134
4.6.4	Exposure, Vulnerability and disaster risk	137
4.7	Conclusions and discussion	138
4.8	Appendix	140
4.8.1	Analytical Results	140
4.8.2	Numerical solution	143
5	Conclusions	147
5.1	Overall conclusion	147
5.2	Potential extensions and outlook	149
	List of Tables	151
	List of Figures	153
	Bibliography	155

Chapter 1

Introduction and Motivation

Dealing with risk and uncertainty in every day life is inherent to the human existence. However the extensive formalisation of random events only dates back to the 17th century. The contributions of mathematician Blaise Pascal were essential for the development of probability theory, which were interestingly motivated by problems arising in gambling. On the other hand his argumentation on the rationality of believing in god (or at least living as if god exists) was one of the first known consideration of expected utility maximisation and sets the origin for the emergence and progression of expected utility theory.¹

In the present day modelling of uncertainty and risk is naturally far more developed as numerous different techniques for the formalisation of these problems and their solution have been established especially since the second half of the 20th century. Before giving an overview of the historic progression of the scientific literature on risk behaviour, we start with definitions of subsequently used relevant terminology. As different research areas evolve in parallel, popular terms often are improperly (and wrongfully) used interchangeably. This holds true for everyday language, but also the scientific language.

For the definition of the term **Risk** itself, we follow the glossary of the IPCC (Masson-Delmotte et al., n.d.) describing it as the result of the combination of hazard, exposure and vulnerability. Risk covers *“the potential for adverse consequences where something of value is at stake and where the occurrence and degree of an outcome is uncertain.”* **Hazard** captures *“the potential occurrence of a natural or human-induced physical event or trend that may cause loss of life, injury, or other health impacts, as well as damage and loss to property, infrastructure, livelihoods, service provision, ecosystems and environmental resources”*, and therefore describes the aspects of risk on the large/aggregated scale. Meanwhile exposure and vulnerability describe facets on smaller scales or individual levels. **Exposure** specifies *“the presence of people (...) or economic, social, or cultural assets in places and settings that could be adversely affected.”* **Vulnerability** is maybe the most abstract aspect of risk and is defined as *“The propensity or predisposition to be adversely affected”* or in other words covers the extent of affectedness in case an individual (or other entity) is hit by an adverse event.

Another term often used in the same vein (especially in the context of socio-ecological models) is **Resilience**. However, as Li et al. (2018) state many different interpretations of resilience have been used in the literature. From a systems approach one of the most practical definition might originate from Holling (1973): *“Resilience is concerned with the capacity of a system to persist and develop in its regime without shifting into another regime.”*² In their discussion of various aspects of resilience, Li et al. (2018) also tie in the terms of mitigation and adaptation. While both try to increase the resilience of a system, their means for success are substantially different. **Mitigation** efforts aim to reduce the probability of adverse events occurring in the first place, whereas **Adaptation** includes preparations and capacity building before the adverse event. Proper adaptation measures lead to effective reactions to a shock and potentially remove the disruptive nature of such an event.

While adaptation and mitigation are often used in context of (larger scale) socio-ecological models,

¹See Lengwiler (2009) for a comprehensive summary of the origins of expected utility theory.

²See Li et al. (2018) p. 311.

on the smaller scale (or individual level) terms like prevention and precaution are more common. As Courbage et al. (2013) describe, **Prevention** measures themselves can be classified into two subcategories. **Loss prevention** (or also called self-protection) has similar objective as mitigation, i.e. to reduce the probability of an adverse event actually occurring. **Loss reduction** (or self-insurance) measures on the other hand have the aim to reduce the impact of a negative shock once it strikes and hence has similarities to the concept of adaptation described above. Tying back to definitions of hazard, exposure and vulnerability, loss prevention/mitigation can be seen as actions to reduce hazard and/or exposure, whereas loss reduction/adaptation targets the vulnerability aspect of risk.

Independent of the terminology used, the points to be answered within the modelling effort in most contexts mostly revolve around the following questions:

- (i) What efforts should be made before an adverse event has occurred? These efforts can include:
 - Measures to reduce the probability of the event occurring.
 - Interventions to reduce the direct impact of the event.
 - Efforts to increase capabilities to deal with the effects of the event ex-post.
- (ii) What is the optimal strategy to deal with the impacts of a negative shock in case it occurs?
- (iii) What are the trade-offs between ex-ante and ex-post measures?

This thesis has the goal to answer these questions and generate new insights in the fields of health and environmental economics. We consider decisions on the individual level as well as optimal policies on larger scales. Thereby the models used are based on a long history of developments in different fields of economics and mathematics and contain different systemic approaches. For the three models presented in this thesis we choose the modelling approach according to specific needs of the research questions posed. The first two models follow the latest developments in optimal control theory, whereas the third model is based on a dynamic programming approach. The historic progress since the second half of the 20th century for the two distinct model classes applied to the modelling of risk behaviour is quite different.

For the branch of literature using optimal control theory, Yaari (1965) and Kamien and Schwartz (1971) were the first to introduce a stochastic element into an optimal control model, i.e. a stochastic time-horizon. While the former considers the optimal consumption over the life-cycle of individuals with uncertain time of death, the latter investigates the handling of stochastic failure of production machinery. Both works build a foundation for large amounts of literature in their respective areas. In health economics, e.g. the development and formalisation of the concept of the value of life (i.e. the willingness to pay for a reduction in the mortality risk) in Rosen (1988) is based on the framework of Yaari (1965) with uncertain end of life. Many other publications like Ehrlich (2000), Schuenemann et al. (2017) or Kuhn et al. (2011) (to name just a few) consider uncertain survival as an incentive for investments into individuals health. Kamien and Schwartz (1971) on the other hand inspired consecutive works of Alam et al. (1976), Eichenbaum et al. (2021) and Gaimon and Thompson (1984). However, both approaches of Yaari (1965) and Kamien and Schwartz (1971) share that the salvage value of the proposed optimal control model is equal to zero.

In the field of environmental economics Reed applied similar frameworks with stochastic time-frames (Reed, 1984; Reed, 1987; Reed, 1989) to the optimal protection of natural resources. Thereby he used non-zero salvage-functions in the process and transformed the models into equivalent deterministic ones. Boukas et al. (1990), Sorger (1991) and Carlson et al. (2012) formalised this framework with a given salvage-function and analysed models of that kind from a theoretical point of view. Although all these findings allow for assessments and analysis of prevention strategies in the face of risk and uncertainty, models with exogenous salvage-functions do not provide information on the optimal adaptation and reaction strategies after a stochastic shock. To resolve this issue multi-stage optimal control models emerged in the literature. Tomiyama (1985) and Tomiyama and Rossana (1989) develop theoretical tools for two-stage optimal control models, however, the time of the switch between the two stages thereby is

a control variable. Hence, these frameworks are not directly applicable for the modelling of a stochastic adverse event. Tsur and Zemel (1995); Tsur and Zemel (1996); Tsur and Zemel (1998) discussed the scenario of a state-dependent stochastic regime shift in an environmental economic setting, yet still used exogenous “post-event” salvage-functions. Polasky et al. (2011) and De Zeeuw and Zemel (2012) were one of the first works considering a stochastic switch between two different regimes with explicit consideration of the solution of the second stage. Nevertheless, their insights were still limited to some degree, as they either needed to analytically solve for the optimal solution of the second stage (which is only possible for certain types of models) or derived numerical solutions using specifically developed numerical methods (again not generally applicable).

Most recently Wrzaczek et al. (2020) proved the general equivalence of two-stage optimal control problems with stochastic switching and deterministic vintage-structured optimal control models and showed how the former can be transformed into the latter. This transformation significantly simplifies the analytical and numerical treatment of this model class and has been applied in several different areas (Kuhn and Wrzaczek, 2021; Buratto et al., 2022). The first part of the thesis aims to substantially contribute to and extend this strain of the literature with an application of the transformation technique to an individual life-cycle model with potential large shocks to health (see Chapter 2).

On the other hand, literature based on dynamic programming started the inclusion of stochastic elements by extending the Hamilton-Jacobi-Bellman (HJB) equations (Bellman, 1954) with Markov-processes. Schechtman (1976) presents the stochastic income fluctuation problem and uses stochastic dynamic programming for the solution. Kydland and Prescott (1982) significantly extend these findings and build the foundation for the extensive Real-Business-Cycle (RBC) literature. This framework is quite flexible and almost any stochastic process, which can be described with a Markov-Process, is relatively easy to incorporate into the basic framework. However, this flexibility results in the downside of comparatively few analytical insights into the optimal strategies and the dependency on numerical results is significant. Despite these downsides we find this framework to be most appropriate way to assess the behaviour of households with respect to disaster risk (see Chapter 4).

For both types of approaches (optimal control as well as dynamic programming) the possibilities to gain theoretical insights become increasingly scarce, as models get more involved. Hence, in a lot of cases we have to rely on the numerical solution of the model to gain insights on the dynamics of the model and the optimal solutions. However, the process of deriving an optimal solution numerically is far from a trivial problem itself. For optimal control models the Pontryagin Maximum Principle (Pontryagin, 1987) provides a set of ordinary differential equations (ODE) with boundary conditions and point-wise optimality conditions of the Hamiltonian.

For the solution of dynamic programming problems the stochastic HJB equations (Bellman, 1954) provide optimality conditions for all possible combinations of state variable, however, in general they involve an unknown value-function. Again, this poses significant difficulties for the analytical inspection and numerical solution of the model.

Standing on the shoulders of the literature discussed above, the aim of this thesis is to enhance the understanding of human decision making in face of risk and uncertainty. Chapter 2 shows how large stochastic shocks to health can be integrated into life-cycle models. Compared to smaller transitory shocks with mostly temporary effects, large health shocks can put an individual’s life on a whole different trajectory. We transform the arising two-stage optimal control model into a vintage-structured optimal control model (following Wrzaczek et al. (2020)) and derive the driving forces of the individuals preventive and precautionary behaviour in anticipation of a potential diagnosis and adaptive decisions after a diagnosis for all potential shock scenarios. We solve the model numerically for a potential cancer diagnosis for an individual in the United States and assess the impacts of a cancer diagnosis on various aspects of the individual’s life.

Chapter 3 remains in the field of health economics, but switches the perspective from the individual to a policy maker as she faces the problem of containing the spread of an infectious disease across a network. Beside lockdown measures to decrease the transmission rates we introduce testing as a potential policy measure. We derive some analytical insights on the lockdown and testing strategies

and assess the group-specific heterogeneous policy measures within a numerical example to illustrate the capabilities of the framework.

For the numerical solution of the optimal control models in these first two chapters we use a gradient-based approach. We base the algorithm on the work of Veliou (2003) and extend it according to the specific needs of our models (e.g. introduce a solution process for state end-constraints or non-trivial feasible regions of the control variables).

In Chapter 4 we return to decision making on the individual level. In this model we investigate the behaviour of households with respect to risks from natural disasters. We investigate the interaction between different strategies to reduce disaster risk including mitigation measures such as relocation of the settlement location as well as loss reduction efforts like insurance and personal damage protection. Furthermore households can increase their resilience by keeping financial savings to increase their adaptive capabilities after a natural disaster. We approach the numerical solution of this model with well established value function iteration (VFI) techniques. As this method is prone to the curse of dimensionality, substantial computational efforts had to be made to obtain solutions in reasonable time. With consequent Monte-Carlo-simulations we assess the impact of different household characteristics with a special focus on the direct and indirect effects of education on disaster risk behaviour.

Finally, Chapter 5 provides some ultimate conclusions and summarises the findings of this thesis. We also discuss some of limitations of the models and present an outlook for potential extensions and future work in related areas.

Chapter 2

Integrating large health shocks into life-cycle models - Modelling impacts of a cancer diagnosis

This chapter in its entirety directly corresponds to the paper Freiburger, Kuhn and Wrzaczek (2022b) in final preparation for submission to Econometrica. The work received financial support from the FWF within the project “Life-cycle behaviour in the face of large shocks to health” (P 30665-G27).

Abstract

Most of the models on the life-cycle utilization of health care assume that individuals are able to foresee the development of their health perfectly. However, health shocks with significant impact (e.g. severe life-threatening conditions, the onset of chronic disease or accidents) should not be averaged into a mean value, as they have the potential to put the life-course onto a different trajectory. In this paper, we introduce a dynamic optimal control framework incorporating a stochastic health shock with individuals allocating their resources to consumption and different kinds of health care over their life-cycle. We distinguish between general health care and shock specific prevention, acute and chronic care. This setup enables us to analyse how the health risk shapes individual behaviour with respect to the different types of health care and how health shocks change the trajectories of consumption and savings. Newly developed transformation techniques allow us to investigate the optimal decisions made in anticipation of a potential health shock and the optimal reaction to all possible shock scenarios. We are able to obtain analytic expressions for the consumption and health care utilization profiles before and after the shock and identify the driving forces. Furthermore, we extend the value of life concept to other aspects of individual health. Finally, we illustrate our findings by calculating a numerical solution calibrated to an individual facing a potential cancer diagnosis in the US.

2.1 Introduction

Typically, life-cycle models of health behaviour take an ex-ante stance, where a representative individual is subject to the depreciation of a health stock (Grossman, 1972), subject to some mortality process (Ehrlich, 2000; Murphy and Topel, 2006; Hall and Jones, 2007; Kuhn et al., 2015), or subject to the accumulation of deficits (Dalgaard and Strulik, 2014). Individuals fix their life-cycle decisions in perfect anticipation of these processes, regardless of whether they are deterministic or stochastic. In a similar vein, the statistical value of a life is based on an ex-ante evaluation of survival (Shepard and

Zeckhauser, 1984; Rosen, 1988; Murphy and Topel, 2006; Hall and Jones, 2007; Kuhn et al., 2010; Kuhn et al., 2011; Kuhn et al., 2015).

While hugely simplifying analytical complexity, the ex-ante perspective amounts to a gross stylization. In reality, health is not developing according to a smooth process which may be shaped to some extent by health investments, but it is rather subject to smaller or greater shocks. Some of these, such as severe life-threatening diseases (e.g. heart attacks, stroke or cancer) or accidents as well as chronic diseases (e.g. diabetes or dementia), have the potential to put the entire life-course on a different trajectory. Even if individuals are anticipating the risk of such a shock, the optimization problem changes in as far as the shock subjects the individual to different constraints (e.g. a permanently lower income stream due to disability or a persistently higher mortality risk after the shock) and, thus, to a different behavioural regime. In consequence, individuals are also prone to alter their behaviour prior to a health shock by engaging in precautionary saving or in preventive actions. The nature and intensity of such actions is shaped by institutions, such as the availability (or not) of health or disability insurance, as well as by policy interventions, such as the subsidization of preventive behaviour.

In this paper, we detail a life-cycle model, allowing us to study large, singular shocks to health. Thus, we consider a life-cycle model with endogenous health and survival in the spirit of Kuhn et al. (2015) but allow for the onset of a disease or accident at some random time s . The individual is only aware of the risk, modelled as a hazard rate, that can be influenced through prevention. If the health shock materializes at s , the health of the individual is affected, implying that (i) acute life-saving and/or disability-preventing health care may be required at s ; (ii) mortality may be permanently elevated in the course of a chronic disease; (iii) a state of disease/disability emerges that may subsequently affect the individual's utility and earnings; (iv) chronic health care may be required to mitigate the adverse longer-term consequences of the shock. Since we model the life of an individual over time (i.e. age) and since the health shock can set in at any time during the life-course, the health state, assets and the individual's consumption and health care choices are age and duration dependent. This goes well beyond most of the analysis to date.

In order to solve the underlying stochastic optimal control model with a random stopping rule (akin e.g. to a problem considered by Boukas et al. (1990)) we transform the model into a vintage optimal control model, following a novel approach by Wrzaczek et al. (2020). This allows us to analyse in a coherent framework the way in which the pre- and post-shock dynamics of both state and control variables are linked through anticipation and retrospection. In so doing and in contrast to most of the previous literature, we make a clear distinction between preventive health care (lowering the hazard rate directly or indirectly), acute health care (lowering the instantaneous impact of the shock on survival and/or subsequent disability) and chronic health care (lowering the disease/disability state after the shock).

Our analysis allows us to generalize the value of life (VOL) and apply it to a setting with a health shock. Specifically, we derive the value of health before and after the health shock, the value of prevention, the value of surviving the shock, and the value of morbidity. We can also calculate an ex-ante value of health. As it turns out, these values can be used to write the first order conditions for the different forms of health care in a compact form with a straightforward interpretation. Furthermore, the valuations are instrumental in understanding the dynamics of the different forms of health care.

In order to illustrate these dynamics, we employ numerical techniques developed by Veliov (2003) to present how the development of health and survival states depends on both age and duration of the disease (following the shock). We illustrate the age-duration specific dynamics by focusing on an application of our model to the onset of cancer as one major type of a health shock. For this purpose we calibrate our model (to reasonable extent) to reflect the dynamics of survival and health care spending relating to cancer, based on US data.

The economic literature on health-related shocks to the life-cycle is disperse. The largest part of this literature analyses the impact of health-related shocks to labour productivity, to health care expenditure, and to survival. Palumbo (1999), De Nardi et al. (2010), and Kopecky and Koreshkova (2014), for instance, focus on the impact of risky health care expenditure during old age on precautionary saving;

while French (2005) and French and Jones (2011) focus on the impact of health shocks on labour supply and retirement.¹ Capatina (2015) provides an overall assessment of how four channels of health-related risk (productivity, medical expenditure, time endowment, and survival) bear on income inequality and precautionary saving. Common to all of these papers is that they model health risks as a sequence of possibly state-dependent shocks over the life-cycle without emphasizing the scope for "catastrophic" shocks with a propensity to drastically shift the life-cycle trajectory. Furthermore and importantly, they do not provide scope for individuals to reduce the probability of shocks and/or mitigate their consequences by purchasing preventive and/or curative health care, or engaging in other health-related actions. Reichling and Smetters (2015) allow for possibly large shocks when showing that stochastic mortality risks may explain the (rational) abstinence from private annuities when these shocks are correlated with medical expenses. This notwithstanding, the focus of their work is again not on the impact of such shocks on the utilization of health care and resulting health outcomes.²

Few works so far have studied how the risk of health shocks bears on an individual's health investments, either in response to a shock or, more importantly, in anticipation of a shock. Cole et al. (2019) consider a setting where (curative) health expenditure and labour productivity are subject to health shocks, the propensity for which depends on health status. The authors examine how the individual's incentive to engage in preventive efforts aimed at improving their health status are shaped by non-discriminatory health insurance and wage-setting. There is no mortality risk and, as the authors themselves remark in their conclusions, the focus is on small (transitory) shocks to health. In Hugonnier et al. (2013) individual productivity and mortality both depend on an underlying health stock à la Grossman (1972). Individuals can invest into this stock of health, which is assumed to be subject to morbidity shocks following a Poisson process. The key distinction to our approach is that their modelling of uniform health investments and a sequence of (at least principally) transitory shocks (apart from death) does not allow them to discriminate between preventive health care (lowering the probability of a shock) and curative health care (reducing the damage from a shock). Neither do they distinguish between the valuation of preventive care as opposed to the value of curative or chronic care,³ distinctions which will feature prominently in our work. Finally, Hugonnier et al. (2013) apply their model to understand the relationship between financial as opposed to health investments, whereas our focus lies on how the life-cycle allocation of preventive and curative care is shaped by the nature of the health shock. The notion of permanent health shocks is only taken up by Laporte and Ferguson (2007) who consider a version of the Grossman (1972) model in which the health stock is subject to a single irreversible shock, the arrival of which follows a Poisson process. They examine how the nature of this shock bears on the ex-ante path of health investments; however, health is assumed to bear on morbidity and, thus, on period utility but not on survival. Indeed, the length of life is assumed to be exogenous and deterministic.

This leads us to conclude that while the theoretical literature on health shocks has made considerable advances in terms of understanding the consequences of (a sequence of) small shocks for life-cycle patterns of labour supply, income, expenditure, and savings/consumption, comparatively little is known yet about the implications of shocks for the demand for health care and for health behaviours in regard to both their preventive and curative aspect. In particular, this applies to large shocks, such as the onset of severe chronic disease (diabetes, heart disease, cancer) or debilitating accidents, which induce permanent rather than transitory shifts in the mortality, morbidity and income patterns over the remaining life-cycle. These issues lie at the heart of the present work.⁴

The remainder of the paper is structured as follows. The next section contains a description of the

¹The literature on health shocks as motivation for precautionary savings ties in with a large literature on the savings response to (general) life-cycle risks (Eeckhoudt et al. (2005) and Eeckhoudt and Schlesinger (2008))

²Smith and Keeney (2005) examine the valuation of health in a setting where individuals face lotteries over their health and income at distinct phases of their life-cycle. While these lotteries may involve large shocks, the authors do not model the timing of these shocks; nor do they endogenize the health risks.

³The same applies to the models in Picone et al. (1998), Fonseca et al. (2013), Jung and Tran (2016) and Yogo (2016).

⁴From an empirical perspective, the behavioural adjustments and consequences for income and well-being of health shocks have been studied extensively. Adjustments have been studied in regard to saving and consumption (Bíró (2013)) as well as in regard to health behaviours, in particular smoking (e.g. Smith et al. (2001); Khwaja et al. (2006); Marti and Richards (2017)).

model. Section 2.3 presents the analytical solution, involving in particular the derivation of various value (of health) terms from the set of relevant shadow prices (Subsection 2.3.1) and their subsequent employment in the first-order conditions (Subsection 2.3.2). Section 2.4 then proceeds to present the necessary foundations for the numerical analysis, with Subsections 2.4.1 through 2.4.4 setting out data, functional specifications as well as details of the solution strategy. Section 2.5 finally examines the numerical solution of our framework starting with a comparison of data and the calibrated model output in Subsection 2.5.1. Subsections 2.5.2 to 2.5.4 break down the numerical consumption, expenditure, health and survival profiles and provide an exploration of the driving forces behind their behaviour as well as illustrations of the numerical assessment of several distinguished values of health. Section 2.6 concludes.

2.2 Model

In this section we present a framework that integrates a large shock to the health of an individual into a life-cycle model. For the timing of the shock we make the following assumptions (which apply throughout the paper).

- (A1)** A large shock to health occurs at some age s , which is random. The probability rate of arrival is known by the individual.
- (A2)** The event at s occurs only once, thus the life-time of an individual can be separated into a stage before and a stage after s .

These two assumptions allow us to formulate the model as a stochastic optimal control model with a random stopping time (see Boukas et al. (1990)) and analyse it in terms of a vintage optimal control model (see Wrzaczek et al. (2020)). In both stages the individual chooses consumption and different types of health care in order to optimize (expected) life-time utility. Denoting by $t \in [0, T]$ the age of the individual, where $T > 0$ gives the maximum feasible age, we then have that $s \in [0, T]$ is the age at which the model switches from stage 1 (where $t < s$) to stage 2 (where $t > s$). In the following, we introduce both stages of the model and specify the rate at which the shock arrives.

Using the index i ($i = 1, 2$) to denote the life-cycle stage that is referred to, we assume that the survival probability $S_1(t)$ in stage 1 is determined by the stage-1 mortality rate $\mu^1(\cdot)$.⁵ This rate is assumed to be equal to the base mortality rate $\mu^b(t, S_1, b_1)$ which depends on age t , decreases in survival $S_1(t)$, and decreases in the quantity $b_1(t)$ of general health care subject to decreasing returns. Here, $b_1(t)$ is a generic measure of all health care that is unrelated to the condition(s) relating to the health shock. As described in Kuhn et al. (2015), the high correlation between survival and health implies that $S_1(t)$ can be interpreted as a proxy for the health status. We thus capture the negative dependency of mortality on health by including $S_1(t)$ as an argument in μ^b .⁶ Altogether, the survival probability evolves with age according to

$$\begin{aligned} \dot{S}_1(t) &= -\mu^1(t, S_1(t), b_1(t))S_1(t) \\ &= -\mu^b(t, S_1(t), b_1(t))S_1(t), \quad S_1(0) = 1. \end{aligned} \quad (2.1)$$

The individual maximizes expected life-time utility

$$\mathbb{E}_s \left[\int_0^s e^{-\rho t} S_1(t) u^1(c_1(t)) dt + e^{-\rho s} V^*(S_1(s), A_1(s), s) \right]. \quad (2.2)$$

⁵Note that we will omit t wherever it is not of particular importance.

⁶This formulation might be surprising at first sight, but following the explanations in Freiberger and Kuhn (2020) and under some weak assumptions it is consistent with a deficit accumulation model as developed in Schuenemann et al. (2017), or a classic Grossman-type (Grossman, 1972) model with a monotonously depreciating health stock over the life-cycle.

The first term contains the aggregated utility from birth up to s , where the period utility $u^1(c_1(t))$ from consumption $c_1(t)$ is weighted by survival $S_1(t)$ and a discount factor $e^{-\rho t}$ (the discount rate ρ is assumed to be exogenous).⁷ The function $u^1(\cdot)$ is assumed to fulfil the classic assumptions of positive but diminishing marginal utility as well as the Inada conditions if consumption tends to zero or infinity. The second part of the expected utility denotes the discounted aggregated utility of the remaining life-time, given that the individual has suffered a health shock at age s . Here, $V^*(\cdot)$ denotes the optimal value (i.e. the value function) of the optimal control problem in the second stage and not only depends on the age s at the occurrence of the shock itself, but also on the survival/health state $S_1(s)$ and the assets $A_1(s)$ at the point of the shock. The expected value is built with respect to the random variable s . The likelihood of a shock occurring at a given age can be influenced by the individual through investments in preventive care h_1 . Let the probability distribution of s be defined by $\mathcal{F}(t) = \mathbb{P}[s \leq t]$. The probability distribution of s can be characterized by the hazard rate η of the shock, which is generally defined by (2.3) and for which equation (2.4) holds:

$$\eta(t) = \frac{\mathcal{F}'(t)}{1 - \mathcal{F}(t)}, \quad (2.3)$$

$$\mathcal{F}(t) = 1 - e^{-\int_0^t \eta(a) da}. \quad (2.4)$$

More specifically, we assume the hazard rate

$$\eta(t) = \eta(t, S_1(t), h_1(t)) \quad (2.5)$$

to depend on age t , to decrease in survival/health $S_1(t)$, and to decrease in the utilization of preventive care $h_1(t)$ (again subject to diminishing returns). One might think, for instance, of h_1 as the propensity to invest in the vaccination against an infectious disease or as the propensity to attend precautionary screenings for cancer or heart disease.

The asset dynamics in stage 1 follow

$$\dot{A}_1(t) = (r(t) + \bar{\mu}(t))A_1(t) + w^1(t) - c_1(t) - p^b(t)b_1(t) - p^1(t)h_1(t), \quad (2.6)$$

$$A_1(0) = 0 \quad \text{and} \quad A_1(T) = 0; \quad (2.7)$$

where in stage 1 assets $A_1(t)$ are annuitized and generate a return $(r(t) + \bar{\mu}(t))$, with $r(t)$ and $\bar{\mu}(t)$ denoting the interest rate and the mortality risk premium on annuities⁸, respectively; where $w_1(t)$ denote stage-1 earnings; where the price for consumption $c_1(t)$ is normalized to one; and where $p^b(t)$ and $p^1(t)$ denote the prices for general and specific preventive health care, $b_1(t)$ and $h_1(t)$, respectively. As usual we assume zero assets at birth and at the end of the maximum life-span T .

In stage 1, the individual chooses the (non-negative) control variables $c_1(t)$, $b_1(t)$ and $h_1(t)$ so as to maximize the objective function (2.2) subject to the constraints (2.1) and (2.5)–(2.7).

Stage 2 is modelled in a similar vein but we now consider a disease stock $E(t, s)$ as an additional state variable that bears on the individual's utility and constraints. For all variables in the second stage t and s describe the age of the individual and the age at which the health shock has occurred respectively. We assume that the condition that sets in at s may be associated with a specific mortality $\mu^m(t, s, E(t, s))$, depending on age, the time of the shock (or onset of disease), and on the disease stock. In addition, the individual continues to be subject to a base mortality $\mu^b(t, S_2(t, s), b_2(t, s))$ which can be reduced by general health investments $b_2(t, s)$. With the total mortality rate μ^2 of the individual in the second

⁷This formulation reaches back to Yaari (1965).

⁸From an individual point of view the annuity rate is exogenously given. However Section 2.4.2 presents details about the actuarially fair annuity rate within our framework.

stage being the sum of μ^b and μ^m , the dynamics of stage-2 survival $S_2(t, s)$ can be written as

$$\begin{aligned}\dot{S}_2(t, s) &:= \frac{dS_2(t, s)}{dt} = -\mu^2(t, s, S_2(t, s), b_2(t, s), E(t, s))S_2(t, s) \\ &= -[\mu^b(t, S_2(t, s), b_2(t, s)) + \mu^m(t, s, E(t, s))] S_2(t, s).\end{aligned}$$

The disease stock $E(t, s)$ evolves according to

$$\dot{E}(t, s) := \frac{dE(t, s)}{dt} = f(t, s, E(t, s), h_2(t, s)) \quad (2.8)$$

where f depends on age, age at the time of the shock, the disease stock itself, and on disease-specific (chronic) health care $h_2(t, s)$, which is aimed at lowering the disease stock (again subject to diminishing marginal effects). Our general formulation of the disease dynamics allows for a range of different interpretations. These include, in particular, the cases of (i) an accident or acute disease at the point of the shock, which leaves the individual disabled initially but where a natural healing process, supported perhaps by health care, leads to a gradual reduction of $E(t, s)$ (understood to be the extent of disability); and (ii) a progressive disease, such as cancer, diabetes or Alzheimer dementia, where $E(t, s)$ tends to increase unless it is kept in check or lowered by the consumption of health care.

To further account for the negative consequences of the onset of disease (or disability), $E(t, s)$ is assumed to lower stage-2 earnings, w^2 , and stage-2 utility u^2 , i.e.

$$\begin{aligned}\frac{\partial u^2(c_2, E)}{\partial E} &\leq 0 & \frac{\partial^2 u^2(c_2, E)}{\partial E^2} &\leq 0, \\ \frac{\partial w^2(t, s, E)}{\partial E} &\leq 0 & \frac{\partial^2 w^2(t, s, E)}{\partial E^2} &\leq 0.\end{aligned}$$

Finally, we assume the (initial) level of the disease state at the time of the shock $t = s$ to be a decreasing function of the general health state, as proxied by $S_1(s)$, and of one-off (acute) health care $d(s)$, i.e. $E(s, s) = B(S_1(s), d(s))$. We assume that acute care affects initial deficits in addition to the probability of surviving the health shock, as the example of cardiac arrest shows, that effective emergency care also affects the long run consequences of the heart attack (see Hassager et al. (2018)).

The dynamics of stage-2 assets $A_2(t, s)$ are similar to those in stage 1 with the following differences. First, earnings are not exogenous but depend on $E(t, s)$, as detailed above. Second, expenditures for chronic health care, purchased at price $p^2(t)$, substitute for preventive health care. Third, the initial stage-2 assets $A_2(s, s)$ (just after the shock) equal the stage-1 assets $A_1(s)$ (just before the shock) net of the expenditure for acute care which is purchased at a price $p^d(s)$. Thus,

$$\dot{A}_2(t, s) = (r(t) + \bar{\mu}(t))A_2(t, s) + w^2(t, s, E(t, s)) - c_2(t, s) - p^b(t)b_2(t, s) - p^2(t)h_2(t, s) \quad (2.9)$$

$$A_2(s, s) = A_1(s) - p^d(s)d(s) \quad \text{and} \quad A_2(T, s) = 0. \quad (2.10)$$

According to the last boundary condition, assets have to equal 0 at the end of life regardless of when the shock has occurred. The aggregated utility during stage 2 consists of the present value of the expected (i.e. survival weighted) utility stream over the remaining life-course

$$P(S_1(s), d(s)) \cdot \int_s^T e^{-\rho t} S_2(t, s) u^2(c_2(t, s), E(t, s)) dt, \quad (2.11)$$

where $P(S_1(s), d(s)) \in [0, 1]$ is the probability that the individual survives the health shock, which similar to the initial value of the disease stock increases in the level of the stage-1 health state $S_1(s)$ and the quantity of acute health care $d(s)$ (subject to diminishing returns). Note that $P(\cdot) < 1$ captures the potential that the individual does not survive the shock (e.g. an accident, a cardiac event or a stroke),

whereas $P(\cdot) = 1$ would reflect a disease that is not mortal upon its onset at s but only potentially so over time (e.g. cancer, diabetes, Alzheimer's disease). Finally, our model also embraces the case of health shocks leading to instantaneous death, $P(\cdot) = 0$. From now on, we will refer to $P(\cdot)$ as the "continuation probability".

The complete model is summarized in Equations (2.12) - (2.20).

$$\max_{c_1(t), h_1(t), b_1(t) \geq 0} \mathbb{E}_s \left[\int_0^s e^{-\rho t} S_1(t) u^1(c_1(t)) dt + e^{-\rho s} V^*(S_1(s), A_1(s), s) \right] \quad (2.12)$$

$$\dot{S}_1(t) = -\mu^1(t, S_1(t), b_1(t)) S_1(t), \quad (2.13)$$

$$\dot{A}_1(t) = (r(t) + \bar{\mu}(t)) A_1(t) + w^1(t) - c_1(t) - p^b(t) b_1(t) - p^1(t) h_1(t), \quad (2.14)$$

$$S_1(0) = 1, \quad A_1(0) = 0, \quad A_1(T) = 0 \quad (2.15)$$

where

$$V^*(S_1(s), A_1(s), s) := \max_{\substack{c_2(t, s), h_2(t, s), \\ b_2(t, s), d(s) \geq 0}} P(S_1(s), d(s)) \cdot \int_s^T e^{-\rho t} S_2(t, s) u^2(c_2(t, s), E(t, s)) dt \quad (2.16)$$

$$\dot{S}_2(t, s) = -\mu^2(t, s, S_2(t, s), b_2(t, s), E(t, s)) S_2(t, s) \quad (2.17)$$

$$\dot{A}_2(t, s) = (r(t) + \bar{\mu}(t)) A_2(t, s) + w^2(t, s, E(t, s)) - c_2(t, s) - p^b(t) b_2(t, s) - p^2(t) h_2(t, s) \quad (2.18)$$

$$\dot{E}(t, s) = f(t, s, E(t, s), h_2(t, s)) \quad (2.19)$$

$$S_2(s, s) = S_1(s), \quad A_2(s, s) = A_1(s) - p^d(s) d(s), \quad A_2(T, s) = 0, \quad E(s, s) = B(S_1(s), d(s)) \quad (2.20)$$

Problem (2.12)-(2.15) can be interpreted as an optimal control model with random stopping time (see Boukas et al. (1990)). For the analysis and for the numerical solution we transform the model into a vintage optimal control model, as this offers additional economic insights as well as an established numerical solution method (see Veliov (2003)).⁹ For the theoretical background and other examples of the transformation method we refer to Wrzaczek et al. (2020). As the presentation of the model in vintage optimal control form is not immediately instructive, we relegate it to appendix 2.7.1. Here, we only note that the vintage formulation implies that all second-stage variables are indexed by both age t and the time of the shock s , which can be interpreted as the arrival-date of a (potential) vintage of the remaining life-course (in disease). Indeed, the notation we have introduced earlier meets this criterion.

The transformation includes the introduction of the following two auxiliary variables, which we will subsequently employ in our calculations and for which interpretations are straightforward. First, $Z_1(t)$ denotes the probability, that an individual has not suffered a health shock up to age t . We will also relate to $Z_1(t)$ as the survival in good health¹⁰. As described above, the arrival rate of the health shock $\eta(t, S_1(t), h_1(t))$ depends on age, health status and preventive health care. This implies that the development of Z_1 can be formulated through the following differential equation

$$\dot{Z}_1(t) = -\eta(t, S_1(t), h_1(t)) Z_1(t) \quad , \quad Z_1(0) = 1. \quad (2.21)$$

Second, we need the auxiliary variable $Z_2(s)$, which is defined by

$$Z_2(s) = Z_1(s) \eta(s, S_1(s), h_1(s)) P(S_1(s), d(s)), \quad (2.22)$$

and can be interpreted as the joint probability of experiencing *and* surviving a shock at age s .¹¹

⁹The solution process is, nevertheless, not trivial as Veliov (2003) sets up a general framework, requiring multifaceted adaptations for the solution of specific problems.

¹⁰It directly holds that $Z_1(t) = 1 - \mathcal{F}(t)$.

¹¹The term probability is not precise as we are analysing a time-continuous model and, strictly speaking, $Z_2(s)$ defines

For further reference, Table 2.1 summarizes the control and state variables in the two life-cycle stages.

<i>Control variables</i>	<i>Stage 1</i>	<i>Stage 2</i>	<i>Shock time s</i>
Consumption	$c_1(t)$	$c_2(t, s)$	-
General Health investments	$b_1(t)$	$b_2(t, s)$	-
Prevention expenditures	$h_1(t)$	-	-
Chronic care	-	$h_2(t, s)$	-
Acute treatment	-	-	$d(s)$
<i>State variables</i>			
Survival probability	$S_1(t)$	$S_2(t, s)$	$S_2(s, s) = S_1(s)$
Assets	$A_1(t)$	$A_2(t, s)$	$A_2(s, s) = A_1(s) - p^d(s)d(s)$
Survival in good health	$Z_1(t)$	-	-
Joint “probability” of shock <i>and</i> survival at s	-	$Z_2(s)$	$Z_2(s) = P(S_1(s), d(s))Z_1(s) \times \eta(s, S_1(s), h_1(s))$
Severity of health deficits	-	$E(t, s)$	$E(s, s) = B(S_1(s), d(s))$

Table 2.1: Summary of all state and control variables in the basic framework

2.3 Analytical results

For the vintage optimal control model, we apply the Maximum Principle presented in Feichtinger et al. (2003) to arrive at a set of necessary optimality conditions. Specifically, we obtain

- a set of differential equations describing the dynamics of the adjoint variables;
- a set of transversality conditions for the adjoint variables, corresponding to state variables with a free endpoint;
- a set of first-order optimality conditions for all control variables at every point in time/age, and for the stage-2 control variables for every possible point in life at which the health shock can occur.

These conditions, together with the state equations (with initial and boundary conditions), are used to find the optimal solution. In the following section, we will employ the first order conditions and the adjoint variables (together with the corresponding differential equations) to identify and characterize the various behavioural channels of the model.

2.3.1 Valuations of Health

Following Rosen (1988), Murphy and Topel (2006), Hall and Jones (2007), Kuhn et al. (2015) and others we investigate the individual willingness to pay for changes in health. However, while the original works focus on a reduction in the mortality risk, i.e. the value of life (VOL), we distinguish between the willingness to pay for changes in a range of different aspects of health.

Definition 1

For the analysis of problem (2.12) - (2.20), we define the following valuations of health, which we will use throughout. Here, \mathcal{V} denotes the value function of problem (2.12)-(2.15).

the probability density function of s multiplied with the continuation probability P . However the term “probability” makes for more intuitive reading.

Value of health ψ_H^i in stage $i = 1, 2$: Willingness to pay for a reduction in the mortality rate (depreciation rate of the survival stock) μ^i in stage $i = 1, 2$. $\psi_H^i := \left(-\frac{dV}{d\mu^i}\right) / \left(\frac{dV}{dA_i}\right)$.

Value of prevention ψ_P : Willingness to pay for a reduction in the risk of a health shock by reducing η , $\psi_P := \left(-\frac{dV}{d\eta}\right) / \left(\frac{dV}{dA_1}\right)$.

Value of acute survival ψ_{AS} : Willingness to pay for an increase in the probability P of surviving the shock, $\psi_{AS} := \left(\frac{dV}{dP}\right) / \left(\frac{dV}{dA_2}\right)$.

Value of morbidity ψ_M : Willingness to pay for a reduction in the disease/disability stock E , $\psi_M := \left(-\frac{dV}{dE}\right) / \left(\frac{dV}{dA_2}\right)$.

In addition, it is convenient to define the value of second-stage life as

$$\psi_{life}^2(t, s) = \int_t^T R(t, \tau) \frac{u^2(\tau, s)}{u_{c_2}^2(\tau, s)} d\tau, \quad (2.23)$$

with

$$R(t, \tau) := \exp\left(-\int_t^\tau r(\tau') + \bar{\mu}(\tau') d\tau'\right),$$

i.e. as the present value at age t of the stream of consumer surplus over the remaining life-course in stage 2, with the return on annuities being applied as discount rate. Note the similarity to the "conventional" value of life in earlier works (e.g. Kuhn et al., 2015).

Based on this, Proposition 1 presents an explicit analytical formulation of the valuations defined above.

Proposition 1

Assume the existence of optimal trajectories of consumption and (the various) health investments in both stages of the individual life-cycle model (2.12)-(2.20) together with an interior solution for the consumption profiles c_1 and c_2 . The valuation terms in Definition 1 can then be written as follows.¹²

Stage-2 valuations:

$$\psi_H^2(t, s) = \int_t^T R_H^2(t, \tau, s) \frac{u^2(\tau, s)}{u_{c_2}^2(\tau, s)} d\tau, \quad (2.24)$$

$$\psi_M(t, s) = \int_t^T R_M^2(t, \tau, s) \left\{ \mu_E^2(\tau, s) \psi_H^2(\tau, s) - w_E^2(\tau, s) - \frac{u_E^2(\tau, s)}{u_{c_2}^2(\tau, s)} \right\} d\tau. \quad (2.25)$$

Stage-1 valuations:

$$\psi_{AS}(t) = \frac{\psi_{life}^2(t, t)}{P(t)}, \quad (2.26)$$

$$\begin{aligned} \psi_H^1(t) = \int_t^T R_H^1(t, \tau) \left\{ \frac{u^1(\tau)}{u_{c_1}^1(\tau)} - \eta_{S_1}(\tau) S_1(\tau) \psi_P(\tau) + \eta(\tau) P(\tau) \frac{u_{c_2}^2(\tau, \tau)}{u_{c_1}^1(\tau)} \times \right. \\ \left. \times [\psi_H^2(\tau, \tau) + P_{S_1}(\tau) S_1(\tau) \psi_{AS}(\tau) - S_1(\tau) B_{S_1}(\tau) \psi_M(\tau, \tau)] \right\} d\tau, \quad (2.27) \end{aligned}$$

$$\psi_P(t) = \int_t^T R_P^1(t, \tau) \left\{ \frac{u^1(\tau)}{u_{c_1}^1(\tau)} + \eta P \frac{u_{c_2}^2(\tau, \tau)}{u_{c_1}^1(\tau)} \psi_{life}^2(\tau, \tau) \right\} d\tau - \frac{u_{c_2}^2(t, t)}{u_{c_1}^1(t)} P \psi_{life}^2(t, t). \quad (2.28)$$

¹²For notational convenience we omit the state and control arguments in all functions and just indicate at which time-point the state and control values should be evaluated, e.g. $u_{c_2}^2(\tau, s) \equiv u_{c_2}^2(c_2(\tau, s), E(\tau, s))$.

The various discount factors are defined by

$$R_P^1(t, \tau) := R(t, \tau) \exp \left(- \int_t^\tau \eta(\tau') P(\tau') \frac{u_{c_2}^2(\tau', \tau')}{u_{c_1}^1(\tau')} d\tau' \right), \quad (2.29)$$

$$R_H^1(t, \tau) := R(t, \tau) \exp \left(- \int_t^\tau \mu_{S_1}^1(\tau') S_1(\tau') + \eta(\tau') P(\tau') \frac{u_{c_2}^2(\tau', \tau')}{u_{c_1}^1(\tau')} d\tau' \right), \quad (2.30)$$

$$R_H^2(t, \tau, s) := R(t, \tau) \exp \left(- \int_t^\tau \mu_{S_2}^2(\tau', s) S_2(\tau', s) d\tau' \right), \quad (2.31)$$

$$R_M^2(t, \tau, s) := R(t, \tau) \exp \left(\int_t^\tau f_E(\tau', s) d\tau' \right). \quad (2.32)$$

The proof is technical and thus relegated to the Appendix 2.7.3. Although the above expressions look involved, the individual terms can be assigned to the forward and backward incentives that shape the qualitative behaviour of the optimal solution. Before discussing the economic interpretations of all valuations in detail we want to stress, that the prerequisites of Proposition 1 are rather weak. Besides the assumption of existence of an optimal solution, it is sufficient to have interior solutions for the consumption paths to be able to derive the presented terms. Consequently all valuations still hold in the absence of some control variables (shocks may only require some of the controls for a proper modelling) or for cases with boundary solutions concerning any types of health investments.

Value of health in the second stage: $\psi_H^2(t, s)$

The value of health in the second stage ψ_H^2 (see equation (2.24)) is closely related to the value of life, i.e. the discounted value of consumer surplus over the remaining life-cycle, as defined in equation (2.23). The only difference is that the value of health takes into additional account a term $\mu_{S_2}^2 S_2$ in the discounting function R_H^2 , reflecting that better health, as measured by S_2 , contributes to lower mortality over the remaining life-course. Given that $\mu_{S_2}^2 < 0$, the new term decreases the discount factor and, thereby, raises the value of health over and above the conventional value of life. Intuitively, this reflects the additional value of health/survival as an asset that by lowering mortality yields a return in terms of additional consumer surplus over the remaining life-course.

Value of acute survival: $\psi_{AS}(t, s)$

The value of marginally increasing the continuation probability, ψ_{AS} , (see equation (2.26)) at the time of the shock directly corresponds to the conventional value of life in the second stage, ψ_{life}^2 .¹³ The weighting of ψ_{life}^2 by $1/P(t)$ implies that the willingness to pay for acute survival increases with the risk of not surviving the health shock.

Value of morbidity: $\psi_M(t, s)$

The willingness to pay for a (marginal) reduction in the disease/disability stock, ψ_M , (see equation (2.25)) depends on the interaction of three effects: (i) the first part $\mu_E^2 \psi_H^2$ captures the impact of the disease/disability stock on mortality or, equivalently, on the depletion of the health stock. A marginal increase in the disease stock increases the mortality risk by μ_E^2 , which consequently leads to a change in second-stage survival, S_2 , which is valued at the willingness to pay ψ_H^2 . (ii) the part $-w_E^2$ contains the impact of the disease/disability on earnings, which directly translates into a willingness-to-pay to lower the disease/disability stock. (iii) The last part $-\frac{u_E^2}{u_{c_2}^2}$ measures the reduction in consumer surplus in the presence of disease/disability, or equivalently the marginal rate of substitution between consumption and degree of illness/disability. The greater the marginal impact of disability on utility, the greater the willingness-to-pay for reductions in morbidity. Finally, we have to adjust the standard discount rate R by accounting for the direction and speed of disease progression, as measured by the

¹³The discounting term $\mu_{S_2}^2 S_2$ is not present here, as a marginal change in $P(t)$ has no effect on survival S_i .

impact f_E of the disease stock on its own accumulation. Thus, the value of future changes in morbidity tends to be discounted more heavily if the disease is accelerating, i.e. $f_E > 0$, and less heavily if it is decelerating, i.e. $f_E < 0$. Intuitively, this suggests that reductions in morbidity are more valuable in the future (present) if the disease is accelerating (decelerating).

Value of prevention: $\psi_P(t)$

The willingness to pay for a reduction in the hazard rate of the health shock, ψ_P , (see equation (2.28)) is equivalent to the net value of remaining in (the healthy) stage 1 as opposed to transiting into (the diseased/disabled) stage 2. Accordingly, the integral term measures the value of remaining in stage 1, which in itself is composed of two distinct factors. The first part $\frac{u^1(c_1(\tau))}{u^1_{c_1}(c_1(\tau))}$ amounts to the consumer surplus for each year that continues to be spent in stage 1, whereas the second part adds the expected value of stage 2 utility should a transition occur at rate $\eta(\tau)$ in some future year $\tau > t$. Note that this value corresponds to the stage-2 value of life, ψ_{life}^2 , weighted with the probability $P(\tau)$ of surviving a health shock at τ . As ψ_{life}^2 is counted in units of stage-2 consumption, a conversion into units of stage-1 consumption takes place by multiplication with $\frac{u^2_{c_2}(t,t)}{u^1_{c_1}(t)}$. Notably, the discount factor R_P^1 applied to the utility stream associated with remaining in stage 1 through t now takes into account the (weighted) risk of a transition into stage 2.

The value of remaining in stage 1 (integral part in equation (2.28)) is then offset against the value of switching to stage 2 at t (last term in (2.28)). This value corresponds to the stage-2 value of life, again weighted with the survival probability, $P(t)$, at the time of the shock and converted into stage-1 values. Note here that the value of avoiding a deadly health shock, for which $P(\tau) = 0$ for all $\tau \in [t, T]$, exactly corresponds to the conventional stage-1 value of life.

Value of health in the first stage: $\psi_H^1(t)$

After having introduced all other valuations of health we can finally analyse the value of health in the first stage ψ_H^1 , which contains multiple terms presented above (see equation (2.27)). In total we separate five distinct impact channels. (i) The stream of stage-1 consumer surplus, $\frac{u^1}{u^1_{c_1}}$. (ii) The value of health/survival in reducing the hazard rate and, thus, preventing the shock, $\eta_{S_1} S_1 \psi_P$. The remaining three parts capture the value of stage-1 health for reaching and living through a stage 2 life-cycle conditional on surviving a shock at age τ . Thus, all three factors are weighted with $\eta P \frac{u^2_{c_2}}{u^1_{c_1}}$, the joint probability of experiencing and surviving the health shock at τ as well as with the conversion factor. (iii) The term ψ_H^2 measures the direct value of health upon entering stage 2; (iv) the term $P_{S_1} S_1 \psi_{AS}$ captures the value of stage-1 health in enhancing acute survival following a shock; and (v) the term $B_{S_1} S_1 \psi_M$ captures the value of stage-1 health in lowering the intensity of disease/disability and, thus, morbidity at the point of the shock. We conclude by noting that the discount factor R_H^1 includes the long run impacts of survival on future mortality $\mu_{S_1}^1 S_1$ like R_H^2 as well as the (weighted) risk of entering stage 2 upon survival of a shock, $\eta P \frac{u^2_{c_2}}{u^1_{c_1}}$.

Ex-ante value of health

From an ex-ante stance the future development of the individual's health is stochastic. Thus the value of health in its general form, comparable to the well-known value of life in other contributions, should account for this uncertainty. Hence we define the (ex-ante) value of health as the expected value of the different values of health in stages 1 and 2, weighted with the corresponding probabilities. This is summarized in the following definition.

Definition 2

Assuming the existence of optimal trajectories of consumption and (the various) health investments in both stages of the individual life-cycle model (2.12)-(2.20), the (ex-ante) value of health can be defined

as

$$\Psi_H(t) := Z_1(t)\psi_H^1(t) + \int_0^t Z_2(s)\psi_H^2(t,s)ds. \quad (2.33)$$

Following from this definition, the ex-ante value of health at age t can also be interpreted as an averaged value of health across individuals who have not experienced a shock up to age t and individuals who, at age t , have experienced different stages of disease progression following a shock experienced at an earlier age $s < t$.

Note that such a measure is useful when it comes to assessing the value of health at population level. In particular for normative purposes, it is considered unethical to distinguish individuals according to their value of life. Thus, the value of life is typically averaged across income strata, health states and often age (for an exception see Aldy and Viscusi, 2008). The ex-ante value of health would provide a value that is averaged across the possible health states at age t , including the potential health-driven inequality in earnings.

2.3.2 First order optimality conditions

As one crucial part of the system of optimality conditions presented in Appendix 2.7.2, the first-order optimality conditions give insight into the economic trade-offs between the different control variables. Using the valuations of health presented in the previous section, we can formulate these FOCs in a compact and intuitive way.

Proposition 2

Assume the existence of optimal trajectories of consumption and (the various) health investments in both stages of the individual life-cycle model (2.12)-(2.20) together with an interior solutions for the (various) choices of health care. The first-order optimality conditions can then be written as follows.

$$\text{Stage 1:} \quad [-\mu_{b_1}^1(t)] \cdot \psi_H^1(t) = p^b(t) \quad (2.34)$$

$$[-\eta_{h_1}(t)] \cdot \psi_P(t) = p^1(t) \quad (2.35)$$

$$\text{Stage 2:} \quad [-\mu_{b_2}^2(t,s)] \cdot \psi_H^2(t,s) = p^b(t) \quad (2.36)$$

$$[-f_{h_2}(t,s)] \cdot \psi_M(t,s) = p^2(t) \quad (2.37)$$

$$\text{At the time of shock } s: \quad [-B_d(s)] \cdot \psi_M(s,s) + P_d(s) \cdot \psi_{AS}(s) = p^d(s) \quad (2.38)$$

Consequently, the optimal allocation of health care involves that for each age/point in time t and for every possible onset of the shock s , the unit price for each type of health care equals the corresponding marginal benefit, consisting of the respective marginal effectiveness and the respective valuation of the health dimension involved. Thus, the price $p^b(t)$ for general first-stage health care, $b_1(t)$, has to equal its marginal impact on mortality ($-\mu_{b_1}^1$) (tantamount to the depreciation rate of health) multiplied with the first-stage value of health, ψ_H^1 . The interpretation for the other types of health care is analogous. Furthermore, we note that the marginal benefits of acute care, d , consist of the sum of two separate terms, as acute care does not only (potentially) increase the chances of surviving the shock but also (potentially) reduce the initial level of morbidity.

The FOCs provide immediate and intuitive information on the (relative) drivers of health care choices. Thus, the individual will demand a higher quantity of health care if it is more effective, if it has a higher value or it has a lower price. Note that the FOCs can also be read as reflecting the optimal trade-off between the different types of health care and consumption. Here, the left-hand side (LHS) of each condition reflects the marginal rate of substitution between the particular type of health care and consumption, whereas the right-hand side (RHS) gives the price ratio, with the price of the consumption good normalized to one. Thus, a higher price for health care would either have to be offset by an increase

in effectiveness and/or an increase in the value of this care, the latter being reflective of greater need. The system of FOCs also allows to trace the allocation across different types of health care in light of relative effectiveness and relative valuations. Consider e.g. the condition

$$\frac{-B_d(s)}{-f_{h_2}(s, s)} + \frac{P_d(s) \cdot \psi_{AS}(s)}{[-f_{h_2}(s, s)] \cdot \psi_M(s, s)} = \frac{p^d(s)}{p^2(s)}$$

as implied by the FOCs (2.37) and (2.38). To understand the intuition, assume first a setting in which the health shock does not impose a risk to survival, implying there is no role for acute care in enhancing acute survival, i.e. $P = 1$ and $P_d(s) = 0$, as would be the case e.g. with cancer. In such a case the condition would tell us that a higher price for acute care (e.g. immediate surgery) as opposed to chronic care (e.g. pharmaceutical therapy) at the point of the onset of the disease would need to be offset by greater effectiveness of acute care in containing the disease (i.e. the progression of cancer). If acute care also bears on survival, the price ratio does not only reflect differences in the effectiveness of care in curbing morbidity but, in addition, the marginal rate of substitution between the (valued) change in acute survival for the acute treatment and the change in morbidity for the chronic treatment. Thus, a higher price for acute care at the point of the shock is supported to the extent that it not only reduces initial morbidity but also improves survival chances by $P_d(s) > 0$. Note that this argument extends to settings, where the provision of acute care, e.g. cancer surgery, may carry a risk to survival, such that $P_d(s) < 0$. Here, the survival risk lowers the willingness to pay for acute care, implying that its utilization is lower relative to chronic care for a given price ratio and/or a given level of utilization is supported only at a lower price.¹⁴ Similar trade-offs between other dimensions of health care, e.g. between general health care and preventive health care in stage 1 or between preventive health care in stage 1 and chronic health care in stage 2, can be constructed by appropriate combination of the relevant FOCs.

Isolating the value terms on the LHS of the FOCs, as e.g. in $\psi_P(t) = \frac{p^1(t)}{[-\eta_{h_1}(t)]}$ for preventive care or $\psi_M(t, s) = \frac{p^2(t)}{[-f_{h_2}(t, s)]}$ for chronic care allows us to interpret the FOCs in terms of the underlying dimensions of health as a final good rather than health care as an intermediate good. Thus, we find that for an optimal allocation, the value of prevention should equal the effective price of prevention, as given by the price of preventive health care adjusted for its effectiveness in curbing the arrival rate of a shock. Similarly, the value of (reducing) morbidity should equal the effective price of lowering morbidity, as given by the price of chronic health care adjusted for its effectiveness in curbing or reverting the progression of the disease. Analogous expressions can be derived for other dimensions of health care. Following Frankovic et al. (2020) who undertake this analysis in the context of survival, we can infer that medical progress that raises the effectiveness of a certain type of health care, prevention say, may be associated with a decline in the value of this dimension of health. Notably, this reflects the greater consumption of preventive health care leading to reductions in the health risk to a level for which any further reduction is less valuable (or in micro-economic terms, the decline in the effective price of prevention relative to consumption is associated with a decline in the marginal rate of substitution between prevention and consumption). We conclude with the following empirical observation: In many practical settings, the valuation of different types of health is difficult to observe. A revealed preference argument would then suggest that the willingness to pay equals the effectiveness-adjusted price of health care, which can be calculated on the basis of observable nominal prices of health care and scientific evidence on medical effectiveness.

Given the multi-dimensionality of health, we can extend the previous argument to examine the relative valuation of different types of health care. Drawing on the FOCS (2.36) and (2.37), for instance, we can write

$$\frac{\psi_M(t, s)}{\psi_H^2(t, s)} = \frac{p^2(t) / [-\mu_{b_2}^2(t, s)]}{p^b(t) / [-f_{h_2}(t, s)]}.$$

¹⁴Recall here our assumption that all controls exhibit diminishing returns.

Thus, the value of lower morbidity relative to the value of survival can be understood to reflect the relationship between the effective price of morbidity relative to the effective price of survival. This has important repercussions from a practical point of view, where many studies in medical evaluation seek to probe into consumer/patient assessment of improvements to the health-related quality of life relative to survival (see e.g. Rowen and Brazier, 2011:for a survey). While such studies are often carried out in the context of an abstract decision-framework, we note that when resulting from market assessments (e.g. when asking expert physicians to state these trade-offs), differences (across individuals, populations or over time) in the relative valuations of health may be as reflective of differences in the underlying preferences as of differences in the relative effectiveness of the different types of health care. Here, again revealed preference analysis may be brought to bear to deduce relative valuations.

2.4 Numerical analysis: an application to cancer

Within the next two sections, we illustrate insights from our model by conducting a calibration exercise and calculating optimal individual behaviour numerically. To provide a concrete context, we study the impact of a (potential) cancer diagnosis on the life-cycle allocation of health care and consumption.

In order to avoid adding the substantial complexity that is involved with the hidden progression of cancer up to the point of a diagnosis, we assume that (i) a diagnosis at age/time s coincides with the onset of cancer. Thus, we disregard the building-up of the cancer-stock prior to s , which is a reasonable assumption when assuming that a diagnosis coincides with the point of first symptoms. As there is no clear empirical correlation¹⁵ between the stage of cancer at the time of diagnosis and the general health status (or the respective age), we suppose (ii) that the individual enters with a constant positive disease stock $E(s, s) = \phi_0 > 0$ independent of its stock of health, $S_1(s)$, at the point of diagnosis. While it is further reasonable to assume that (iii) a diagnosis at s is not associated with an acute risk to survival, such that $P(S_1, d) \equiv 1$, we also assume for the purpose of this analysis that (iv) there is no role of acute care at the time of diagnosis, i.e. $d(s) \equiv 0$. This is justified when assuming that the diagnosis is early enough to rule out a significant role for surgery. To avoid the complexity involved with the timing of diagnosis, we abstract from (v) screening measures. Finally, in order to limit the number of states and controls we assume (vi) that healthy and unhealthy behaviours such as smoking, drinking, eating habits or exercising that affect the risk of cancer are captured by "generic" investments in health, b_1 . Note that these feed into the hazard rate $\eta(t, S_1(t), h_1(t))$ through their impact on the health stock/survival and thereby generate health benefits beyond cancer, a reasonable assumption. We, thus, assume (vii) that there is no further role for specific cancer prevention, such that $h_1(t) \equiv 0$. We also assume that age does not have a direct bearing on the hazard rate, implying that we work with $\eta(S_1(t))$.

Altogether, we obtain a slightly reduced model, which nevertheless enables us to analyze the impacts of cancer on the utilization of health care before and after the diagnosis, including a detailed characterization of the utilization pattern of chronic care, which from now on we refer to as cancer care.

2.4.1 Data sources

We calibrate our model to the data of an individual in the United States in the 2010s, using a number of data sources for the input parameters.¹⁶ We assume that an individual receives earnings from age 20, which is also the age at which the individual begins to make life-cycle decisions.

Input parameters

One of the few fully exogenous inputs is the base earnings profile $\{w_1(t) | t \in [20, 110]\}$, which in our

¹⁵Goodwin et al. (1986) find a positive correlation between the stage of cancer and the age at diagnosis for some types of cancer, while for other types this correlation turns out to be negative. Even for breast cancer alone the results turn out to be unclear. While Satariano et al. (1986) and Mandelblatt et al. (1991) find a positive relationship between age and severity, Yancik et al. (2001) find it to be non-significant.

¹⁶Due to limitations in data availability we are unable to use all data from the same year.

model we assume to be the average earnings age-profile for the US in 2011, as taken from the National Transfer Accounts (NTA) database¹⁷. We account for the fact, that there are no public pensions in our model and all health care expenditures are paid out of pocket, by adding the net health and pension transfer profiles (which are also contained in the NTA database) to the base working income to obtain a profile of disposable income.

We abstain from trying to fit the (non-health) consumption profile from the NTA database, as it shows the typical hump-shape over the life-cycle. Hansen and İmrohorođlu (2008) have shown that such a pattern can only be explained consistently within a life-cycle model when assuming that annuity markets are absent (or severely incomplete).¹⁸ However, we quite deliberately prefer to model a complete annuity market for the explicit purpose of establishing a flat consumption profile as a benchmark against which to assess the impact of health shocks on consumption and identify the underlying channels. Here, an imperfect annuity market would obscure some of the effects. This notwithstanding, we aim for a level of consumption that is in line with the average consumption over the life-cycle.¹⁹

In further pursuit of establishing a flat benchmark consumption profile, we opted to set two of the other exogenous inputs, the interest rate $r(t)$ and the discount rate ρ , both equal to 3%. If we eliminated the possibility of a health shock in our model, this would then lead to a completely flat consumption profile, with the annuity rate perfectly covering the mortality risk in this scenario. Therefore $r(t) = \rho$ enables us to directly identify the impact of the existence of a potential health shock in the consumption profile before and after the cancer diagnosis.

Calibration goals

We use health expenditure data from the NTA-Database (available for the year 2011) and combine it with age-specific information about the share of cancer specific health care expenditures from `healthdata.org`²⁰ to construct age profiles for (i) the general (non cancer) health expenditure and (ii) cancer specific expenditures. We compare these against the average general health expenditure and average cancer care expenditure profiles generated by our model and aim to match the latter with their empirical counterparts.²¹

Furthermore, we use the age-specific mortality profile for the US in 2011 from the human mortality database²² as a basis for establishing survival profiles. Using the mortality rate directly to calculate the corresponding survival profile within our model, we obtain the equivalent of the average survival $\bar{S}(t) := Z_1(t)S_1(t) + \int_0^t Z_2(s)S_2(t,s)ds$ in our model. To obtain the appropriate data against which to compare first-stage survival in our model, we take age-specific cancer mortality rates from the SEER-database²³ and subtract them from the corresponding general mortality rates. From the resulting age-profile of non-cancer mortality we construct a cancer-free survival profile as the appropriate comparison for the S_1 profile in the model.

We then employ the cancer-free survival profile from the data together with age-specific rates of cancer incidence²⁴, again taken from the SEER-database, to calibrate the hazard rate $\eta(S_1)$, which we assume to depend only on the survival state. Finally, we use information from the SEER-database about cancer-specific survival depending on the duration after the cancer diagnosis to calibrate the cancer

¹⁷ www.ntaccounts.org, see Lee and Mason (2011) and United Nations (2013) for details.

¹⁸ From the Euler-equation (2.42) we see that the absence of annuities ($\bar{\mu} = 0$) implies that increasing mortality μ^i with age ultimately leads to a decline in consumption. Annuities eliminate this risk (or, as in our case, even overcompensate it), so there is no significant force anymore that would shift consumption to younger ages, implying a hump-shaped pattern cannot be obtained.

¹⁹ The US exhibit a significant life-cycle deficit, implying that life-time consumption is significantly higher than life-time income. Thus, the raw data contradicts our assumption of zero assets at the end of life. To fulfil the equivalent condition of life-time consumption and life-time income being equal, we have decided to raise income by 15% for our calibration.

²⁰ Institute for Health Metrics and Evaluation (IHME) (2016) (Accessed 2020-01-13)

²¹ We define average health expenditure at age t as $\bar{b}(t) := Z_1(t)b_1(t) + \int_0^t Z_2(s)b_2(t,s)ds$ and, similarly, average cancer care expenditure as $\bar{h}_2(t) := \int_0^t Z_2(s)h_2(t,s)ds$.

²² *Human Mortality Database* (accessed 2019-10-04)

²³ *SEER*Explorer: An interactive website for SEER cancer statistics [Internet]*

²⁴ We take average incidence rates between 2012 and 2016.

specific mortality rates for four different (rough) age-groups over the first ten years after the diagnosis. For further details about the estimation and calibration strategy for the general and cancer-specific mortality rates (μ^b and μ^m) and the cancer incidence rate (η) we refer to appendix 2.7.6.

2.4.2 Remarks on the annuity market

In most models annuity markets are assumed to be either absent or complete with the equilibrium annuity rate being equal to the mortality rate. Our setting is more complex, as it contains two mortality rates reflecting different health regimes: one with cancer and one without. Thus the rate of return for annuities $\bar{\mu}(t)$ crucially depends on the way the annuity market is structured.

In this calibration exercise we assume that the insurer has no information about whether or when an individual has been diagnosed with cancer. As a result the annuity rate at time t turns out to be the expected (or averaged in a population context) mortality rate, which can be calculated as

$$\bar{\mu}(t) = \frac{Z_1(t)(-\dot{S}_1(t)) + \int_0^t Z_2(s)(-\dot{S}_2(t, s))ds}{Z_1(t)S_1(t) + \int_0^t Z_2(s)S_2(t, s)ds} = \frac{Z_1 S_1 \mu^1 + \int_0^t Z_2 S_2 \mu^2 ds}{Z_1 S_1 + \int_0^t Z_2 S_2 ds}. \quad (2.39)$$

The nominator adds up all deaths across the two groups with and without cancer and relates them to the total population in the denominator.²⁵

2.4.3 Solution strategy

The numerical solution of a two-stage optimal control problem with random switching time is a far-from-trivial problem. As already indicated in the problem introduction the transformation into a vintage-structured optimal control problem is the first step in the solution process. The transformation (see Wrzaczek et al. (2020) for further details) allows our numerical strategy to rest on an existing gradient-based optimization algorithm, as described by Veliov (2003). However, particular features of our model, e.g. the asset end-point constraints and the balanced annuity market, as well as variation by orders of magnitude across some of the gradients of the controls²⁶ made further non-trivial adaptations to the numerical method necessary.

2.4.4 Functional specifications

To generate a numerical solution of our model, as summarized in equations (2.12) - (2.20), we need to specify the following functional forms.

Utility

Following Hall and Jones (2007) and many others, we employ an adjusted CRRA-utility function for the instantaneous utility from consumption:

$$u^1(c) = \frac{c^{1-\sigma}}{1-\sigma} + \bar{u}, \quad 0 < \sigma \neq 1.$$

Here, we assume \bar{u} to be a sufficiently large constant, which guarantees $u^1(c) > 0$ for all reasonable consumption levels. For the second period, we assume that the cancer stock affects the utility of

²⁵One of multiple alternative assumptions would be that the insurer has perfect information about a cancer diagnosis. In this case the annuity rate would be equal to the mortality rate for each individual. Investigating the implications of different annuity markets is an interesting task in its own right that goes beyond the scope of the present paper.

²⁶The gradients of stage 1 are initially weighted way higher than that of stage 2, what results in the effects of the gradient in the second stage being levered out. Without adjustments this would imply a faster convergences of the controls in the first stage and could lead to premature termination of the algorithm as either (i) the gradients of the second stage are directly below a reasonable numerical threshold or (ii) the improvements in the objective value for adjustment of second-stage-controls are swallowed by numerical inaccuracies due to the small directional gradient step combined with their resp. small weights in the objective function.

consumption in a multiplicative form:

$$u^2(c, E) = u^1(c)v(E) = u^1(c) \exp \{ \kappa_0 \cdot E^{\kappa_1} \},$$

with $v(E) \in [0, 1]$ and $v'(E) < 0$ for $\kappa_0 < 0$. The mixed derivative $u_{cE}^2 = u_c^1 v'(E)$ is negative, implying that a higher cancer stock reduces the marginal utility of (non health care) consumption. Note that this is in line with empirical evidence (see Finkelstein et al. (2013)).

Mortality

For this numerical presentation, we assume that the non-cancer mortality rate μ^b does not depend on the state of survival, enabling us to better disentangle cancer specific mortality in the second stage. The parameters γ_i and α_i will be calibrated to reproduce the survival profile without a cancer diagnosis (see appendix 2.7.6).

$$\begin{aligned} \mu^b(t, S_1, b_1) = \mu^b(t, b_1) = g(t)b_1^{\varepsilon(t)} & \quad g(t) = \exp \{ \gamma_0 + \gamma_1 t + \gamma_2 t^2 + \gamma_3 t^3 \} \\ \varepsilon(t) = \alpha_0 + \alpha_1 t \quad (< 0) & \end{aligned}$$

Cancer incidence

Comparing the survival data adjusted for cancer mortality and the cancer incidence rates from the SEER-database we find that the function in equation (2.40) delivers the best fit, while still using just three parameters.

$$\eta(t, S_1) = \eta(S_1) = \frac{\beta_0}{1 + \beta_1 \left(\frac{1-S_1}{S_1} \right)^{\beta_2}} \quad (2.40)$$

Cancer stock

For the progression of the cancer stock, we decided to draw on the biological development and spread of cancerous cells. Talkington and Durrett (2015) present multiple established processes for the untreated spread of cancerous cells. We use the simplest version (which is sufficient for our purposes) and hence assume that the number of cancerous cells grows over time at the constant rate δ_0 . Following the work of Heuser et al. (1979), who estimated the average doubling time for breast cancer cells to be 327 days, we propose a rate of $\delta_0 = 0.774$ (per year). In a normative step, we set the initial number of cancerous cells when diagnosed constant to 1, i.e. $B(S, d) = \phi_0 = 1.0$.

For the effects of cancer care we lean on the development of health deficits introduced by Dalgaard and Strulik (2014). Depending on the intensity of cancer care h_2 (which exhibits diminishing returns modelled with $\delta_2 \in (0, 1)$) and the available technology (captured by δ_1), we propose that the growth can be slowed down or also turned negative to hopefully eradicate cancerous cells in the long-run.

$$f(t, s, E, h_2) = \delta_0 E - \delta_1 h_2^{\delta_2} E^{\delta_3}.$$

However, in contrast to Dalgaard and Strulik (2014) we propose that the effectiveness of cancer care also increases with the number of cancerous cells and is therefore multiplied by E^{δ_3} with $\delta_3 > 0$.

Cancer specific mortality

For the cancer specific mortality function $\mu^m(t, s, E)$, we eliminate the dependence on age t and assume that the whole mortality process is driven by the number of cancerous cells E and the age at which cancer was diagnosed. As we find in the data, mortality increases with the age at diagnosis. Hence we propose the following functional form:

$$\mu^m(t, s, E) = \mu^m(s, E) = \psi_0 \cdot E \cdot \exp \left\{ \psi_1 \cdot \left(\frac{s}{T} \right)^{\psi_2} \right\} \quad (2.41)$$

We estimate these parameters ψ_i using the cancer incidence and survival data as described in appendix 2.7.6.

Prices for health care

After the abstraction from preventive and acute care for this numerical exercise, there remain two prices for health care goods and services (each expressed in units of the consumption good):

- *Price of general health care p^b*
As we will detail in the appendix, we write mortality directly as a function of health care expenditures, $p^b b_1$ for the purpose of this calibration. We can therefore set the price equal to one $\implies p^b = 1$.
- *Price of cancer care p^2*
Due to the multiple types and combinations of cancer treatment, it turns out to be most practical to also measure cancer care directly in its monetary units. Hence we have decided to set the price for cancer care equal to one $\implies p^2 = 1$.

Table 2.2 summarizes the functional forms and parameter choices. We wish to stress, that we did not attempt to obtain a full calibration of the model, since this would have required the introduction of further state and control variables to represent the involved nature of cancer risk, prevention, development and treatment in the first place. Instead we aimed for a reduced model formulation, which nevertheless replicates key patterns of cancer progression and cancer care and thereby enables exceptional new insights into the behavioural patterns regarding health care (general and cancer-specific) and consumption.

Note that some parameter values shown in Table 2.2 are not the result of an automated calibration process, but are manually chosen to improve the calibration fit. The high complexity of the model and intricacy of the solution process in addition to relatively long computation times for one set of parameters did not allow us to conduct standard calibration methods. Hence the parameters $\delta_1, \delta_2, \delta_3$ and κ_0, κ_1 are not chosen through a process minimizing a strictly defined quality of calibration criterion (e.g. minimizing the maximum absolute difference in the cancer mortality rates), but are manually set to obtain a qualitatively appropriate fit of the several profiles presented.

$u^1(c) = \frac{c^{1-\sigma}}{1-\sigma} + \bar{u}$	σ	1.13 ⁺
	\bar{u}	11.355 ⁺
$\varepsilon(t) = \alpha_0 + \alpha_1 \cdot t$	α_0	-0.2700*
	α_1	0.0667*
$g(t) = e^{\gamma_0 + \gamma_1 t + \gamma_2 t^2 + \gamma_3 t^3}$	γ_0	-6.690*
	γ_1	-0.982*
$(\mu^b(t, b) = \mu_0(t) \cdot g(t) \cdot b^{\varepsilon(t)})$	γ_2	15.846*
	γ_3	-7.560*
Interest rate	r	0.03
Discount rate	ρ	0.03
$\eta(t, S) = \frac{\beta_0}{1 + \beta_1 (\frac{1-S_1}{S_1})^{\beta_2}}$	β_0	0.0243**
	β_1	0.0087**
	β_2	-2.020**
$B(S, d) = \phi_0$	ϕ_0	1.0
$f(t, s, E, h_2) = \delta_0 E - \delta_1 h_2^{\delta_2} E^{\delta_3}$	δ_0	0.7740 ⁺
	δ_1	2.0
	δ_2	0.03
	δ_3	1.3
$\mu^m(t, s, E) = \psi_0 \cdot E \cdot \exp \left\{ \psi_1 \cdot \left(\frac{s}{T} \right)^{\psi_2} \right\}$	ψ_0	0.1705**
	ψ_1	1.8475**
	ψ_2	1.5574**
$v(E) = \exp \{ \kappa_0 \cdot E^{\kappa_1} \}$	κ_0	ln(0.7)
	κ_1	1.2

Table 2.2: Summary of functional specifications and parameters in the model. Parameters indicated by ** are estimated before the solution process using only the empirical data. Parameters indicated with * result from the calibration process within the solution process. Parameters marked with + are chosen from the literature. Parameters without indications are either educated guesses (ρ, r, κ_i) or are manually chosen to improve the calibration ($\delta_1, \delta_2, \delta_3$).

2.5 Numerical results

Turning to the numerical results, we will first show to which degree we were able to replicate the data and thereby meet our calibration goals (Section 2.5.1). In a next step we discuss and compare the profiles of consumption before and after a cancer diagnosis and use the Euler-equations to identify how these profiles are impacted by the different aspects of the (potential) diagnosis (Section 2.5.2). Section 2.5.3 presents the profiles of general health investments and cancer care together with the resulting developments of the health state in both stages and the disease stock capturing the progression of cancer. In so doing, we also examine the numerical evaluations of the various valuations of health presented in Section 2.3.1 and their respective decomposition. Finally Section 2.5.4 presents the Euler-type equations for the different types of health care both in the most general formulation as well as their numerical evaluations for the optimal solution. These calculations allow us (as for the consumption profiles) to pin down the several impacts a potential diagnosis has on the individual's decision making. Throughout the whole section we summarise the key outcomes in corollaries 1–6.

2.5.1 Expenditure and survival patterns: Model vs. data

Figures 2.1 and 2.2 compare for a range of variables the outcomes of our model with data from the US, as discussed in the sections above. More specifically, Figure 2.1 summarizes the different types of expenditures in our model. The upper left panel shows the expected values for general (non-cancer) health expenditures as well as for the spending on cancer care. Overall, the expected expenditure profiles follow the data reasonably well and the corresponding spending shares, shown in the upper right panel, paint a similar picture. As was discussed before, for the purpose of better isolating transmission channels, we have quite deliberately opted for a model with an annuity market. For this reason, our model is not suited to replicate the US consumption profile from the data, but the lower left panel shows that the expected consumption profile from the model meets our goal of matching average consumption. Finally, the lower right panel illustrates the composition of the age-profiles of income (labour and transfers).

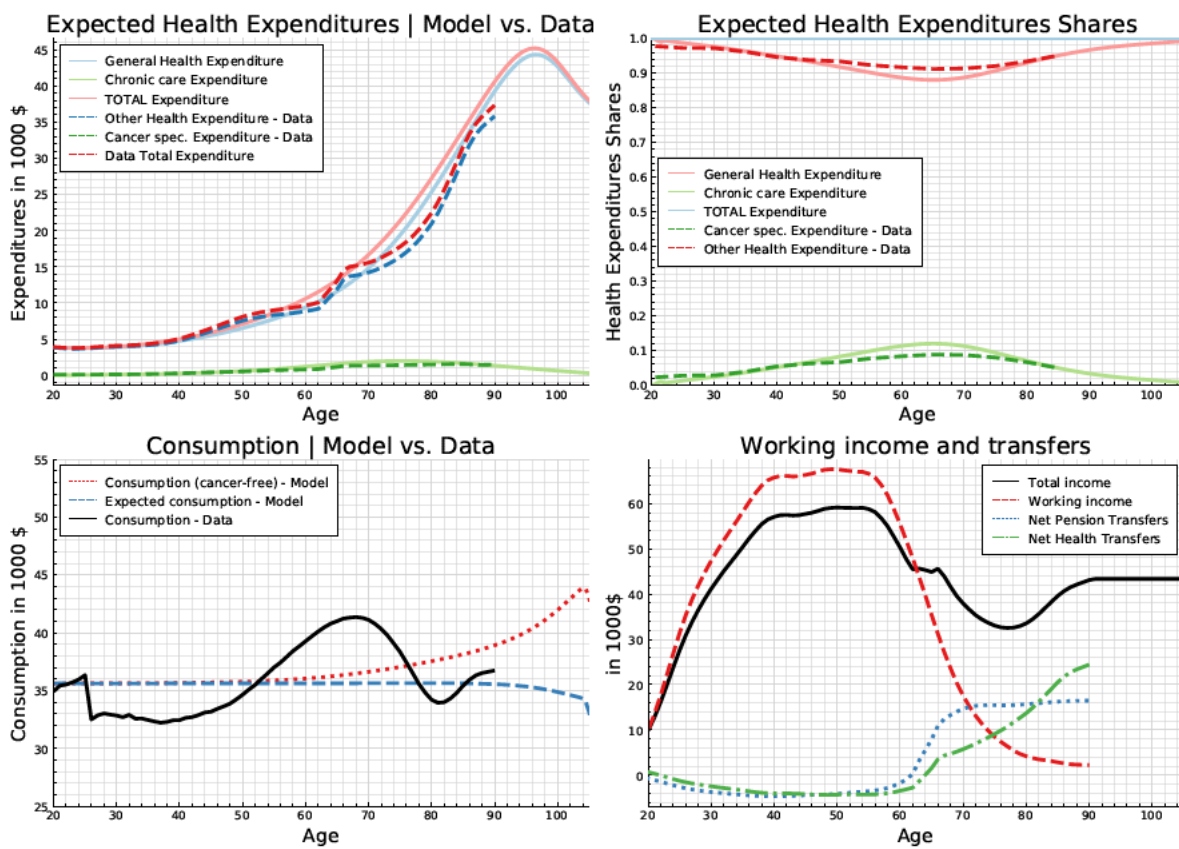


Figure 2.1: Calibration results: Comparison between expenditures predicted by the model and expenditure data

In Figure 2.2 we focus on the fit of the survival and mortality data. The upper left panel shows that we meet remarkably well both average survival, $\bar{S}(t) = Z_1(t)S_1(t) + \int_0^t Z_2(s,t)S_2(s,t)ds$ and survival conditional on remaining cancer-free, $S_1(t)$. Furthermore we can see that the cancer-free survival, $Z_1(t)$, profile of our model meets the corresponding profile derived from the data relatively well, even if slightly underestimating it.²⁷ The lower left panel shows cancer-specific mortality data over the first 10 years after a diagnosis for 4 different age-groups. Note that we plot the logarithm of the mortality rate to account for the strong differences in magnitude of these values. The model and data profiles match very well despite some discrepancy for the youngest and oldest age groups around the point of diagnosis. Finally, the lower right panel shows the total mortality rates before and after a cancer diagnosis.²⁸ The

²⁷This could be due to numerical inaccuracies, as data on cancer incidence was only available in 5-year intervals.

²⁸We again took a logarithmic scale as mortality rates towards the end of life are in a different order of magnitude compared to mortality rates in younger ages.

good fit of mortality before a cancer diagnosis is reflecting the fit of the survival profiles.

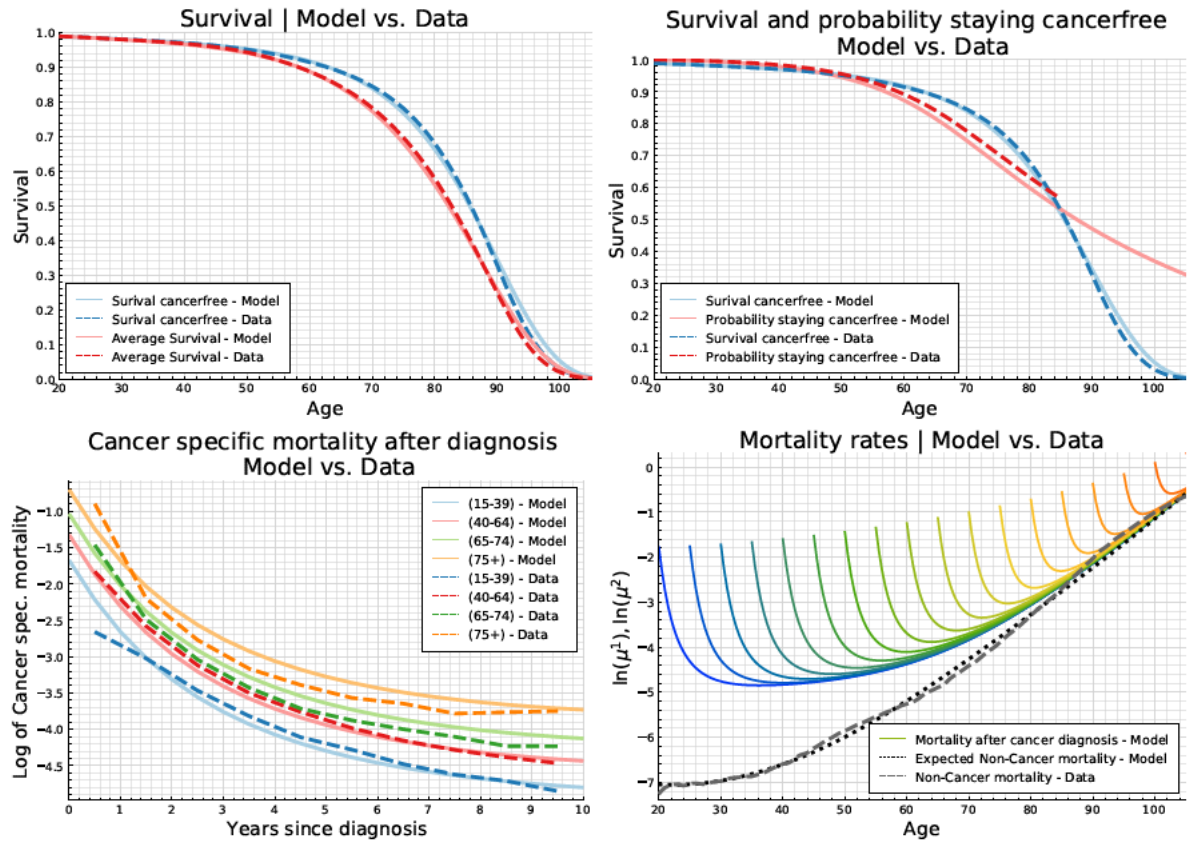


Figure 2.2: Calibration results: Comparison of survival and mortality profiles between the model and the data

2.5.2 Consumption profiles

Consider now, the consumption profiles, as plotted in Figure 2.3. The grey dashed line depicts the consumption profile in the absence of cancer. The solid lines represent the consumption profile following a cancer diagnosis at a given age, where for illustrative purposes we move the age at diagnosis in five year steps. Finally, the black dotted line illustrates the consumption level chosen immediately after a diagnosis, the distance between the dotted and dashed line representing the instantaneous consumption change following a cancer diagnosis.

The consumption profile in the absence of cancer is relatively flat and only increases towards the end of life from 35000\$ per year up to around 45000\$ in the very late stages of life. In contrast, the optimal choice of consumption following a diagnosis reveals strong fluctuations. Right after the diagnosis individuals decrease their consumption if the diagnosis occurs relatively early in life. This immediate drop is followed by an increase over the next two years which reduces the gap to consumption levels of cancer-free individuals of the same age to insignificant levels. This increase is followed by a slightly U-shaped pattern over the remaining life-course (partially increasing the consumption difference again). For early diagnosed individuals (approximately before age 70 in our setting) the levels of consumption vary between about 28000\$ and 35000\$ per year. If the diagnosis occurs after the age of 75, individuals boost their consumption directly after the cancer diagnosis and the consumption path follows a constant decline over the remaining life-course. The upward jump is increasing with the age at diagnosis and reaches up to levels of 55000\$ per year. Towards the end of life consumption decreases to 26000\$ per year.

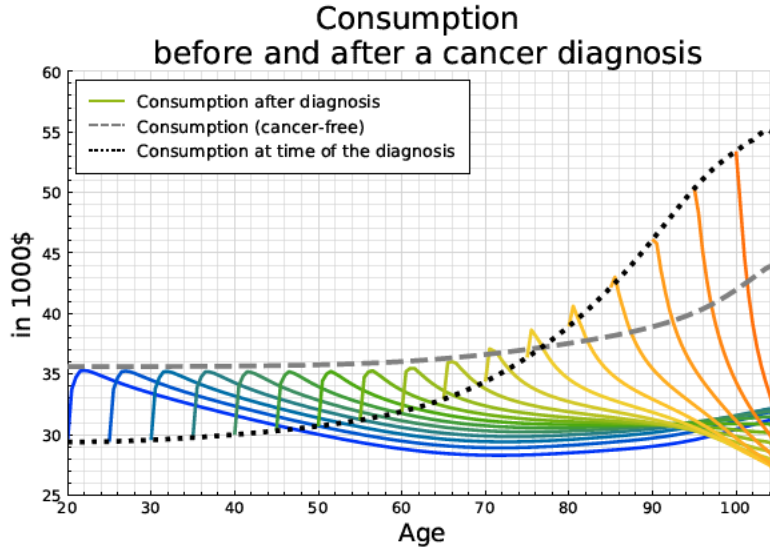


Figure 2.3: Consumption profiles in the first and in the second stage

To understand the widely different consumption patterns of cancer-free, early, and late diagnosed individuals, we derive the Euler-type equations (2.42) and (2.43) for the dynamics of consumption.²⁹ For the derivations we use the first-order optimality conditions for consumption together with the dynamics of the adjoint variables of assets (see appendix 2.7.5).

$$\frac{\dot{c}_1}{c_1} = \frac{u_{c_1}^1}{-u_{c_1 c_1}^1 c_1} \left[r - \rho + \bar{\mu} - \mu^1 - \eta + \eta P(S_1, d) \frac{u_{c_2}^2}{u_{c_1}^1} \right], \quad (2.42)$$

$$\frac{\dot{c}_2}{c_2} = \frac{u_{c_2}^2}{-u_{c_2 c_2}^2 c_2} \left[r - \rho + \bar{\mu} - \mu^2 + \frac{u_{c_2 E}^2}{u_{c_2}^2} f \right] \quad (2.43)$$

For both stages we are able to derive the elements of a standard Euler equation in the presence of partially insured mortality risk. These involve the difference between the interest rate, r , and discount rate, ρ , which drops out in our calibration. Additionally, the mortality risk, μ^i , shifts consumption toward younger ages in both stages to the extent it is not offset by annuity returns, $\bar{\mu}$.³⁰ In the intuitive case of higher mortality following a cancer diagnosis (see lower right panel of Figure 2.2) and under our assumption of a non-state contingent annuity premium $\bar{\mu}(t)$, as defined in equation (2.39), it likely holds that:

$$\mu^1(t, S_1, b_1) \leq \bar{\mu}(t) \leq \mu^2(t, s, S_2(t, s), b_2(t, s), E(t, s))$$

Hence the annuity rate overcompensates the mortality rate before a diagnosis leading to a deferral of consumption, while it only partially compensates the force of mortality after a diagnosis leading to an advancement of consumption immediately after a shock. However, as Figure 2.5 will show for the case of cancer, the second consideration may not hold over the whole life-cycle. When the cancer stock is reduced dramatically after the first few years the total mortality rate is returning back to the cancer-free mortality levels, while the average mortality $\bar{\mu}$ is still high, due to the higher total mortality of more recently diagnosed cancer patients. Hence, survivors of cancer have an incentive yet again to defer

²⁹Note that all Euler equations (here and in the following sections) are presented in their general form, and the interpretations also apply to a general health shock. Here, we discuss them in the context of the numerical results based on our calibration for cancer and discuss which terms are non-existent in our specified set-up.

³⁰As already indicated in Section 2.4 in absence of the risk of a cancer diagnosis this would imply completely flat (and identical) consumption profiles before and after the shock. Hence we can attribute all variation in consumption over the life-course to different aspects of the potential health shock/cancer diagnosis.

consumption in their late years.

The remaining factors in equations (2.42) and (2.43) differ and we analyse them separately along some numerical illustrations in Figures 2.4 and 2.5. These figures show the different terms $F_i(t)$ in the Euler-equations aggregated from the beginning of the life-cycle (resp. from the time of diagnosis) up to each point in time t . Considering, for instance, stage-1 consumption $c_1(t)$ and defining $\frac{c_1(t)}{c_1(t_0)} =: \sum F_i(t)$ we can write $c_1(t) = c_1(t_0) \exp\{\int_{t_0}^t \sum F_i(s) ds\}$ and, thus, $\log(c_1(t)) - \log(c_1(t_0)) = \sum \int_{t_0}^t F_i(s) ds$. Hence, the aggregated terms are direct expressions of the (log) difference between current consumption $c_1(t)$ and initial consumption $c_1(t_0)$. We apply the same reasoning to all other control variables in the subsequent sections.

First stage

In the first stage the uninsured risk of suffering a health shock, η , shifts consumption towards younger ages, but in contrast to the mortality risk, there is an additional effect associated with a health shock. The term $\eta P \frac{u_{c_2}^2}{u_{c_1}^1}$ shows that the desire for consumption smoothing tends to shift consumption back to later ages. The term contains the product of the hazard rate, η , the probability of surviving the shock, P , and the ratio of the marginal utilities relating to the consumption levels $c_2(t, t)$ and $c_1(t)$ immediately after and before the shock. The rationale behind this shift lies in the incentive to accumulate precautionary savings which can be used to maintain a given level of consumption following the shock.

Three thought experiments can illustrate how the two offsetting effects combine. If the health shock has no effect on the household at all, $P = 1$ and $c_2(t, t) = c_1(t)$ hold. As a result, $\eta = \eta P(S_1, d) \frac{u_{c_2}^2}{u_{c_1}^1}$ and, as expected, the two terms cancel out in the c_1 -dynamics. For a second experiment, we propose a health shock with no chance of survival ($P = 0$). Here, the hazard rate η has the same impact as the mortality rate μ^1 as there is no more desire for consumption smoothing. Finally, suppose that a health shock lowers the marginal utility of consumption such that $u_{c_2}^2 < u_{c_1}^1$ for $c_2(t, t) = c_1(t)$. For any $P \leq 1$ we then have that $\eta > \eta P(S_1, d) \frac{u_{c_2}^2}{u_{c_1}^1}$, implying that the incentive to advance consumption dominates the precautionary savings motive. However, our numerical analysis for the case of cancer shows that in general a drop in consumption after the diagnosis is possible (e.g. for early diagnosed individuals). In combination with the negative impact of the cancer stock on the marginal utility, this implies ambiguity for $u_{c_2}^2 \gtrless u_{c_1}^1$. Hence we are unable to make general statements about the combined effect and the desire for precautionary saving could potentially be more pronounced than the incentive to advance consumption.

In Figure 2.4, we illustrate the different parts of equation (2.42) aggregated up to each time-point t . We can see that the desire to advance consumption resulting from the term $-\eta$ (dashed purple line) dominates the incentive for precautionary savings, $\eta P(S_1, d) \frac{u_{c_2}^2}{u_{c_1}^1}$ (dash-dotted green line). However the difference between the two only becomes significant after the ages around 50, but strongly increases towards the end of life. As $P = 1$, the combined effect can be simplified to $\eta \cdot (u_{c_2}^2/u_{c_1}^1 - 1)$. Hence the increasing cancer incidence rate η predominantly defines the magnitude of the combined effect, while its sign is determined by the divergence of the marginal utility ratio $\frac{u_{c_2}^2}{u_{c_1}^1}$ from 1. Thus the combined effect being non-positive for all ages indicates, that the impact of the cancer stock on the marginal utility still implies $u_{c_2}^2 < u_{c_1}^1$, although $c_2(t, t) < c_1(t)$ for diagnoses in younger ages. Consequently, desire for precautionary savings is curbed compared to the health shock risk and we would obtain a tendency for consumption to decline over the cancer-free life-cycle (from this combined effect alone). However, as Figure 2.4 shows, the growing saving incentive that arises from the increasing gap between the annuity return and mortality risk for individuals who remain cancer-free (dotted blue line) overcompensates this effect and leads to an overall increasing consumption profile (black solid line, which we also have already seen in Figure 2.3).³¹ We summarise our findings in Corollary 1.

³¹This analysis for the consumption profile already highlights how the introduction of a stochastic health shocks affects the individual on multiple levels and how intricate the effort to disentangle the different driving forces can be.

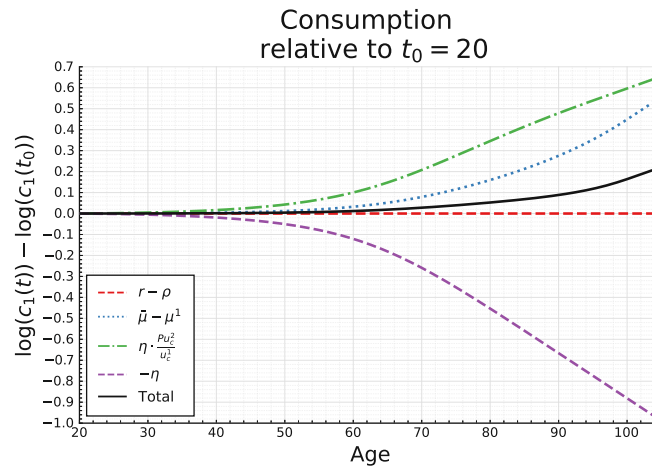


Figure 2.4: Illustrations of the impacts of the different parts of the Euler equation for consumption before the diagnosis

Corollary 1 (Consumption before diagnosis)

Compared to an individual not facing the risk of a cancer diagnosis, its introduction implies an increasing consumption profile for cancer free individuals. This results from the incentive to defer consumption (as the annuity rate overcompensates their mortality risk) dominating the desire to advance consumption (as the marginal utility of consumption is higher when cancer-free).

Second stage

In the second stage there is no risk of a further shock, but the individual takes into account the development of the shock-specific deficit stock E , i.e. the cancer stock. This leads to the term $\frac{u_{c_2}^2 E}{u_{c_2}^2} f$ in the growth rate of consumption in equation (2.43). In the more intuitive case³² of $u_{c_2 E}^2 < 0$, where the marginal utility of consumption declines with the deficit stock, two possible scenarios can arise: (i) consumption is shifted towards earlier ages where the marginal utility is still high if the deficit stock increases over time $f > 0$, or (ii) consumption is deferred if the deficit stock decreases after the shock $f < 0$.

For illustration, we consider in Figure 2.5 the consumption profiles following a diagnosis at ages 30 and 70, respectively. As we have already seen (lower right panel in Figure 2.2), the mortality risk after the diagnosis is significantly increased for several years and consequently is only partially offset by the annuity rate. This implies advanced consumption right after the diagnosis as depicted by the dotted blue line in both panels in Figure 2.5. However, individuals diagnosed early in life can have an incentive to defer consumption if they survive long enough for their mortality risk to reapproach that of cancer-free individuals, such that the term $\bar{\mu} - \mu^2$ becomes positive. Altogether this results in a U-shaped consumption pattern illustrated by the dotted blue line in the left panel.

As we will present later (see Figure 2.7), the cancer stock decreases in a convex pattern regardless of the age at diagnosis, hence $f < 0$ during the first years after the diagnosis. Therefore, $\frac{u_{c_2}^2 E}{u_{c_2}^2} f > 0$ captures the incentive of individuals to postpone their consumption in line with their recovery from cancer (dash-dotted green line). In our calibration this would imply a steep increase in consumption in the first few years after the diagnosis, followed by a fairly constant profile as the deficit stock is significantly decreased and has limited impact (similar for early and late diagnoses).

Adding up both terms to obtain the total effect, we are able to explain why the consumption profiles of early and late diagnosed individuals feature such pronounced qualitative differences. In the left panel

³²The marginal utility of consumption is higher for lower levels of the deficit stock. This holds for the multiplicative specification of the stage-2 utility function in our calibration and is empirically supported by Finkelstein et al. (2013).

we see, that for $s = 30$ the desire to defer consumption dominates the incentive to advance consumption in the first two years resulting in an initial increase in consumption in total (solid black line). As the impact of the cancer stock vanishes, the difference between annuity returns and mortality risk adds the U-shape for the consumption profile over the remaining life-course. For $s = 70$, the above average mortality risk after a cancer diagnosis becomes the dominant driver after the first 6 months. Following an initial boost, consumption then declines until the end of life (solid black line), as post-cancer mortality never decreases to levels, where it falls short of the annuity rate.

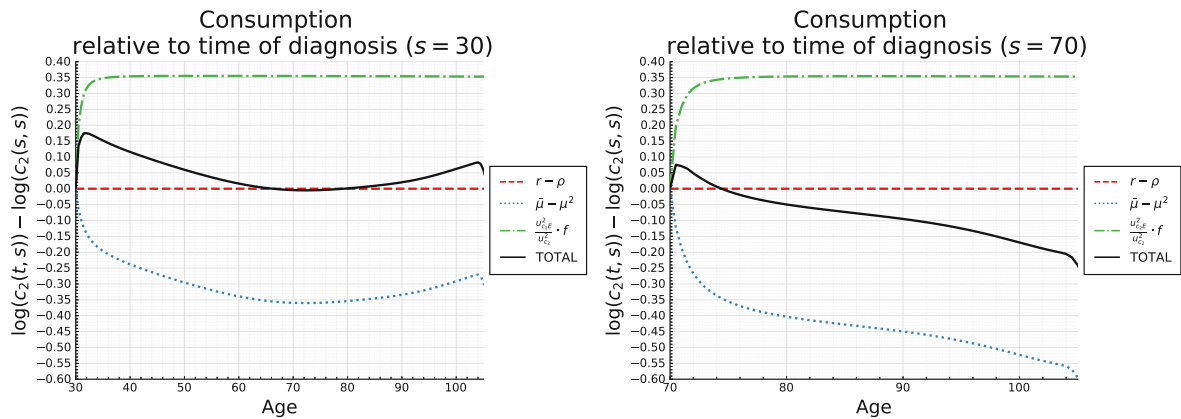


Figure 2.5: Illustrations of the impacts of the different parts of the Euler equations for consumption after the diagnosis.

Corollary 2 (Consumption after diagnosis)

The increased mortality risk following a cancer diagnosis implies, that individuals strongly advance their consumption. This effect can be dampened or even overcompensated by the incentive to postpone consumption to ages where a reduction in the intensity of cancer allows for a higher marginal utility of consumption.

2.5.3 Health expenditure and state profiles

In this section we focus on the health expenditure choices and the resulting profiles of the health/survival stock and cancer deficits. We also calculate, for our numerical specification, the valuations of health from Definition 1 and decompose them according to Proposition 1, noting that they will play a crucial role in the analysis of the health care Euler-type equations in Section 2.5.4.

Health expenditure profiles

For the life-cycle allocation of general health expenditures, as shown in the upper left panel in Figure 2.6, a cancer diagnosis does not imply a great difference regarding the qualitative shape over the life-cycle. Quantitatively, however, we obtain some significant differences. For early diagnoses up to age 70 health expenditures initially drop by around 2000\$ per year, which amounts to around 50% of the expenditures before a diagnosis for the youngest groups. This gap persists for the remaining lifetime following a cancer diagnosis and even widens after the age of 70. On the other hand a diagnosis already late in life leads to an even more sizeable immediate reduction in general health care spending (in absolute values).

The instantaneous reduction in general health care spending (difference between the dashed and dotted lines) can be traced back to the sudden and large increase in spending on cancer care that is triggered by a diagnosis (see the upper right panel of Figure 2.6). For all age groups, the initial expenditures are highest immediately following the diagnosis and then decline rapidly within around ten years after the diagnosis where the residual level can be interpreted as regular screenings to avoid the reemergence of cancer. Strikingly, (initial) expenditures on cancer-care (dotted line) slightly

increase with age from around 67000\$ at young ages to 72000\$ if the diagnosis occurs at age 60. As the time horizon quickly shortens when cancer is diagnosed closer to the end of life, the benefits of cancer care are less pronounced, and as a result the initial spending level declines steeply down to insignificant values. Expenditures for cancer care might appear rather high, but are in line with the average per-patient costs, that are reimbursable for breast cancer. Blumen et al. (2016) estimate the allowed costs to lie between roughly 60000\$ and 134000\$ within the first 12 months after a diagnosis (depending on the stage of cancer) and between 13000\$ and 70000\$ through the second year.

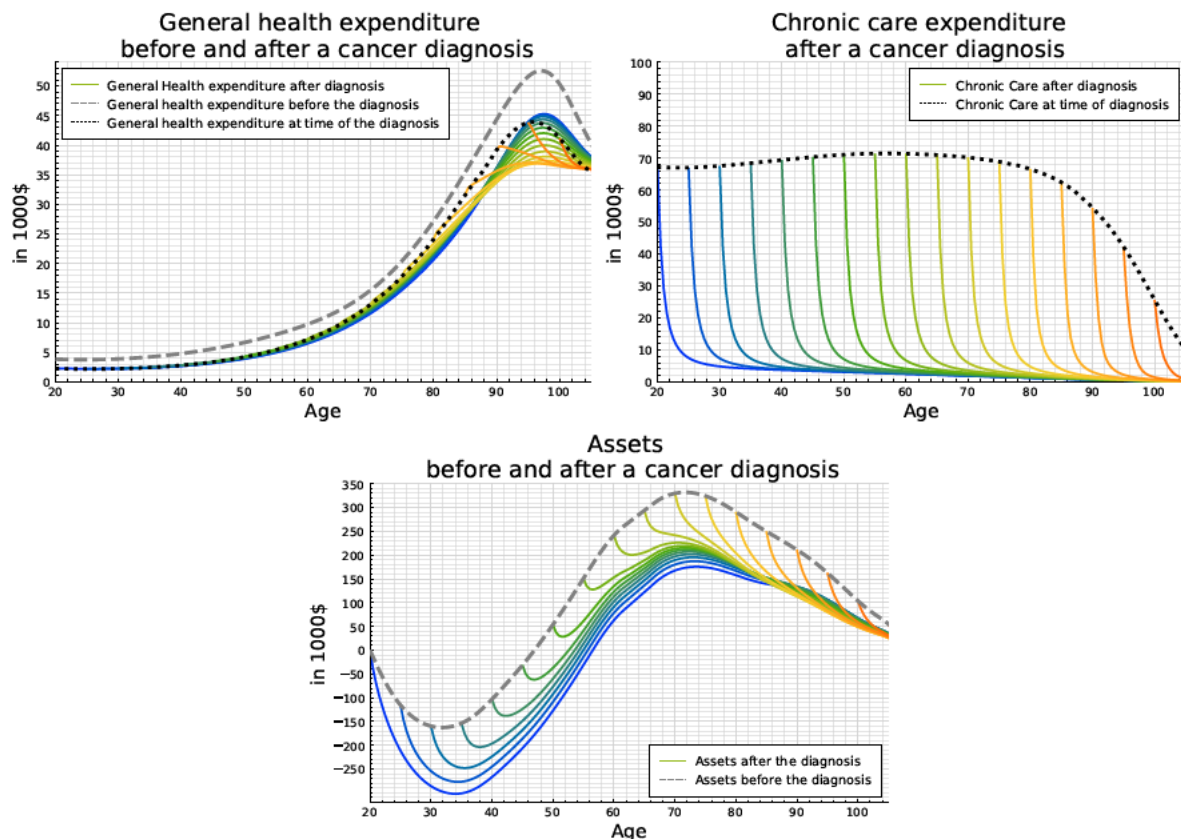


Figure 2.6: First row: General and cancer specific health expenditures before and after the diagnosis. Second row: Financial assets before and after the diagnosis.

The lower panel of Figure 2.6 shows the financial burden associated with the high expenditures for cancer care. Right after the diagnosis, assets decline steeply, where young individuals (holding negative assets already (dashed line)) go further into debt in order to finance cancer treatment, despite concomitantly reducing general health expenditure and consumption, as in Figure 2.3. We need to stress, however, that the structure of the annuity market plays a crucial role. As we assume that cancer patients are not identified by the market, they obtain the same annuity rate as a representative member of the population. Despite their higher mortality risk individuals with a cancer diagnosis can thus still go into debt to finance cancer care, facing the same full interest rate as the average population. In contrast, adjustments in the annuity price to the individual health state and associated mortality would strongly compromise the individual's scope for undertaking the investments that are necessary for treating cancer effectively.

Health state profiles

In Figure 2.7 we see the outcomes related to general health care and chronic care spending, as reflected in the profiles of the health/survival state and the cancer stock. In the left panel we see the survival/health state profiles in the absence of a cancer diagnosis (dashed grey line) and those following a cancer diagnosis

(solid grey-scaled/coloured lines) at various ages (again in 5-year steps). Most prominently we see a rapid decline in survival immediately after a diagnosis, reflecting high mortality risk during the first years after a diagnosis. While the decline in survival becomes less distinct after ten years (for most age groups), even for diagnoses early in life, the survival path never returns to the one of cancer-free individuals.

In the right panel we see that the (optimal) pattern of cancer care is effective in diminishing the cancer stock for all ages at diagnosis. This suggests that for our specification, cancer kills patients relatively early on and is otherwise transformed into a chronic disease, where a small residual stock is not eliminated but rather held in check by chronic care. This can be interpreted as a situation in which cancer can, indeed, be reduced to negligible levels (and, thus, be considered to be eliminated for all practical purposes) but where the risk of its resurgence requires a small amount of care in the form of regular screenings. The residual cancer stock also induces a mortality risk (depending on the age at diagnosis), which covers the chance of dying through a potential relapse. This also explains intuitively, why the survival profiles never return to the qualitative shape of cancer-free individuals. Meanwhile the impact of the residual cancer stock on consumption utility is insignificant in the long run. For our calibration, after ten years the remaining cancer stock decreases utility by less than one percent.

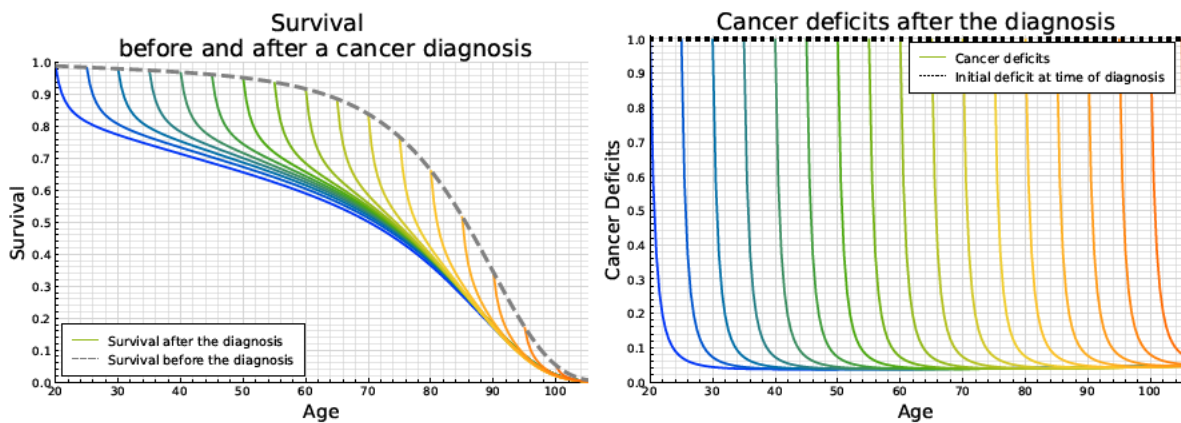


Figure 2.7: Health/survival state and cancer deficits before and after the diagnosis

Valuations of Health

Before we explore the Euler-type equations for general health and chronic care expenditures, we will make a small detour providing numerical illustrations of the valuations of the various components of health derived in Section 2.3.1. These profiles play a crucial role for the shape of health investment profiles as we already indicated in Section 2.3.2 discussing the first order optimality conditions.

In the upper left panel of Figure 2.8 we compare the different valuations. The solid black line shows the value of health (VOH) in the first, cancer-free stage ψ_H^1 , which starts slightly above 16 million \$ at age 20 and follows a concave-convex decline over the life-cycle. The ex-ante or average VOH, which takes account of the prevalence of cancer within each particular age-class (see definition 2), is illustrated by the dotted (blue/dark) line and differs only slightly from the first-stage VOH. The dash-dotted (green) line indicates the value of acute survival, which covers the willingness to pay for an increase in the continuation probability. Considering this value might look counter-intuitive at first, as in the case of a progressive disease such as cancer we do not have a mortality risk at the point of diagnosis. Following equation (2.26), however, the value of acute survival also covers the initial VOH after the shock if $P \equiv 1$. Hence, we can see that the VOH drops dramatically after a diagnosis before age 70, this difference being much less pronounced for diagnoses later in life.

The upper right panel shows the development of the VOH in the second stage, i.e. the stage with cancer, for different ages at diagnosis as compared to the VOH in stage one. Apart from an age gradient there is only little variation in the second-stage VOH with respect to the age at diagnosis. Given the

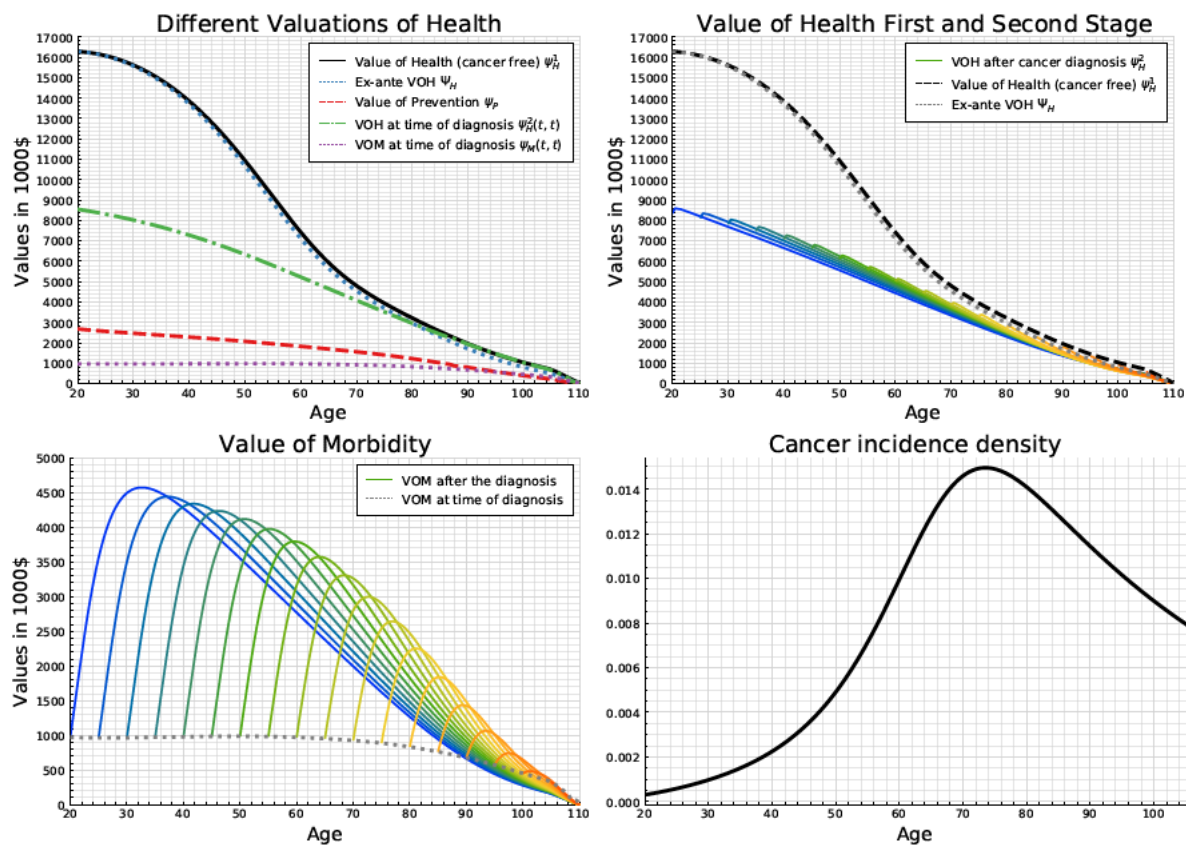


Figure 2.8: Valuations of Health

strong decline in the VOH at diagnosis for all ages up to 70, questions about the comparatively small difference between the ex-ante VOH and ψ_H^1 might arise. The answer originates in the low probabilities of getting diagnosed with cancer early in life, as can be seen in the lower right panel in Figure 2.8. This implies a low weight of ψ_H^2 in the calculation of the ex-ante VOH (see equation (2.33)). In contrast, the difference in the first and second stage VOH has become small for the age-groups older than 70, which exhibit significant prevalence of cancer. Thus the difference between the 'cancer-free' VOH and the ex-ante VOH remains remarkably small over the whole life-cycle. This has an interesting policy implication:

Corollary 3 (Expected Value of Health)

Even health shocks that are wide-spread (such as cancer) and have large impacts on the VOH from an individual perspective, exert little influence on the VOH from a population perspective (as represented by the ex-ante or average VOH) if their prevalence distribution is centred on high ages.

Returning to the upper left panel, the dashed (red) line indicates the value of prevention, i.e. the willingness to pay for a reduction of the hazard rate. This value is significantly smaller, but still starts above 2.5 million \$ at age 20 and is still around one million \$ at the age of 85. Finally the dotted (purple/light) line represents the value of reducing morbidity (VOM) right at the time of the diagnosis. This value stays fairly constant around one million \$ up to age 70 and then declines slowly towards the end of the time-horizon. The lower left panel in Figure 2.8 shows the VOM after a cancer diagnosis for varying ages at diagnosis. Strikingly, the value of morbidity increases over the first 10-12 years to values more than fourfold the VOM at the time of diagnosis, followed by a continuous decline over the remaining life-course. These patterns can be understood when recalling from the FOC (2.37) that in the optimum the VOM equals the effective price of reducing the cancer stock by one unit, $\psi_M = \frac{p^2}{-f_{h_2}}$. As cancer care becomes less effective with a diminishing cancer stock, this implies an increasing effective

price of reducing cancer (further). For an optimal allocation, this, in turn, implies that the VOM (i.e. the marginal rate of substitution between reductions in the cancer stock and consumption) must increase as well. Once the cancer stock has been reduced to its residual value after around ten years, the effective price of controlling it remains constant, and the VOM (and efforts towards controlling cancer) now declines in line with the reduction of the remaining life expectancy.

In the next step we discuss the VOH before a diagnosis (left panel) and the value of prevention (VOP) (right panel) and their respective decompositions for an optimal allocation within our numerical setting in Figure 2.9. The solid (black) lines equal the sum of the respective sub-parts and correspond to the curves in the upper left panel in Figure 2.8. For the VOH in the left panel the conventional VOL (dashed red line) adds the main share over the whole life-cycle. This part thereby decreases nearly linearly with age and accounts for most of the VOH after the age of 70. Up to that age the benefit of health, as measured by S_1 , through better prevention of cancer (dotted blue line) contributes significantly and explains more than a third of the total cancer-free VOH at age 20. The remaining component of the VOH captures the effect of first-stage health, S_1 , on second-stage health, S_2 . At values below 1 million \$ this part is relatively small compared to the others, but still far from insignificant in absolute terms.

Corollary 4 (Value of Health before diagnosis)

A potential cancer diagnosis significantly adds to the value of health for relatively young ages if health has a preventive effect on the hazard rate.

In the right panel, we decompose the value of prevention into its three components. The direct (dashed red line) and indirect (dotted blue line) benefits of postponing a cancer diagnosis get partially offset by the utility which would have been generated after a diagnosis (dash-dotted green line). As the absolute value of the losses are less than even the direct gains alone, we obtain an overall positive value of prevention over the life-cycle.

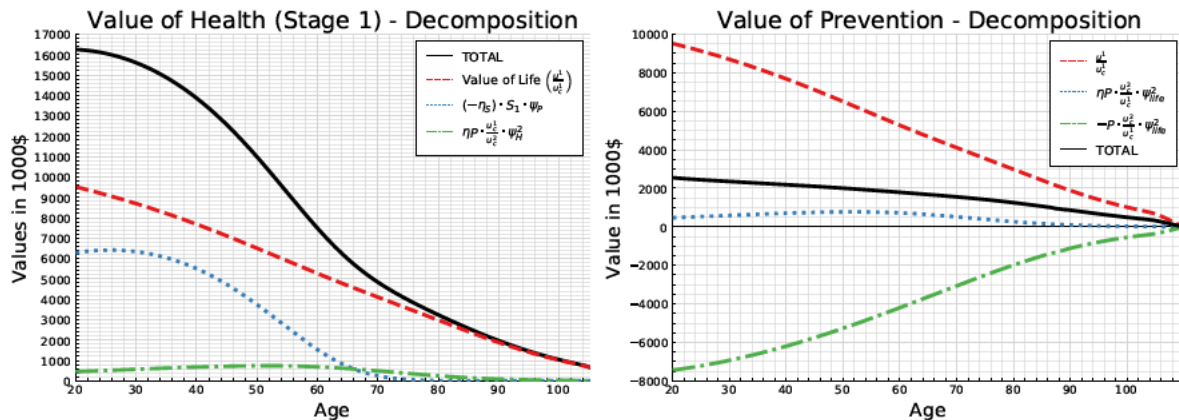


Figure 2.9: Valuations of Health

2.5.4 Euler equations for health care investments

Using the first-order optimality conditions for the different types of health expenditures presented in Proposition 2, which hold for each separate point in time, we can derive Euler type equations for general health care (before and after the diagnosis), chronic care and also preventive care³³. The full derivations are again relegated to appendix 2.7.5.

³³Again we want to stress that the interpretations also hold for the case of a general health shock. In this setting they are discussed together with the numerical results in the case of cancer, since this allows a more intuitive understanding and gives a feeling of the signs and sizes of the different channels. The results for preventive care however are purely theoretical as we abstracted from this type of care in our numerical setting.

The Euler equations for all types of health investments share a similar pattern as they consist of three main terms. (a) Intuitively the dynamics of a specific type of health care are connected to the dynamics of the valuation of health, which enters the respective FOC. E.g. for life-cycle pattern of general health care investments, the value of health is the decisive valuation, while for chronic care, the value of morbidity is a determining factor. (b) Changes in the efficiency of health investments over time due the impacts of other variables or external factors like age or technology can contribute significantly towards the qualitative pattern over the life-course. (c) The price development of different health care types imposes further incentives to either postpone investments to later ages or conduct them earlier in life.

General health expenditure

The Euler equations for general health expenditure before and after a cancer diagnosis in equations (2.44) and (2.45) directly illustrate these three main contributing factors. Note that all terms are finally scaled with $\frac{-\mu_{b_i}^i}{\mu_{b_i}^i b_i}$ respectively, which is the equivalent to the inverse of the intertemporal elasticity of substitution in the consumption Euler equation.

$$\begin{aligned} \frac{\dot{b}_1}{b_1} = & \frac{-\mu_{b_1}^1}{\mu_{b_1}^1 b_1} \left[-\frac{\mu_{b_1 S_1}^1}{\mu_{b_1}^1} \mu^1 S_1 + \frac{\mu_{b_1 t}^1}{\mu_{b_1}^1} - \frac{\dot{p}^b}{p^b} + \right. \\ & \left. + r + \bar{\mu} + \mu_{S_1}^1 S_1 - \frac{u^1/u_{c_1}^1}{\psi_H^1} + \eta_{S_1} S_1 \frac{\psi_P}{\psi_H^1} - \eta P \frac{u_{c_2}^2}{u_{c_1}^1} \left\{ \frac{(\psi_H^2 - \psi_H^1)}{\psi_H^1} + \frac{P_{S_1} S_1 \psi_{life}^2}{P \psi_H^1} + (-B_{S_1}) S_1 \frac{\psi_M}{\psi_H^1} \right\} \right] \\ & = \frac{\psi_H^1(t)}{\psi_H^1(t)} \end{aligned} \quad (2.44)$$

$$\begin{aligned} \frac{\dot{b}_2}{b_2} = & \frac{-\mu_{b_2}^2}{\mu_{b_2}^2 b_2} \left[-\frac{\mu_{b_2 S_2}^2}{\mu_{b_2}^2} \mu^2 S_2 + \frac{\mu_{b_2 t}^2}{\mu_{b_2}^2} - \frac{\dot{p}^b}{p^b} + r + \bar{\mu} + \mu_{S_2}^2 S_2 - \frac{u^2/u_{c_2}^2}{\psi_H^2} \right] \\ & = \frac{\psi_H^2(t,s)}{\psi_H^2(t,s)} \end{aligned} \quad (2.45)$$

We will now discuss the economic interpretations of the three driving forces (a), (b), and (c) and their respective decompositions in detail.

(a) **Value of Health:** The rate of change of the VOH, $\psi_H^1(t)/\psi_H^1(t)$ resp. $\psi_H^2(t,s)/\psi_H^2(t,s)$, plays a crucial part in the Euler equations. Using the results of Proposition 1 these rates can be decomposed into several parts. Parts (i) and (ii) discussed below show up in both stages and the interpretations are relevant for general health expenditures in both stages. Meanwhile the effects discussed in (iii)-(vi) are only present before the cancer diagnosis.

- (i) The discount rate of the VOH ($r + \bar{\mu} + \mu_{S_i}^i S_i \gtrless 0$) incentivizes later health investments for high market returns ($r + \bar{\mu}$). This effect is partially offset as the individuals take the effect of higher health on the mortality rate into account ($\mu_{S_i}^i S_i \leq 0$).
- (ii) $-\frac{u^i/u_{c_i}^i}{\psi_H^i} \leq 0$ captures the depreciation of the value of life (relative to the VOH) and motivates health investments earlier in life.
- (iii) $\eta_{S_1} S_1 \frac{\psi_P}{\psi_H^1} \leq 0$ implies advancement of health investments towards younger ages, since S_1 has an impact on the hazard rate and decreases the probability of a cancer diagnosis. The extent of this effect depends on the value of prevention relative to the value of health.
- (iv) $-\eta P \frac{u_{c_2}^2}{u_{c_1}^1} \frac{(\psi_H^2 - \psi_H^1)}{\psi_H^1} \gtrless 0$ infers that if the value of health is smaller in the second stage compared to the first stage (as we have seen in Figure 2.8), health investments are less attractive in the present (in case of a shock, health is instantaneously valued less) and consequently delayed to

later stages in life. This factor becomes more pronounced at ages at which a cancer diagnosis is more likely (increased η).

- (v) $-\eta P \frac{u_{c_2}^2}{u_{c_1}^2} \frac{P_{S_1} S_1}{P} \frac{\psi_{iife}^2}{\psi_H^2} \leq 0$ accounts for the positive effect of health S_1 on the continuation probability and results in another force shifting health investment to younger ages. *This aspect is equal to zero in our scenario, as health has no impact on the continuation probability*
- (vi) $-\eta P \frac{u_{c_2}^2}{u_{c_1}^2} (-B_{S_1}) S_1 \frac{\psi_M}{\psi_H} \leq 0$ lastly represents that as initial deficits are lower, if an individual is in good health at time of the diagnosis, individuals have another incentive to keep their health at high levels over the life-cycle. This results in health investments being advanced towards younger ages. *This aspect is equal to zero in our scenario, as health has no impact on the initial deficits.*

(b) Effectiveness: Changes in the effectiveness of health investments over the life-cycle also play a decisive role for the shape of the profile. This aspect contains two separate parts:

- (i) $-\frac{\mu_{b_i S_i}^i}{\mu_{b_i}^i} \mu^i S_i$ covers that if health investments tend to have a higher impact for already depleted health ($\mu_{b_i S_i}^i > 0$), people are more likely to defer until their health is depleted to increase their health expenditures. *This aspect is equal to zero in our scenario, as health has no impact on the base mortality rate.*
- (ii) $\frac{\mu_{b_i t}^i}{\mu_{b_i}^i}$ shows that people have an incentive to postpone their health investment to later ages, if marginal effectiveness of health expenditures increase with age ($\mu_{b_i t}^i < 0$). This could also include a set-up, in which individuals expect health technologies to improve during their lifetime.

(c) Price: Expected increases of the prices of health care over time push health investments towards earlier stages in life (and vice versa) as the term $-\frac{p^b}{p}$ indicates. *This aspect is equal to zero in our scenario, as prices are assumed to be constant.*

Similarly to the consumption profiles we want to illustrate these theoretical driving forces using our numerical calibration. As already indicated above, some of the terms are equal to zero due to the assumptions made for the scenario of a cancer diagnosis and are therefore omitted in the presentation in the figures below. Like in Figure 2.4 and 2.5, we present the terms in the Euler equations aggregated up to timepoint t as their sum corresponds to the difference of log-expenditures at point t and age $t_0 = 20$. Figure 2.10 shows the numerical evaluations of the different (non-zero) parts and their aggregated sum for general health expenditure of cancer-free individuals.

We can easily identify the two main driving forces behind the overall increasing health expenditures over the life-cycle. First, intuitively the sum of market and annuity interest alone would impose a strong delay in health expenditure (dashed red/dark grey line) especially towards the end of life, where the annuity rate strongly increases. The second strong incentive to postpone health investments is the increasing efficiency of general health care for older ages (dotted orange/light grey line), which by itself would imply nearly linearly increasing health expenditures after the age of 45. Consequently the effect of interest is more pronounced in the early ages and for ages 100+, while the impact of the increasing effectiveness is stronger between ages around 45 to 100.

However the aggregated effect of these two terms is dampened by two other factors. While the advancing effect of the preventive aspect of higher health is only present up to age 70 (dash-dotted green/grey line), the depreciation of the value of life (dotted blue/grey line) becomes increasingly strong over the life-course. This leads to the health expenditures in total becoming stagnant at the end of the time horizon (the solid black line becomes flat). Finally we see that the last term containing the impact of a change in the VOH through diagnosis (dashed purple/grey line) has no significant impact over the whole life-course.³⁴

³⁴The insignificant impact can be explained similarly to the small difference between the expected and the cancer-free VOH.

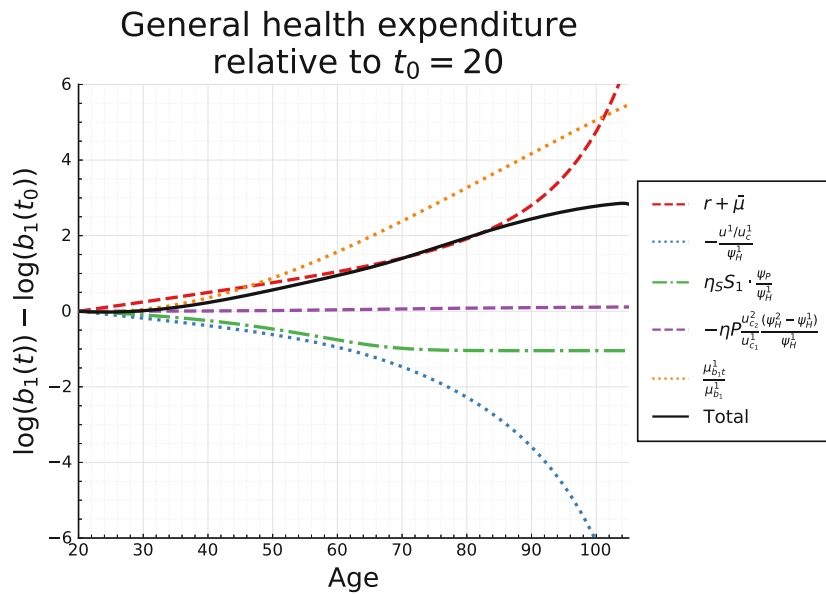


Figure 2.10: Illustrations of the impacts of the different parts of the Euler equations for general health expenditure before a cancer diagnosis

In Figure 2.11 we continue with the analogous decomposition for general health expenditure after a cancer diagnosis. In the left panel we illustrate the decomposition for an early diagnosed individual at age 30, while in the right panel the diagnosis occurs later in life at age 70. Although we already identified in Figure 2.6 that the age-profile for health expenditures after a cancer diagnosis follows a qualitatively similar pattern to the one of cancer free patients, this decomposition still helps us to identify the smaller differences. First of all, there are only three terms remaining for health expenditures after the diagnosis

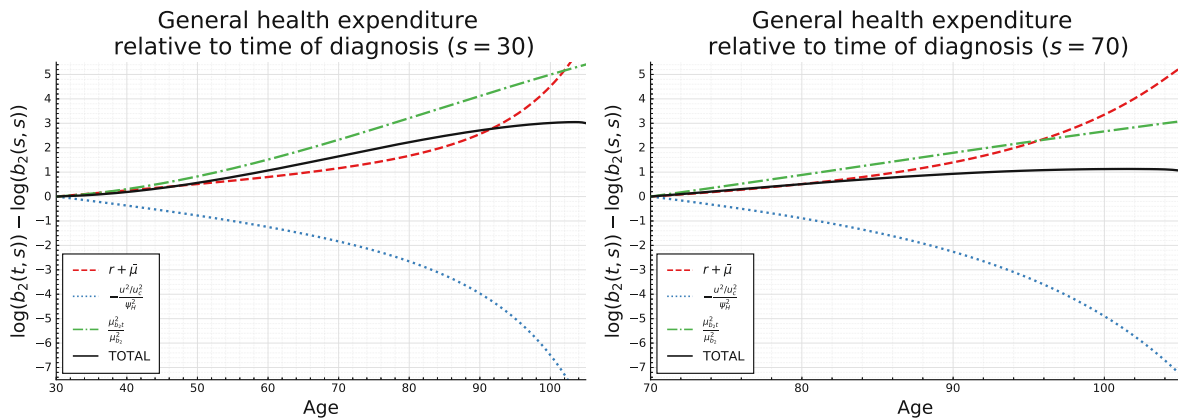


Figure 2.11: Illustrations of the impacts of the different parts of the Euler equations

compared to the previous five terms. However the three with the strongest magnitude are still present. Again the interest and annuity rate (dashed (red) line) incentivises individuals to defer health care, however now the postponing effect of increasing effectiveness of health care (dash-dotted (green) line) is the most distinct reason behind overall increasing health expenditures (solid black line). The sole dampening effect of the otherwise strongly increasing profile results from the depreciation of the value of health (dotted (blue) line). The absence of an additional advancing force (like the preventive aspect of health before the diagnosis) at least partially explains, why the total effect is actually slightly higher for ages after 60 in case of an early diagnosis compared to the cancer free profile.³⁵

³⁵Furthermore we would like to stress that these figures evidently show how an analytical derivation and assessment of

Corollary 5 (General Health Expenditure)

The main driving forces behind increasing health expenditures over the life-course are the market and annuity interest rate and the increasing effectiveness of health care in older ages. In the presence of a potential cancer diagnosis the dampening effect of the depreciation of the value of life is enhanced by an additional term, which covers the preventive effect of health and motivates advancement of general health expenditure. This term is only present before the cancer diagnosis and only shapes the behaviour of cancer-free individuals.

Cancer specific chronic care expenditure

As the last part of our numerical analysis we focus on cancer care after the diagnosis, present the corresponding Euler equation and again evaluate the decomposition along the optimal numeric solution. The dynamics of chronic care can differ quantitatively between the different shock scenarios, however the qualitative shaping forces are the same independent of the age at diagnosis. Similar to the general health expenditures all terms in equation (2.46) are scaled with $\frac{-f_{h_2}}{f_{h_2}h_2h_2}$, the equivalent to the inverse of the intertemporal elasticity of substitution in the consumption Euler equation.

$$\frac{\dot{h}_2}{h_2} = \frac{-f_{h_2}}{f_{h_2}h_2h_2} \left[\underbrace{r + \bar{\mu} - f_E + \frac{w_E^2}{\psi_M} + \frac{u_E^2/u_{c_2}^2}{\psi_M} - \mu_E^2 \frac{\psi_H^2}{\psi_M}}_{= \frac{\psi_M(t,s)}{\psi_M(t,s)}} + \frac{f_{h_2}t}{f_{h_2}} + \frac{f_{h_2}E f}{f_{h_2}} - \frac{p^2}{p^2} \right] \quad (2.46)$$

Furthermore we also obtain the same three main parts in the dynamics: (a) value of morbidity dynamics, (b) chronic care efficiency aspects and (c) the price developments.

(a) Value of Morbidity: As we have seen in Figure 2.8, the value of morbidity follows a surprising profile after the diagnosis. Consequently decomposing its rate of change $\frac{\psi_M(t,s)}{\psi_M(t,s)}$ can enhance our understanding extensively:

- (i) The discount rate of the value of morbidity ($r + \bar{\mu} - f_E$) again leads to chronic care postponement for high market rates ($r + \bar{\mu}$). If deficits accumulate slower (faster) if they are already on a high level, i.e. $f_E < 0$ ($f_E > 0$) individuals further delay (accelerate) their chronic care expenditures.
- (ii) $\frac{w_E^2}{\psi_M} \leq 0$ captures the impact of the deficit stock on the wage, what also incites individuals to conduct chronic care earlier in life or closer to the time of the diagnosis. *This term is equal to zero in our scenario, as deficits are assumed to have no impact on working income.*
- (iii) $\frac{u_E^2/u_{c_2}^2}{\psi_M} \leq 0$ is another reason for individuals to invest into chronic care earlier, as the deficits have a negative impact on the utility. The extent of the shift of chronic care towards younger ages depends on the ratio of marginal utility gains through deficit reduction and consumption.
- (iv) $-\mu_E^2 \frac{\psi_H^2}{\psi_M} \leq 0$ implies advancements of chronic care as increasing deficits impose an additional mortality risk. The strength of this factor depends on the value of health relative to the value of morbidity.

(b) Effectiveness: Changes in the effectiveness of chronic care over the life-course also play a decisive role for the shape of the patterns:

- (i) $\frac{f_{h_2}E f}{f_{h_2}} \geq 0$ covers the effect of the cancer stock on the marginal impact of chronic care. The sign of this term depends on several factors, e.g. first the sign of the mixed derivative f_{h_2E} might be ambiguous. Depending on the specific health shock, chronic care might be more effective if a certain deficit stock has been accumulated ($f_{h_2E} < 0$). On the other hand it is also possible, that an higher accumulated deficit stock makes it harder to eliminate the

dominating terms in the Euler equations would be fairly impossible. The numerical evaluation nevertheless give great insight into to driving behavioural forces.

remaining deficits as the treatment is effective ($f_{h_2E} \geq 0$). Second it is crucial whether deficits are accumulated or reduced at every point in time as it defines the sign of f .

- (ii) Again we might expect $\frac{f_{h_2t}}{f_{h_2}} \geq 0$ to lead to a delay in chronic care investments, since it is reasonable for chronic care to be more effective for older ages. Still we cannot make a statement about the sign of this term in general. *This term is equal to zero in our scenario, as the progression of cancer is assumed to be independent of age.*

- (c) **Price:** Expected increases of the price of cancer care over time can push cancer care closer to the diagnosis (and vice versa) as the term $-\frac{p^2}{p^2}$ indicates. *This term is equal to zero in our scenario, as prices are assumed to be constant.*

In Figure 2.12 we present the numerical profiles aggregated from the time of diagnosis analogue to the previous analyses. First of all we find, that the qualitative patterns of each decompositional term are similar regardless the age at diagnosis, however the absolute values and consequently the composite effect can be different.

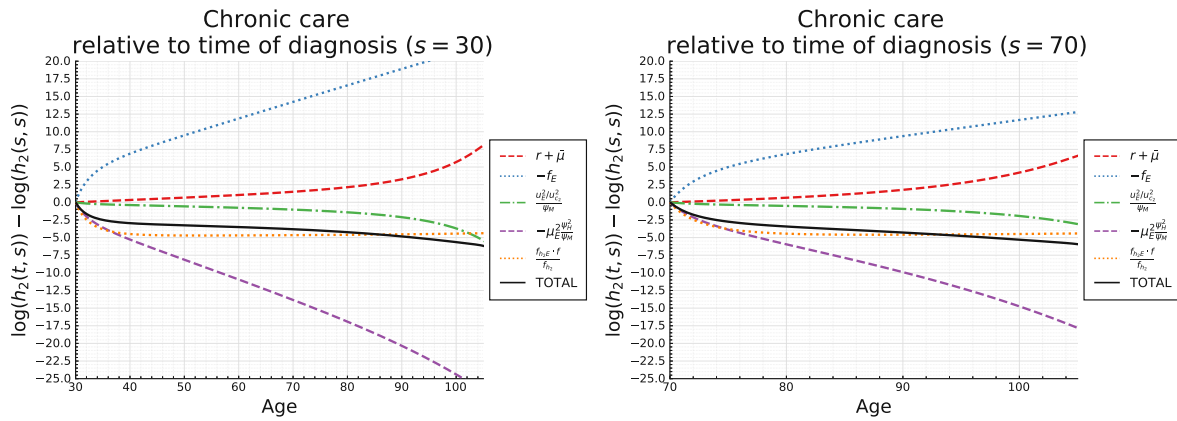


Figure 2.12: Illustrations of the impacts of the different parts of the Euler equations for cancer specific care after a diagnosis.

Similarly to all types of health expenditures presented so far, the total interest rate $r + \bar{\mu}$ alone would lead to deferral of chronic care especially late in life (dashed red/dark grey line). Strikingly, as the marginal impact of deficits on the efficiency of chronic care is (for the optimal chronic care profile) stronger than the cancer growth rate δ_0 , the term $-f_E$ implies a delay in cancer care, especially in the first years after the diagnosis (dotted blue/dark grey line). Consequently through this channel we would expect a chronic care profile even more concentrated right after the diagnosis, if the efficiency of chronic care was independent or even increasing with a decreasing cancer stock.

On the other hand the negative impacts of the cancer stock are an incentive for the advancement of cancer treatments. Notably the term concerning the additional mortality risk (dashed violet/grey line) implies a continuously decreasing chronic care pattern after the diagnosis. This effect is strengthened by the term containing the impact of cancer deficits on utility, which, however, is comparatively small and only increases in magnitude at very late ages (dash-dotted green/light grey line).³⁶ As a last part the dotted orange/light grey line shows, that the decreasing efficiency of cancer treatments for decreasing cancer stock leads to investments closer to the diagnosis, but has no significant impact on the qualitative shape of the pattern after the first ten years. Adding all terms up, we obtain the total profile (solid black line), which highlights the dominance of negative terms over the positive ones. The strongest decrease in the cancer expenditure profile occurs right after the diagnosis, which is followed by a period of relatively flat resp. slightly diminishing expenditures.

³⁶Note that this reassures, that changes in our educated guesses for the parameters of the utility function likely do not have significant impact on the overall structure of the optimal solution.

Corollary 6 (Cancer treatment)

A spike in cancer care right after the diagnosis is mainly driven by the high additional mortality risk and decreasing efficiency of cancer care for a decreasing cancer stock. However the combined impact is attenuated as the direct effect of the cancer stock on cancer progression (under the optimal cancer care regime) acts in an opposing way. Compared to the other types of health care, the total interest rate significantly affects the cancer care profile only towards older ages 70+.

Preventive care

Lastly we want to analyse the theoretical results we can derive for preventive care, which we unfortunately cannot underline with our numerical solution as for the other types of health care. However the interpretations of the analytical results still hold true in general. In equation (2.47) we present the Euler type equation for preventive care.

$$\frac{\dot{h}_1}{h_1} = \frac{-\eta_{h_1}}{\eta_{h_1 h_1} h_1} \left[\underbrace{r + \bar{\mu} + \eta P \frac{u_{c_2}^2}{u_{c_1}^1} - \frac{u^1/u_{c_1}^1}{\psi_P}}_{=\frac{\dot{\psi}_P}{\psi_P}} + \left(\rho - \frac{\dot{P}}{P} - \frac{\frac{d}{dt} V^*(t)}{V^*(t)} \right) \frac{u_{c_2}^2}{u_{c_1}^1} \frac{P \psi_{life}^2}{\psi_P} - \frac{\eta_{h_1 S_1} \mu^1 S_1}{\eta_{h_1}} + \frac{\eta_{h_1 t}}{\eta_{h_1}} - \frac{\dot{p}^1}{p^1} \right] \quad (2.47)$$

Analogous to the other health care measures the Euler equation consists of the equivalent of the inverse of the intertemporal elasticity of substitution as a scaling factor and the three main impact factors: (a) value of prevention dynamics, (b) preventive care efficiency aspects, and (c) the price developments. While the latter two are straight forward as for the other types of health care, the decomposition of the former can be significantly more involved. This follows intuitively from preventive care (resp. the value of prevention) generating costs in the cancer-free present, while its impacts and benefits are on the probabilistic side.

(a) Value of Prevention: The rate of change of the value of prevention $\frac{\dot{\psi}_P(t)}{\psi_P(t)}$ can be split up in the following parts:

- (i) The discount rate of the value of prevention ($r + \bar{\mu} + \eta P \frac{u_{c_2}^2}{u_{c_1}^1} > 0$) again leads to chronic care postponement for high market rates ($r + \bar{\mu}$). Additionally the desire for precautionary savings $\eta P \frac{u_{c_2}^2}{u_{c_1}^1}$ (which we also discussed in Section 2.5.2 regarding the consumption profiles) further delays preventive care.
- (ii) $-\frac{u^1/u_{c_1}^1}{\psi_P} < 0$ captures the the value of life lost, when being diagnosed with cancer, relative to the value of prevention and motivates advancements of preventive care to earlier ages.
- (iii) $\left(\rho - \frac{\dot{P}}{P} - \frac{\frac{d}{dt} V^*(t)}{V^*(t)} \right) \frac{u_{c_2}^2}{u_{c_1}^1} \frac{P \psi_{life}^2}{\psi_P} \geq 0$ is the most involved part. It contains the comparison between the value of life in the second stage and the value of prevention. The sign of this term depends on whether the utility discount rate (ρ) is higher then the sum of the rate of change in the continuation probability and the optimal objective value in the second stage $\left(\frac{\dot{P}}{P} + \frac{\frac{d}{dt} V^*(t)}{V^*(t)} \right)$. However it is more than plausible that $\frac{\frac{d}{dt} V^*(t)}{V^*(t)} < 0$ in general, since the length of the remaining time horizon after a health shock shrinks as t increases. Furthermore as $\frac{\dot{P}}{P} \leq 0$ is more likely than not for most diseases, the total effect is probably positive in many cases (without general applicability) and implies deferral of preventive investments towards older ages, where the cancer diagnosis (or general health shock) has a more significant impact on the individual.

(b) Effectiveness: Changes in the effectiveness of preventive care over the life-cycle also play a decisive role for the shape of the profile:

- (i) $-\frac{\eta_{h_1 S_1} \mu^1 S_1}{\eta_{h_1}}$ indicates, that depending on whether preventive care is more ($\eta_{h_1 S_1} < 0$) or less ($\eta_{h_1 S_1} > 0$) effective while in better health, individuals are either motivated to conduct preventive care earlier or later in life.
- (ii) $\frac{\eta_{h_1 t}}{\eta_{h_1}}$ shows that the expectation of higher effectiveness of preventive care in the future ($\eta_{h_1 t} < 0$) leads to postponement of preventive care to older ages.

(c) **Price:** Expected increases of the price of preventive care over time push investments into preventive actions towards earlier stages in life (and vice versa) as the term $-\frac{p_t^1}{p^1}$ indicates.

Corollary 7 (Preventive Care)

The Euler equation for preventive care is the most involved within all types of health care. Beside the price, effectiveness and interest rate impacts similar to other types of health care, the demand for precautionary savings potentially delays preventive efforts. While the depreciation of the value of life incentivises earlier investments in preventive care, a counter acting deferral effect towards later ages where the shock has more significant impact is likely to exist.

2.6 Conclusions

We have constructed a life-cycle model in which individuals respond to the risk of a singular, life-changing shock to their health (e.g. heart attack, stroke, cancer, diabetes, disabling accident) by investing besides their general health expenditures in a range of distinct forms of health care: preventive care to reduce, directly or indirectly, the arrival rate of the shock; acute care to lower instantaneous survival and the extent of morbidity/disability at the point of the shock; and chronic care to lower mortality and morbidity in the follow-up of the shock. We solve the complex underlying stochastic optimal control problem with a random time horizon by applying an innovative transformation into an age-structured control model. This enables us to derive (i) intuitive expressions for the first-order conditions for the choice of health care based on their respective (monetary) valuations; and (ii) Euler equations that reflect both age and duration of disease after the onset of the shock and forward-looking behaviour before the arrival of the shock. The first-order conditions first of all affirm the intuitive notion, individuals choose their optimal level of all distinct types of health care, so that the marginal costs of any additional effort match the monetary valuation of the respective aspect of health. On the other hand they also provide a basis for empirical derivations of the valuations of different facets of health following a revealed preference argument.

Calibrated to US data, our model illustrating the risk of onset of cancer provides a very good fit between (i) general and cancer-specific health care expenditure, (ii) age-specific survival and age-specific prevalence of cancer; (iii) age- and duration-specific cancer mortality rates, as well as (iv) average cancer and non-cancer specific mortality. Under the assumption that assets are exclusively held as annuities the return of which is based on the average mortality rate, relatively young individuals respond to a cancer diagnosis by adjusting their consumption in the light of an increased mortality risk, from a gradually increasing pattern to a U-shaped pattern (after a brief initial adjustment phase). To finance high temporary cancer-specific care they reallocate resources from consumption and health expenditures putting both respective profile on a lower level compared to cancer free individuals. Given the possibility to do so in a perfect annuity market, they also run up substantial debt in order to finance cancer-specific health care. With the duration of the disease, cancer-specific health care is reduced, reflecting an intuitive tendency of such care to evolve from urgent life-protecting care into less intensive chronic care, conditional on the individual's survival. A hike in consumption early-on for individuals diagnosed late in life coincides with a larger drop in general health expenditures and a smaller investment in cancer-specific care, showing that a cancer diagnosis affects individuals differently at different points within the life-cycle. However consumption of individuals diagnosed during old age subsequently steeply reduces to levels below those that would be realized for healthy individuals.

We calculate various expressions for the value of health. Here, the value of health before a diagnosis of cancer starts from a level of around 16.3 Mio \$ at age 20 and then declines steadily. Our analysis shows that up to around age 70 a cancer diagnosis lowers the value of health dramatically. While both the cancer-free value of health and the value of health at the point of diagnosis decline with age, the decline is much more pronounced for the cancer-free value of health. From around age 85 onwards, the cancer-free value of health no longer differs much from the value at the onset of diagnosis, reflecting the much diminished remaining life time in either case. Contrasting the cancer-free value of health with the ex-ante value of health, which involves a weighting with the prevalence function, we find little difference between the two. This may come as a surprise given the strong drop in the value of health for individuals diagnosed at young ages. Notably, however, for these age groups a cancer is an unlikely diagnosis, implying a low weight, whereas for higher ages, the two values of health have converged.

Considering the components of the cancer-free value of health, we find that while the largest part falls on the value of survival (tantamount to the conventional value of life), the preventive value of good health (i.e. the value of prevention weighted with the health-related reduction in the incidence of cancer) makes up for a significant proportion of the total value up to age 65. Standing at slightly above 6 Mio \$ up to the early thirties, it compares to a value of survival around 9.5 Mio \$ and makes up more than a third of the total value of health at age 20. From age 50 onwards, however, the preventive value of good health diminishes quickly and vanishes almost entirely for the highest ages.

Finally, turning to the Euler equations, we see that under our assumption that the rate of time preference equals the interest rate, two offsetting forces emerge as drivers of consumption in the healthy state: On the one hand, given that the individual remains healthy, the average mortality component (including expected mortality due to cancer) exceeds (to increasing extent) the non-cancer mortality as a source of risk, allowing the individual to increase consumption over the remaining life-course. On the other hand, a desire for consumption smoothing leads to the advancement of consumption into the healthy stage, which subsequently translates into a tendency to reduce consumption of the life-course. As it turns out, the former effect is slightly stronger for our calibration, leading to an increase in consumption with age for individuals who remain cancer-free. After the onset of cancer, again two main forces determine the consumption choice: Assuming that the marginal utility of consumption is reduced through cancer, there is a tendency for individuals to defer consumption into the future when (conditional on survival) they expect to experience a higher marginal utility from consumption. As it turns out, for the most part this tendency tends to be offset by the desire of individuals to consume instantaneously given the high mortality risk, following the onset of cancer.

The demand for health care generally develops under the presence of two forces: On the one hand, the return to annuities tends to imply an increase, whereas the writing off of life-years from the respective value of health tends to imply a decline. For general health investments, these effects tend to be moderated by age-related changes to their effectiveness. While for cancer-free individuals the shock-preventive aspect of health implies an additional incentive to advance general health investment, this factor is absent for individuals after a diagnosis. Overall, in our calibration there is a moderate increase in general health investments over the life-course both in the absence and, from a much lower level, in the presence of cancer. While the former three forces also tend to be at play for cancer-specific health care, there are two additional effects: On the one hand, individuals have an incentive to delay the consumption of cancer care, given it is more effective for a disease that has progressed already; on the other hand, the mortality risk that is increasing with the progression of the disease provides a strong incentive to advance cancer care; and this is, indeed, the dominating effect.

While in this paper we have developed a rich framework for the study of large, singular health shocks, there is considerable scope for extensions and applications. First, we are planning to provide analysis for other types of health shocks, in particular cardio-vascular events that involve an instantaneous threat to survival and, thus, warrant the consumption of acute health care. Second, especially in the case of cancer, there is an important issue about the lag between onset and diagnosis, which we plan to address in future work. Third, there is large scope for employing the model to study the role of the annuity market and the role of health and disability insurance. Finally, we are considering extensions of the

model to involve multiple shocks and differences in severity of the health shock as a second dimension of uncertainty.

2.7 Appendix

2.7.1 The model as a vintage optimal control model

The transformation from the stochastic formulation defined by equations (2.12)–(2.20) into a vintage structured optimal control problem follows the technique developed by Wrzaczek et al. (2020). We will present the two crucial steps here, for the proof and detailed derivations we will however refer to the work of the original authors.

In the first step we transform the stochastic formulation into an equivalent deterministic version. The dynamics of the state variables thereby do not change, but we need to introduce the new state variable $Z_1(t)$, which is already presented in Section 2.2 and its dynamics are described in equation (2.21). Following the description in Wrzaczek et al. (2020) we obtain the new deterministic objective function shown below.

$$\begin{aligned} & \max_{\substack{c_1(t), h_1(t), b_1(t) \geq 0 \\ c_2(t, s), h_2(t, s), b_2(t, s) \geq 0 \\ d(t) \geq 0}} \int_0^T e^{-\rho t} \left[Z_1(t) S_1(t) u^1(c_1(t)) + \right. \\ & \left. + \underbrace{Z_1(t) \eta(t, S_1(t), h_1(t)) P(S_1(t), d(t)) \cdot \left(\int_t^T e^{-\rho(\tau-t)} S_2(\tau, t) u^2(c_2(\tau, t), E(\tau, t)) d\tau \right)}_{:= \tilde{V}(t)} \right] dt \end{aligned}$$

The integrand consists of two parts. The first part is the probability of surviving in good health $Z_1(t)$ multiplied with the expected utility derived from consumption at this age $S_1(t) u^1(c_1(t))$. The second part captures the "probability" of suffering the health shock at age t ($\rightarrow Z_1(t) \eta(t, S_1(t), h_1(t))$) times the probability of surviving the shock ($\rightarrow P(S_1(t), d(t))$) times the expected aggregated utility over the remaining life-time after health shock at age t . To abbreviate some equations below we define this term as $\tilde{V}(t)$ (as it directly corresponds to the objective value of the second stage).

For the next step in the transformation process we need to introduce the variable $Q(t)$ as the expected utility at age t with the expectation taken over the distribution of all possible prior shocks $s \leq t$, i.e.

$$Q(t) = \int_0^t Z_2(s) S_2(t, s) u^2(c_2(t, s), E(t, s)) ds \quad (2.48)$$

with $Z_2(s)$ being defined as in equation (2.22). Wrzaczek et al. (2020) then show, that the equation (2.49) for the aggregated utility over all possible second stage scenarios holds,

$$\int_0^T e^{-\rho t} \tilde{V}(t) dt = \int_0^T e^{-\rho t} Q(t) dt \quad (2.49)$$

This is illustrated by a diagram similar to a Lexis-diagram (see figure 2.13). The left respectively right panel in figure 2.13 illustrate the different integrations techniques on the left respectively right side of equation (2.49). In both panels (corresponding to the resp. side in equation (2.49)) the utility is aggregated over the triangle below the 45°-line. In the left panel the integration first takes place along each shock scenario from the age at the health shock until the end of the maximum life-span. In a second step all scenarios weighted with their occurrence probability of $Z_1(t) \eta(t, S_1(t), h_1(t))$ get aggregated. On the right side, for each age t , first the utility generated in each scenario, where the shock has occurred at an age $s < t$, is aggregated and in the second step the integration over all ages takes place.

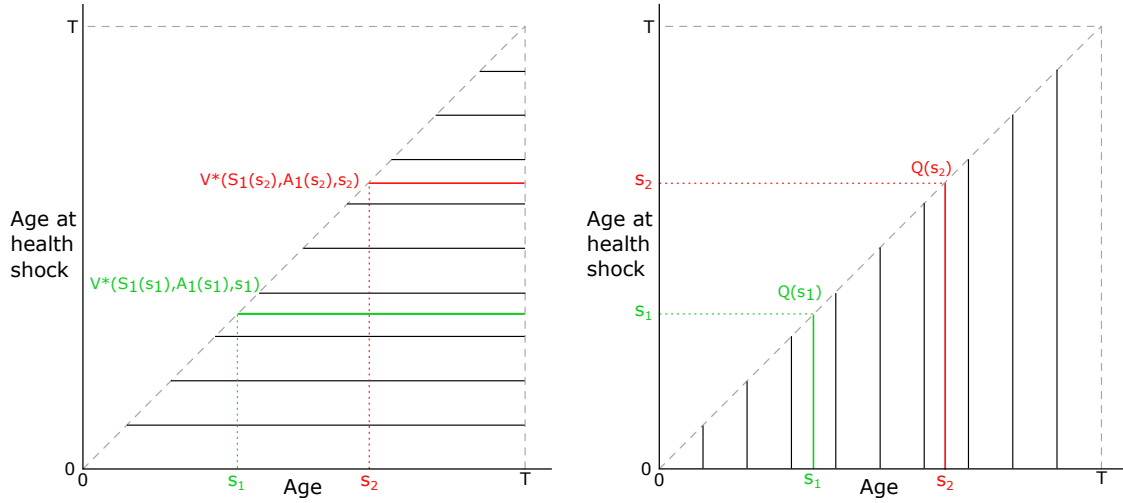


Figure 2.13: Graphical illustration of the change in the order of integration

In a final step we need to introduce another artificial dimension to the variable $Z_2(s)$ as follows:

$$\frac{dZ_2(t, s)}{dt} = 0, \quad t \geq s,$$

$$Z_2(s, s) = Z_1(s)\eta(s, S_1(s), h_1(s))P(S_1(s), d(s)), \quad \forall s \geq 0.$$

This is a pure technical point (i.e. it eliminates the time lag) and ensures that the optimisation problem fits the problem framework of Feichtinger et al. (2003). Finally we can summarise the whole problem formulation in equations (2.50)–(2.66).

$$\max_{\substack{c_1(t), h_1(t), b_1(t) \geq 0 \\ c_2(t, s), h_2(t, s), b_2(t, s) \geq 0 \\ d(t) \geq 0}} \int_0^T e^{-\rho t} [Z_1(t)S_1(t)u^1(c_1(t)) + Q(t)] dt \quad (2.50)$$

$$\text{s.t.} \quad \dot{S}_1(t) = -\mu^1(t, S_1(t), b_1(t))S_1(t) = -\mu^b(t, S_1(t), b_1(t))S_1(t) \quad (2.51)$$

$$S_1(0) = 1, \quad (2.52)$$

$$\dot{A}_1(t) = (r(t) + \bar{\mu}(t))A_1(t) + w^1(t) - c_1(t) - p^b(t)b_1(t) - p^1(t)h_1(t) \quad (2.53)$$

$$A_1(0) = 0, \quad A_1(T) = 0 \quad (2.54)$$

$$\dot{Z}_1(t) = -\eta(t, S_1(t), h_1(t))Z_1(t) \quad (2.55)$$

$$Z_1(0) = 1, \quad (2.56)$$

$$\frac{dS_2(t, s)}{dt} = -\mu^2(t, s, S_2(t, s), b_2(t, s), E(t, s))S_1(t) \quad (2.57)$$

$$= -[\mu^b(t, S_2(t, s), b_2(t, s)) + \mu^m(t, s, E(t, s))] S_2(t, s), \quad t \geq s \quad (2.58)$$

$$S_2(s, s) = S_1(s), \quad \forall s \geq 0 \quad (2.59)$$

$$\frac{dA_2(t, s)}{dt} = (r(t) + \bar{\mu}(t))A_2(t, s) + w^2(t, s, E(t, s)) - c_2(t, s) - p^b(t)b_2(t, s) - p^2(t)h_2(t, s) \quad (2.60)$$

$$A_2(s, s) = A_1(s) - p^d(s)d(s), \quad , \quad A_2(T, s) = 0, \quad \forall s \geq 0 \quad (2.61)$$

$$\frac{dZ_2(t, s)}{dt} = 0 \quad (2.62)$$

$$Z_2(s, s) = Z_1(s)\eta(s, S_1(s), h_1(s))P(S_1(s), d(s)), \quad \forall s \geq 0 \quad (2.63)$$

$$\frac{dE(t, s)}{dt} = f(t, s, E(t, s), h_2(t, s)) \quad (2.64)$$

$$E(s, s) = B(S_1(s), d(s)), \quad \forall s \geq 0 \quad (2.65)$$

$$Q(t) = \int_0^t Z_2(t, s)S_2(t, s)u^2(c_2(t, s), E(t, s)) \, ds \quad (2.66)$$

2.7.2 Necessary optimality conditions

Using the vintage structured model formulation (2.50)–(2.66) we can derive the necessary optimality conditions following the theoretical work of Feichtinger et al. (2003). These conditions consist of (i) a system of differential equations (2.67)–(2.71) for first stage co-state variables, (ii) a system of partial differential equations (2.72)–(2.78) and (iii) a set of first order optimality conditions (accounting for boundary solutions for the controls) (2.79)–(2.92). Note there are not endpoint condition for $\lambda_A(t)$ and $\xi_A(t, s)$ as we included the terminal conditions $A_1(T) = 0$ and $A_2(T, s) = 0 \forall s \geq 0$ in the problem formulation.

$$\begin{aligned} \dot{\lambda}_S = & (\rho + \mu^1 + \mu_{S_1}^1 S_1)\lambda_S + Z_1\eta_{S_1} [\lambda_Z(t) - P\xi_Z(t, t)] - \xi_S(t, t) - \xi_E(t, t)B_{S_1} - \\ & - Z_1 [u^1 + \eta P_{S_1}\xi_Z(t, t)] \end{aligned} \quad (2.67)$$

$$\lambda_S(T) = 0 \quad (2.68)$$

$$\dot{\lambda}_A = (\rho - r - \bar{\mu})\lambda_A - \xi_A(t, t) \quad (2.69)$$

$$\dot{\lambda}_Z = (\rho + \eta)\lambda_Z - \eta P\xi_Z(t, t) - S_1 u^1 \quad (2.70)$$

$$\lambda_Z(T) = 0 \quad (2.71)$$

$$\frac{d\xi_S(t, s)}{dt} = (\rho + \mu^2 + \mu_{S_2}^2 S_2)\xi_S - Z_2 u^2 \quad (2.72)$$

$$\xi_S(T, T) = 0 \quad (2.73)$$

$$\frac{d\xi_A(t, s)}{dt} = (\rho - r - \bar{\mu})\xi_A \quad (2.74)$$

$$\frac{d\xi_Z(t, s)}{dt} = \rho\xi_Z - S_2 u^2 \quad (2.75)$$

$$\xi_Z(T, T) = 0 \quad (2.76)$$

$$\frac{d\xi_E(t, s)}{dt} = (\rho - f_E)\xi_E + \mu_E^2 S_2 \xi_S - w_E^2 \xi_A - S_2 Z_2 u_E^2 \quad (2.77)$$

$$\xi_E(T, T) = 0 \quad (2.78)$$

$$0 \geq Z_1 S_1 u_{c_1}^1 - \lambda_A \quad (2.79)$$

$$0 = (Z_1 S_1 u_{c_1}^1 - \lambda_A) \cdot c_1 \quad (2.80)$$

$$0 \geq -Z_1 \eta_{h_1} (\lambda_Z - P(S_1, d)\xi_Z) - \lambda_A p^1 \quad (2.81)$$

$$0 = (-Z_1\eta_{h_1}(\lambda_Z - P(S_1, d)\xi_Z) - \lambda_A p^1) \cdot h_1 \quad (2.82)$$

$$0 \geq -\mu_{b_1}^1 S_1 \lambda_S - \lambda_A p^b \quad (2.83)$$

$$0 = (-\mu_{b_1}^1 S_1 \lambda_S - \lambda_A p^b) \cdot b_1 \quad (2.84)$$

$$0 \geq \xi_E(t, t)B_d + Z_1\eta\xi_Z(t, t)P_d - \xi_A(t, t)p^d \quad (2.85)$$

$$0 = (\xi_E(t, t)B_d + Z_1\eta\xi_Z(t, t)P_d - \xi_A(t, t)p^d) \cdot d \quad (2.86)$$

$$0 \geq Z_2 S_2 u_{c_2}^2 - \xi_A(t, s) \quad (2.87)$$

$$0 = (Z_2 S_2 u_{c_2}^2 - \xi_A(t, s)) \cdot c_2 \quad (2.88)$$

$$0 \geq \xi_E(t, s)f_{h_2} - \xi_A(t, s)p^2 \quad (2.89)$$

$$0 = (\xi_E(t, s)f_{h_2} - \xi_A(t, s)p^2) \cdot h_2 \quad (2.90)$$

$$0 \geq -\mu_{b_2}^2 S_2 \xi_S - \xi_A(t, s)p^b \quad (2.91)$$

$$0 = (-\mu_{b_2}^2 S_2 \xi_S - \xi_A(t, s)p^b) \cdot b_2 \quad (2.92)$$

2.7.3 Proof of proposition 1 on the valuations of health

Following the calculations of Rosen (1988) we derive the following terms for the valuations:

$$\psi_H^1(t) = \frac{S_1 \lambda_S}{\lambda_A} \quad (2.93)$$

$$\psi_H^2(t, s) = \frac{S_2 \xi_S}{\xi_A} \quad (2.94)$$

$$\psi_P(t) = \frac{Z_1(\lambda_Z - P\xi_Z(t, t))}{\lambda_A} \quad (2.95)$$

$$\psi_{AS}(t) = \frac{Z_1(t)\eta(S_1(t), h_1(t))\xi_Z(t, t)}{\xi_A(t, t)} \quad (2.96)$$

$$\psi_M(t, s) = \frac{-\xi_E(t, s)}{\xi_A(t, s)} \quad (2.97)$$

Since we assume an interior solution for the consumption profiles, the first order conditions for consumption (2.79) and (2.87) have to be fulfilled in strict form and we can derive the following equations.

$$\lambda_A(t) = Z_1(t)S_1(t)u_{c_1}(c_1(t))$$

$$\xi_A(t) = Z_2(t, s)S_2(t, s)u_{c_2}(c_2(t, s), E(t, s))$$

$$\forall t \in [0, T] \text{ and } s \in [0, t]$$

Furthermore we can easily derive the following relationship, which will be useful in the calculations below.

$$\lambda_A(t) = \frac{\xi_A(t, t)}{\eta(t)P(t)} \frac{u_{c_1}^1(t)}{u_{c_2}^2(t, t)}. \quad (2.98)$$

In a next step we summaries the explicit solutions for the costate ODEs and PDEs.

ξ_Z - dynamics

From equation (2.75) together with the terminal condition $\xi_Z(T, s) = 0$ it directly follows that

$$\xi_Z(t, s) = \int_t^T e^{-\rho(\tau-t)} S_2(\tau, s) u^2(\tau, s) d\tau.$$

 ξ_S - dynamics

From equation (2.72) together with the terminal condition $\xi_S(T, s) = 0$ it directly follows that

$$\begin{aligned} \xi_S(t, s) &= \int_t^T e^{-\int_t^\tau [\rho + \mu^2 + \mu^2 S_2] d\tau'} Z_2(\tau, s) u^2(\tau, s) d\tau \\ &= \int_t^T e^{-\rho(\tau-t)} Z_1(s) \eta(s) P(s) u^2(\tau, s) \underbrace{e^{-\int_t^\tau \mu^2 d\tau'}}_{\frac{S_2(\tau, s)}{S_2(t, s)}} e^{-\int_t^\tau \mu^2 S_2(\tau', s) d\tau'} d\tau \\ &= \frac{Z_1(s) \eta(s) P(s)}{S_2(t, s)} \int_t^T e^{-\rho(\tau-t)} S_2(\tau, s) u^2(\tau, s) \underbrace{e^{-\int_t^\tau \mu^2 S_2(\tau', s) d\tau'}}_{\geq 1} d\tau \\ &\geq \frac{Z_1(s) \eta(s) P(s)}{S_2(t, s)} \xi_Z(t, s). \end{aligned}$$

 ξ_A - dynamics

As there is no terminal condition for ξ_A we can only derive the following equation from equation (2.74) which will still be crucial further down below.

$$\xi_A(t, s) = \xi_A(s, s) \cdot e^{\int_s^t (\rho - r - \bar{\mu}) ds'} = \xi_A(T, s) \cdot e^{-\int_t^T (\rho - r - \bar{\mu}) ds'}$$

 ξ_E - dynamics

The dynamics for $\xi_E(t, s)$ described in equation (2.77) together with the terminal condition $\xi_E(T, s) = 0$ yield

$$\xi_E(t, s) = \int_t^T e^{-\int_t^\tau [\rho - f_E(\tau', s)] d\tau'} [w_E^2(\tau, s) \xi_A(\tau, s) + Z_2(\tau, s) S_2(\tau, s) u_E^2(\tau, s) - \mu_E^2(\tau, s) S_2(\tau, s) \xi_S(\tau, s)] d\tau$$

 λ_Z - dynamics

Similar to the ξ_Z dynamics from equation (2.70) together with the terminal condition $\lambda_Z(T) = 0$ it directly follows that

$$\lambda_Z(t) = \int_t^T e^{-\int_t^\tau (\rho + \eta) ds'} [\eta(\tau) P(\tau) \xi_Z(\tau, \tau) + S_1(\tau) u^1(\tau)] d\tau$$

 λ_A - dynamics

For our calculation of the λ_A -dynamics we first use equation (2.98) to modify the λ_A -equation what results in

$$\begin{aligned} \dot{\lambda}_A &= (\rho - r - \bar{\mu}) \lambda_A - \xi_A(t, t), \\ \dot{\lambda}_A &= (\rho - r - \bar{\mu}) \lambda_A - \eta(t) P(t) \frac{u_{c_2}^2(t, t)}{u_{c_1}^1(t)} \lambda_A(t), \\ \dot{\lambda}_A &= \left(\rho - r - \bar{\mu} - \eta(t) P(t) \frac{u_{c_2}^2(t, t)}{u_{c_1}^1(t)} \right) \lambda_A(t). \end{aligned}$$

Now similar to the ξ_A -dynamics, without a terminal condition, we can still derive the equation below:

$$\lambda_A(s) = \exp\left(\int_t^s \rho - r - \bar{\mu}(s') - \eta(s')P(s') \frac{u_{c_2}^2(s', s')}{u_{c_1}^1(s')} ds'\right) \lambda_A(t).$$

λ_S - dynamics

Finally from equation (2.67) together with the terminal condition $\lambda_S(T) = 0$ we can derive the following representation of $\lambda_S(t)$.

$$\begin{aligned} \lambda_S(t) = & \int_t^T e^{-\int_t^\tau (\rho + \mu^1(\tau') + \mu_{S_1}^1 S_1) d\tau'} \left[Z_1 u^1 + \xi_E(\tau, \tau) B_{S_1} + \xi_S(\tau, \tau) - \right. \\ & \left. - Z_1 \eta_{S_1} [\lambda_Z(\tau) - P(\tau) \xi_Z(\tau, \tau)] + \eta P_{S_1}(\tau) Z_1(\tau) \xi_Z(\tau, \tau) \right] d\tau \end{aligned}$$

Coming back to the different valuation of health, we are going to start with the simpler ones of the second stage and proceed to the more involved ones of the first stage.

Value of health in the second stage $\psi_H^2(t, s)$

$$\begin{aligned} \psi_H^2(t, s) &= \frac{S_2(t, s) \xi_S(t, s)}{\xi_A(t, s)} = \frac{S_2(t, s) \frac{Z_2(t, s)}{S_2(t, s)} \int_t^T e^{-\rho(\tau-t)} S_2(\tau, s) u^2(\tau, s) e^{-\int_t^\tau \mu_{S_2}^2 S_2(\tau', s) d\tau'} d\tau}{\xi_A(t, s)} \\ &= \int_t^T \left(e^{-\rho(\tau-t)} Z_2(\tau, s) S_2(\tau, s) u^2(\tau, s) e^{-\int_t^\tau \mu_{S_2}^2 S_2(\tau', s) d\tau'} \frac{1}{\xi_A(\tau, s)} e^{\int_t^\tau \rho - r - \bar{\mu} d\tau'} \right) d\tau \\ &= \int_t^T e^{-\int_t^\tau r + \bar{\mu} + \mu_{S_2}^2 S_2(\tau', s) d\tau'} \frac{Z_2(\tau, s) S_2(\tau, s) u^2(c_2(\tau, s), E(\tau, s))}{Z_2(\tau, s) S_2(\tau, s) u_{c_2}^2(c_2(\tau, s), E(\tau, s))} d\tau \\ &= \int_t^T e^{-\int_t^\tau r + \bar{\mu} + \mu_{S_2}^2 S_2(\tau', s) d\tau'} \frac{u^2(c_2(\tau, s), E(\tau, s))}{u_{c_2}^2(c_2(\tau, s), E(\tau, s))} d\tau \end{aligned}$$

Value of morbidity $\psi_M(t, s)$

$$\begin{aligned} \psi_M(t, s) &= \frac{-\xi_E(t, s)}{\xi_A(t, s)} = \frac{\int_t^T \left(e^{-\int_t^\tau \rho - f_E d\tau'} [-w_E^2(\tau, s) \xi_A(\tau, s) - Z_2(\tau, s) u_E^2(\tau, s) + \mu_E^2(\tau, s) S_2(\tau, s) \xi_S(\tau, s)] \right) d\tau}{\xi_A(t, s)} \\ &= \int_t^T e^{-\int_t^\tau \rho - f_E d\tau'} \left[-w_E^2(\tau, s) - \frac{Z_2(\tau, s) S_2(\tau, s) u_E^2(\tau, s)}{\xi_A(\tau, s)} + \mu_E^2(\tau, s) \frac{S_2(\tau, s) \xi_S(\tau, s)}{\xi_A(\tau, s)} \right] e^{\int_t^\tau \rho - r - \bar{\mu} d\tau'} d\tau \\ &= \int_t^T e^{-\int_t^\tau r + \bar{\mu} - f_E d\tau'} \left[-w_E^2(\tau, s) - \frac{u_E^2(\tau, s)}{u_{c_2}^2(\tau, s)} + \mu_E^2(\tau, s) \psi_H^2 \right] d\tau \end{aligned}$$

Value of acute survival $\psi_{AS}(t)$

$$\begin{aligned} \psi_{AS}(t) &= \frac{Z_1(t) \eta(t) \xi_Z(t, t)}{\xi_A(t, t)} = \frac{Z_1(t) \eta(t) \int_t^T e^{-\int_t^\tau \rho d\tau'} S_2(\tau, t) u^2(\tau, t) d\tau}{\xi_A(t, t)} \\ &= Z_1(t) \eta(t) \int_t^T e^{-\int_t^\tau \rho d\tau'} e^{\int_t^\tau \rho - r - \bar{\mu} d\tau'} \frac{S_2(\tau, t) u^2(\tau, t)}{\xi_A(\tau, t)} d\tau \\ &= Z_1(t) \eta(t) \int_t^T e^{-\int_t^\tau (r + \bar{\mu}) d\tau'} \frac{S_2(\tau, t) u^2(\tau, t)}{Z_2(\tau, t) S_2(\tau, t) u_{c_2}^2(\tau, t)} d\tau \end{aligned}$$

$$\begin{aligned}
&= Z_1(t)\eta(t) \int_t^T e^{-\int_t^\tau (r+\bar{\mu})d\tau'} \frac{u^2(\tau, t)}{Z_1(t)\eta(t)P(t)u_{c_2}^2(\tau, t)} d\tau \\
&= \frac{1}{P(t)} \int_t^T \left[e^{-\int_t^\tau (r+\bar{\mu})d\tau'} \frac{u^2(c_2(\tau, t), E(\tau, t))}{u_{c_2}^2(c_2(\tau, t), E(\tau, t))} \right] d\tau \\
&=: \frac{1}{P(t)} \psi_{life}^2(t, t)
\end{aligned}$$

Value of prevention ψ_P

$$\begin{aligned}
\psi_P(t) &= \frac{Z_1(t)(\lambda_Z(t) - P\xi_Z(t, t))}{\lambda_A(t)} = \frac{Z_1(t)\lambda_Z(t)}{\lambda_A(t)} - \frac{Z_1(t)P\xi_Z(t, t)}{\lambda_A(t)} = \Pi_1 - \Pi_2 \\
\Pi_1 &= \frac{Z_1(t)\lambda_Z(t)}{\lambda_A(t)} \\
&= \frac{1}{\lambda_A(t)} \int_t^T e^{-\int_t^\tau \rho ds} [Z_1(\tau)\eta(\tau)P(\tau)\xi_Z(\tau, \tau) + S_1(\tau)Z_1(\tau)u^1(\tau)] d\tau \\
&= \int_t^T e^{-\int_t^\tau \rho ds} e^{\int_t^\tau (\rho - r - \bar{\mu} - \eta P \frac{u_{c_2}^2}{u_{c_1}^1}) ds} \left[\frac{Z_1(\tau)\eta(\tau)P(\tau)\xi_Z(\tau, \tau)}{\lambda_A(\tau)} + \frac{S_1(\tau)Z_1(\tau)u^1(\tau)}{\lambda_A(\tau)} \right] d\tau \\
&= \int_t^T e^{-\int_t^\tau (r + \bar{\mu} + \eta P \frac{u_{c_2}^2}{u_{c_1}^1}) ds} \left[\frac{Z_1(\tau)\eta(\tau)P(\tau)\xi_Z(\tau, \tau)}{\frac{\xi_A(\tau, \tau)}{\eta(\tau)P(\tau)} \frac{u_{c_1}^1(\tau)}{u_{c_2}^2(\tau, \tau)}} + \frac{S_1(\tau)Z_1(\tau)u^1(\tau)}{S_1(\tau)Z_1(\tau)u_{c_1}^1(\tau)} \right] d\tau \\
&= \int_t^T e^{-\int_t^\tau (r + \bar{\mu} + \eta P \frac{u_{c_2}^2}{u_{c_1}^1}) ds} \left[\frac{u^1(\tau)}{u_{c_1}^1(\tau)} + \eta(\tau)P(\tau) \frac{u_{c_2}^2(\tau, \tau)}{u_{c_1}^1(\tau)} P(\tau) \frac{Z_1(\tau)\eta(\tau)\xi_Z(\tau, \tau)}{\xi_A(\tau, \tau)} \right] d\tau \\
&= \int_t^T e^{-\int_t^\tau (r + \bar{\mu} + \eta P \frac{u_{c_2}^2}{u_{c_1}^1}) ds} \left[\frac{u^1(\tau)}{u_{c_1}^1(\tau)} + \eta(\tau)P(\tau) \frac{u_{c_2}^2(\tau, \tau)}{u_{c_1}^1(\tau)} \psi_{life}^2(\tau, \tau) \right] d\tau \\
\Pi_2 &= \frac{Z_1(t)P(t)\xi_Z(t, t)}{\lambda_A(t)} = \frac{Z_1(t)P(t)\xi_Z(t, t)}{\frac{\xi_A(t, t)}{\eta(t)P(t)} \frac{u_{c_1}^1(t)}{u_{c_2}^2(t, t)}} \\
&= \frac{u_{c_2}^2(t, t)}{u_{c_1}^1(t)} P(t) \frac{Z_1(t)\eta(t)\xi_Z(t, t)}{\xi_A(t, t)} P(t) = \frac{u_{c_2}^2(t, t)}{u_{c_1}^1(t)} P(t) \psi_{life}^2(t, t) \\
\psi_P(t) &= \int_t^T e^{-\int_t^\tau (r + \bar{\mu} + \eta P \frac{u_{c_2}^2}{u_{c_1}^1}) ds} \left[\frac{u^1(\tau)}{u_{c_1}^1(\tau)} + \eta(\tau)P(\tau) \frac{u_{c_2}^2(\tau, \tau)}{u_{c_1}^1(\tau)} \psi_{life}^2(\tau, \tau) \right] d\tau - \frac{u_{c_2}^2(t, t)}{u_{c_1}^1(t)} P(t) \psi_{life}^2(t, t)
\end{aligned}$$

Value of health in the first stage $\psi_H^1(t)$

$$\begin{aligned}
\psi_H^1(t) &= \frac{S_1(t)\lambda_S(t)}{\lambda_A(t)} = \frac{S_1(t)}{\lambda_A(t)} \int_t^T e^{-\int_t^\tau (\rho + \mu^1(\tau') + \mu_{S_1}^1 S_1) d\tau'} [Z_1 u^1 - Z_1 \eta_{S_1} [\lambda_Z(\tau) - P(\tau)\xi_Z(\tau, \tau)] + \\
&\quad + \xi_E(\tau, \tau) B_{S_1} + \xi_S(\tau, \tau) + \eta P_{S_1}(\tau) Z_1(\tau) \xi_Z(\tau, \tau)] d\tau \\
&= \int_t^T e^{-\int_t^\tau (\rho + \mu^1 + \mu_{S_1}^1 S_1) d\tau'} e^{\int_t^\tau \mu^1 d\tau'} e^{\int_t^\tau (\rho - r - \bar{\mu} - \eta P \frac{u_{c_2}^2}{u_{c_1}^1}) d\tau'} \left[\frac{S_1(\tau)}{\lambda_A(\tau)} Z_1 u^1 - \right.
\end{aligned}$$

$$\begin{aligned}
& - \frac{S_1(\tau)}{\lambda_A(\tau)} \eta_{S_1} Z_1 [\lambda_Z(\tau) - P(\tau) \xi_Z(\tau, \tau)] + \frac{S_1(\tau)}{\lambda_A(\tau)} \xi_E(\tau, \tau) B_{S_1} + \frac{S_1(\tau)}{\lambda_A(\tau)} \xi_S(\tau, \tau) + \frac{S_1(\tau)}{\lambda_A(\tau)} \eta_{P_{S_1}(\tau)} Z_1(\tau) \xi_Z(\tau, \tau) \Big] d\tau \\
& = \int_t^T e^{-\int_t^\tau (r + \bar{\mu} + \mu_{S_1}^1 S_1 + \eta P \frac{u_{c_2}^2}{u_{c_1}^1}) d\tau'} \left[\frac{u^1(\tau)}{u_{c_1}^1(\tau)} - \eta_{S_1} S_1 \frac{Z_1 [\lambda_Z(\tau) - P(\tau) \xi_Z(\tau, \tau)]}{\lambda_A(\tau)} + \right. \\
& \quad \left. + \eta P \frac{u_{c_2}^2}{u_{c_1}^1} \left(B_{S_1} S_1(\tau) \frac{\xi_E(\tau, \tau)}{\xi_A(\tau, \tau)} + \frac{S_2(\tau, \tau) \xi_S(\tau, \tau)}{\xi_A(\tau, \tau)} + P_{S_1}(\tau) S_1(\tau) \frac{Z_1(\tau) \eta(\tau) \xi_Z(\tau, \tau)}{\xi_A(\tau, \tau)} \right) \right] d\tau \\
& = \int_t^T e^{-\int_t^\tau (r + \bar{\mu} + \mu_{S_1}^1 S_1 + \eta P \frac{u_{c_2}^2}{u_{c_1}^1}) d\tau'} \left[\frac{u^1(\tau)}{u_{c_1}^1(\tau)} - \eta_{S_1} S_1 \psi_P(\tau) + \right. \\
& \quad \left. + \eta P \frac{u_{c_2}^2}{u_{c_1}^1} \left\{ (-B_{S_1}) S_1(\tau) \psi_M(\tau, \tau) + \psi_H^2 + P_{S_1}(\tau) S_1(\tau) \psi_{AS}(\tau) \right\} \right] d\tau \\
& = \int_t^T e^{-\int_t^\tau (r + \bar{\mu} + \mu_{S_1}^1 S_1 + \eta P \frac{u_{c_2}^2}{u_{c_1}^1}) d\tau'} \left[\frac{u^1(\tau)}{u_{c_1}^1(\tau)} + (-\eta_{S_1}) S_1 \psi_P(\tau) + \right. \\
& \quad \left. + \eta P \frac{u_{c_2}^2}{u_{c_1}^1} \left\{ \psi_H^2 + \frac{P_{S_1}(\tau) S_1(\tau)}{P(\tau)} \psi_{life}^2 + (-B_{S_1}) S_1(\tau) \psi_M(\tau, \tau) \right\} \right] d\tau
\end{aligned}$$

2.7.4 Proof of proposition 2 on the first order optimality conditions

Under the assumption of interior solutions the FOCs (2.79), (2.81), (2.83), (2.85), (2.87), (2.89) and (2.91) all have to hold in strict form. Simple rearrangements of these equations and substituting the different valuations of health in equations (2.93)-(2.97) result in the equations presented in proposition 2.

2.7.5 Derivation of the Euler equations

In this section we present the derivation of all the Euler equations presented in the numerical analysis.

c_1 - Dynamics

$$\begin{aligned}
\lambda_A(t) &= Z_1(t) S_1(t) u_{c_1}^1(c_1(t)) \\
\frac{\dot{\lambda}_A(t)}{\lambda_A(t)} &= \frac{\dot{Z}_1(t)}{Z_1(t)} + \frac{\dot{S}_1(t)}{S_1(t)} + \frac{\frac{d}{dt} [u_{c_1}^1(c_1(t))]}{u_{c_1}^1(c_1(t))} \\
\rho - r - \bar{\mu} - \eta P \frac{u_{c_2}^2}{u_{c_1}^1} &= -\eta - \mu^1 + \frac{u_{c_1 c_1}^1 c_1}{u_{c_1}^1 c_1} \frac{\dot{c}_1}{c_1} \\
\frac{\dot{c}_1}{c_1} &= \frac{u_{c_1}^1}{-u_{c_1 c_1}^1 c_1} \left[r - \rho + \bar{\mu} - \mu^1 - \eta + \eta P \frac{u_{c_2}^2}{u_{c_1}^1} \right]
\end{aligned}$$

c_2 - Dynamics

$$\begin{aligned}
\xi_A(t, s) &= Z_2(t, s) S_2(t, s) u_{c_2}^2(c_2(t, s), E(t, s)) \\
\frac{\dot{\xi}_A(t, s)}{\xi_A(t, s)} &= \frac{\dot{Z}_2(t, s)}{Z_2(t, s)} + \frac{\dot{S}_2(t, s)}{S_2(t, s)} + \frac{\frac{d}{dt} [u_{c_2}^2(c_2(t, s), E(t, s))]}{u_{c_2}^2(c_2(t, s), E(t, s))}
\end{aligned}$$

$$\rho - r - \bar{\mu} = 0 - \mu^2 + \frac{u_{c_2}^2 c_2}{u_{c_2}^2} \dot{c}_2 + \frac{u_{c_2}^2 E}{u_{c_2}^2} f$$

$$\frac{\dot{c}_2}{c_2} = \frac{u_{c_2}^2}{-u_{c_2}^2 c_2} \left[r - \rho + \bar{\mu} - \mu^2 + \frac{u_{c_2}^2 E}{u_{c_2}^2} f \right]$$

b₂ - Dynamics

$$p^b(t) = [-\mu_{b_2}^2(t)] \cdot \psi_H^2(t)$$

$$\frac{\dot{p}^b(t)}{p^b(t)} = \frac{\frac{d}{dt} [-\mu_{b_2}^2(t)]}{-\mu_{b_2}^2(t)} + \frac{\dot{\psi}_H^2(t)}{\psi_H^2(t)}$$

$$\frac{\dot{p}^b}{p^b} = \frac{\mu_{b_2}^2 b_2 \dot{b}_2}{\mu_{b_2}^2 b_2} + \frac{\mu_{b_2}^2 S_2 \dot{S}_2}{\mu_{b_2}^2} + \frac{\mu_{b_2}^2}{\mu_{b_2}^2} + \frac{\dot{\psi}_H^2(t)}{\psi_H^2(t)}$$

$$\frac{\dot{b}_2}{b_2} = \frac{-\mu_{b_2}^2}{\mu_{b_2}^2 b_2} \left[\frac{\dot{\psi}_H^2(t)}{\psi_H^2(t)} - \frac{\dot{p}^b}{p^b} - \frac{\mu_{b_2}^2 S_2}{\mu_{b_2}^2} \mu^2 S_2 + \frac{\mu_{b_2}^2}{\mu_{b_2}^2} \right]$$

$$\dot{\psi}_H^2 = (r + \bar{\mu} + \mu_{S_2}^2 S_2) \psi_H^2 - \frac{u^2}{u_{c_2}^2}$$

$$\frac{\dot{b}_2}{b_2} = \frac{-\mu_{b_2}^2}{\mu_{b_2}^2 b_2} \left[r + \bar{\mu} + \mu_{S_2}^2 S_2 - \frac{u^2/u_{c_2}^2}{\psi_H^2} - \frac{\dot{p}^b}{p^b} - \frac{\mu_{b_2}^2 S_2}{\mu_{b_2}^2} \mu^2 S_2 + \frac{\mu_{b_2}^2}{\mu_{b_2}^2} \right]$$

b₁ - Dynamics

$$p^b(t) = [-\mu_{b_1}^1(t)] \cdot \psi_H^1(t)$$

$$\frac{\dot{p}^b(t)}{p^b(t)} = \frac{\frac{d}{dt} [-\mu_{b_1}^1(t)]}{-\mu_{b_1}^1(t)} + \frac{\dot{\psi}_H^1(t)}{\psi_H^1(t)}$$

$$\frac{\dot{p}^b}{p^b} = \frac{\mu_{b_1}^1 b_1 \dot{b}_1}{\mu_{b_1}^1 b_1} + \frac{\mu_{b_1}^1 S_1 \dot{S}_1}{\mu_{b_1}^1} + \frac{\mu_{b_1}^1}{\mu_{b_1}^1} + \frac{\dot{\psi}_H^1(t)}{\psi_H^1(t)}$$

$$\frac{\dot{b}_1}{b_1} = \frac{-\mu_{b_1}^1}{\mu_{b_1}^1 b_1} \left[\frac{\dot{\psi}_H^1(t)}{\psi_H^1(t)} - \frac{\dot{p}^b}{p^b} - \frac{\mu_{b_1}^1 S_1}{\mu_{b_1}^1} \mu^1 S_1 + \frac{\mu_{b_1}^1}{\mu_{b_1}^1} \right]$$

$$\dot{\psi}_H^1 = \left(r + \bar{\mu} + \mu_{S_1}^1 S_1 + \eta P \frac{u_{c_2}^2}{u_{c_1}^1} \right) \psi_H^1 - \frac{u^1}{u_{c_1}^1} + \eta_{S_1} S_1 \psi_P - \eta P \frac{u_{c_2}^2}{u_{c_1}^1} \left\{ \psi_H^2 + \frac{P_{S_1} S_1}{P} \psi_{ife}^2 + (-B_{S_1}) S_1 \psi_M \right\}$$

$$\frac{\dot{\psi}_H^1}{\psi_H^1} = r + \bar{\mu} + \mu_{S_1}^1 S_1 - \frac{u^1/u_{c_1}^1}{\psi_H^1} + \eta_{S_1} S_1 \frac{\psi_P}{\psi_H^1} - \eta P \frac{u_{c_2}^2}{u_{c_1}^1} \left\{ \frac{(\psi_H^2 - \psi_H^1)}{\psi_H^1} + \frac{P_{S_1} S_1}{P} \frac{\psi_{ife}^2}{\psi_H^1} + (-B_{S_1}) S_1 \frac{\psi_M}{\psi_H^1} \right\}$$

$$\frac{\dot{b}_1}{b_1} = \frac{-\mu_{b_1}^1}{\mu_{b_1}^1 b_1} \left[r + \bar{\mu} + \mu_{S_1}^1 S_1 - \frac{u^1/u_{c_1}^1}{\psi_H^1} + \eta_{S_1} S_1 \frac{\psi_P}{\psi_H^1} - \right]$$

$$-\eta P \frac{u_{c_2}^2}{u_{c_1}^1} \left\{ \frac{(\psi_H^2 - \psi_H^1)}{\psi_H^1} + \frac{P S_1 S_1 \psi_{life}^2}{P \psi_H^1} + (-B_{S_1}) S_1 \frac{\psi_M}{\psi_H^1} \right\} - \frac{\dot{p}^b}{p^b} - \frac{\mu_{b_1}^1 S_1}{\mu_{b_1}^1} \mu^1 S_1 + \frac{\mu_{b_1}^1 t}{\mu_{b_1}^1} \Big]$$

h₂ - Dynamics

$$p^2(t) = [-f_{h_2}(t, s)] \cdot \psi_M(t, s)$$

$$\frac{\dot{p}^2(t)}{p^2(t)} = \frac{\frac{d}{dt} [-f_{h_2}(t, s)]}{[-f_{h_2}(t, s)]} + \frac{\dot{\psi}_M(t, s)}{\psi_M(t, s)}$$

$$\frac{\dot{p}^2}{p^2} = \frac{f_{h_2 h_2} h_2}{f_{h_2}} \frac{\dot{h}_2}{h_2} + \frac{f_{h_2 t}}{f_{h_2}} + \frac{f_{h_2 E} \dot{E}}{f_{h_2}} + \frac{\dot{\psi}_M(t, s)}{\psi_M(t, s)}$$

$$\frac{\dot{h}_2}{h_2} = \frac{-f_{h_2}}{f_{h_2 h_2} h_2} \left[\frac{\dot{\psi}_M(t, s)}{\psi_M(t, s)} + \frac{f_{h_2 t}}{f_{h_2}} + \frac{f_{h_2 E} f}{f_{h_2}} - \frac{\dot{p}^2}{p^2} \right]$$

$$\dot{\psi}_M = (r + \bar{\mu} - f_E) \psi_M + w_E^2 + \frac{u_E^2}{u_{c_2}^2} - \mu_E^2 \psi_H^2$$

$$\frac{\dot{h}_2}{h_2} = \frac{-f_{h_2}}{f_{h_2 h_2} h_2} \left[r + \bar{\mu} - f_E + \frac{w_E^2}{\psi_M} + \frac{u_E^2 / u_{c_2}^2}{\psi_M} - \mu_E^2 \frac{\psi_H^2}{\psi_M} + \frac{f_{h_2 t}}{f_{h_2}} + \frac{f_{h_2 E} f}{f_{h_2}} - \frac{\dot{p}^2}{p^2} \right]$$

h₁ - Dynamics

$$p^1(t) = [-\eta_{h_1}(t)] \cdot \psi_P(t)$$

$$\frac{\dot{p}^1(t)}{p^1(t)} = \frac{\frac{d}{dt} [-\eta_{h_1}(t)]}{[-\eta_{h_1}(t)]} + \frac{\dot{\psi}_P(t)}{\psi_P(t)}$$

$$\frac{\dot{p}^1}{p^1} = \frac{\eta_{h_1 h_1} h_1}{\eta_{h_1}} \frac{\dot{h}_1}{h_1} + \frac{\eta_{h_1 S_1} \dot{S}_1}{\eta_{h_1}} + \frac{\eta_{h_1 t}}{\eta_{h_1}} + \frac{\dot{\psi}_P}{\psi_P}$$

$$\frac{\dot{h}_1}{h_1} = \frac{-\eta_{h_1}}{\eta_{h_1 h_1} h_1} \left[\frac{\dot{\psi}_P}{\psi_P} - \frac{\eta_{h_1 S_1} \mu^1 S_1}{\eta_{h_1}} + \frac{\eta_{h_1 t}}{\eta_{h_1}} - \frac{\dot{p}^1}{p^1} \right]$$

$$\begin{aligned} \psi_P(t) &= \int_t^T e^{-\int_t^\tau (r + \bar{\mu} + \eta P \frac{u_{c_2}^2}{u_{c_1}^1}) ds} \left[\frac{u^1(\tau)}{u_{c_1}^1(\tau)} + \eta(\tau) P(\tau) \frac{u_{c_2}^2(\tau, \tau)}{u_{c_1}^1(\tau)} \psi_{life}^2(\tau, \tau) \right] d\tau - \frac{Z_1(t) P(t) \xi_Z(t, t)}{\lambda_A(t)} \\ &= \Pi_1 - \Pi_2 \end{aligned}$$

$$\dot{\Pi}_1 = \left(r + \bar{\mu} + \eta P \frac{u_{c_2}^2}{u_{c_1}^1} \right) \Pi_1 - \frac{u^1}{u_{c_1}^1} - \eta P \frac{u_{c_2}^2}{u_{c_1}^1} \psi_{life}^2$$

$$\dot{\Pi}_2 = \left(\frac{\dot{Z}_1}{Z_1} + \frac{\dot{P}}{P} + \frac{\frac{d}{dt} \xi_Z(t, t)}{\xi_Z} - \frac{\dot{\lambda}_A}{\lambda_A} \right) \Pi_2$$

$$= \left(-\eta + \frac{\dot{P}}{P} + \frac{\dot{\xi}_Z}{\xi_Z} + \frac{\frac{\partial}{\partial s} \xi_Z}{\xi_Z} - \left(\rho - r - \bar{\mu} - \eta P \frac{u_{c_2}^2}{u_{c_1}^1} \right) \right) \Pi_2$$

$$\frac{\dot{\xi}_Z}{\xi_Z} = \frac{\rho \xi_Z - S_2 u^2}{\xi_Z} = \rho - \frac{S_2 u^2}{\xi_Z} = \rho - S_2 u^2 \frac{Z_1 \eta}{\xi_A \psi_{AS}} = \rho - \frac{Z_1 \eta S_2 u^2}{Z_2 S_2 u_{c_2}^2 \psi_{AS}} = \rho - \frac{Z_1 \eta S_2 u^2}{Z_1 \eta P S_2 u_{c_2}^2 \psi_{AS}} = \rho - \frac{u^2}{u_{c_2}^2 \psi_{life}^2}$$

$$\begin{aligned} \dot{\Pi}_2 &= \left(r + \bar{\mu} - \eta + \eta P \frac{u_{c_2}^2}{u_{c_1}^1} + \frac{\dot{P}}{P} + \frac{\frac{\partial}{\partial s} \xi_Z}{\xi_Z} \right) \Pi_2 - \frac{u^2}{u_{c_2}^2 \psi_{life}^2} \Pi_2 \\ &= \left(r + \bar{\mu} - \eta + \eta P \frac{u_{c_2}^2}{u_{c_1}^1} + \frac{\dot{P}}{P} + \frac{\frac{\partial}{\partial s} \xi_Z}{\xi_Z} \right) \Pi_2 - \frac{u^2}{u_{c_1}^1} P \end{aligned}$$

$$\begin{aligned} \dot{\psi}_P(t) &= \dot{\Pi}_1 - \dot{\Pi}_2 = \left(r + \bar{\mu} + \eta P \frac{u_{c_2}^2}{u_{c_1}^1} \right) \Pi_1 - \frac{u^1}{u_{c_1}^1} - \eta P \frac{u_{c_2}^2}{u_{c_1}^1} \psi_{life}^2 - \\ &\quad - \left(\left(r + \bar{\mu} - \eta + \eta P \frac{u_{c_2}^2}{u_{c_1}^1} + \frac{\dot{P}}{P} + \frac{\frac{\partial}{\partial s} \xi_Z}{\xi_Z} \right) \Pi_2 - \frac{u^2}{u_{c_1}^1} P \right) \\ &= \left(r + \bar{\mu} + \eta P \frac{u_{c_2}^2}{u_{c_1}^1} \right) \psi_P - \frac{(u^1 - P u^2)}{u_{c_1}^1} - \left(\frac{\dot{P}}{P} + \frac{\frac{\partial}{\partial s} \xi_Z}{\xi_Z} \right) \frac{u_{c_2}^2}{u_{c_1}^1} P \psi_{life}^2 \end{aligned}$$

Since it also holds that

$$\xi_Z(t, t) = V^*(t, S_1(t), A_1(t), Z_1(t)) =: V^*(t)$$

we can alternatively derive

$$\dot{\psi}_P(t) = \left(r + \bar{\mu} + \eta P \frac{u_{c_2}^2}{u_{c_1}^1} \right) \psi_P - \frac{u^1}{u_{c_1}^1} + \left(\rho - \frac{\dot{P}}{P} - \frac{\frac{d}{dt} V^*(t)}{V^*(t)} \right) \frac{u_{c_2}^2}{u_{c_1}^1} P \psi_{life}^2$$

$$\begin{aligned} \frac{\dot{\psi}_P}{\psi_P} &= r + \bar{\mu} + \eta P \frac{u_{c_2}^2}{u_{c_1}^1} - \frac{(u^1 - P u^2)/u_{c_1}^1}{\psi_P} - \left(\frac{\dot{P}}{P} + \frac{\frac{\partial}{\partial s} \xi_Z}{\xi_Z} \right) \frac{u_{c_2}^2}{u_{c_1}^1} \frac{P \psi_{life}^2}{\psi_P} \\ &= r + \bar{\mu} + \eta P \frac{u_{c_2}^2}{u_{c_1}^1} - \frac{u^1/u_{c_1}^1}{\psi_P} + \left(\rho - \frac{\dot{P}}{P} - \frac{\frac{d}{dt} V^*(t)}{V^*(t)} \right) \frac{u_{c_2}^2}{u_{c_1}^1} \frac{P \psi_{life}^2}{\psi_P} \end{aligned}$$

$$\begin{aligned} \frac{\dot{h}_1}{h_1} &= \frac{-\eta_{h_1}}{\eta_{h_1} h_1} \left[r + \bar{\mu} + \eta P \frac{u_{c_2}^2}{u_{c_1}^1} - \frac{(u^1 - P u^2)/u_{c_1}^1}{\psi_P} - \left(\frac{\dot{P}}{P} + \frac{\frac{\partial}{\partial s} \xi_Z}{\xi_Z} \right) \frac{u_{c_2}^2}{u_{c_1}^1} \frac{P \psi_{life}^2}{\psi_P} + \frac{\eta_{h_1} S_1 \dot{S}_1}{\eta_{h_1}} + \frac{\eta_{h_1} t}{\eta_{h_1}} - \frac{p^1}{p^1} \right] \\ &= \frac{-\eta_{h_1}}{\eta_{h_1} h_1} \left[r + \bar{\mu} + \eta P \frac{u_{c_2}^2}{u_{c_1}^1} - \frac{u^1/u_{c_1}^1}{\psi_P} + \left(\rho - \frac{\dot{P}}{P} - \frac{\frac{d}{dt} V^*(t)}{V^*(t)} \right) \frac{u_{c_2}^2}{u_{c_1}^1} \frac{P \psi_{life}^2}{\psi_P} + \frac{\eta_{h_1} S_1 \dot{S}_1}{\eta_{h_1}} + \frac{\eta_{h_1} t}{\eta_{h_1}} - \frac{p^1}{p^1} \right] \end{aligned}$$

2.7.6 Estimation and calibration strategies

Estimation of cancer incidence rate $\eta(S_1)$

For the estimation of the cancer incidence rate we adjust the total age-specific mortality rate obtained

from the human mortality database (University of California, Berkeley (USA) and Max Planck Institute for Demographic Research (Germany) (2020)) for the average age-specific cancer mortality provided in the SEER-database (Surveillance Research Program, National Cancer Institute (2020)). Using this adjusted mortality profile, we can derive the empirical equivalent of the S_1 -profile in our model. Using the age-specific cancer incidence rate (again from the SEER database), the functional form

$$\eta(t, S) = \frac{\beta_0}{1 + \beta_1 \left(\frac{1-S_1}{S_1}\right)^{\beta_2}}$$

turns out to provide an appropriate fit. Consequently we use a general least square fitting function for non-linear functions to obtain estimates for β_0 , β_1 , and β_2 .

Estimation of cancer specific mortality rate $\mu^m(s, E)$

The estimation procedure the parameters of the cancer specific mortality rate is more involved. Using the duration dependent cancer specific survival data for different ages at diagnosis (4 broad age groups), we first derive the corresponding empirical information about the cancer mortality rate. The remaining procedure is less straight forward. Compared to the estimation of the cancer incidence rate, we do not have an empirical counterpart for the development of the cancer stock E . First note, that we can derive the empirical duration specific mortality only between full years. However especially within the first two years after a diagnosis, there can be significant differences in cancer mortality risk and expenditure between the beginning and the end of the first year. Hence the derived empirical mortality rates are in a continuous time setting more appropriate estimates for mortality at 6 months, 1.5 years, 2.5 years, etc. Furthermore our specified cancer mortality function

$$\mu^m(t, s, E) = \psi_0 \cdot E \cdot \exp \left\{ \psi_1 \cdot \left(\frac{s}{T}\right)^{\psi_2} \right\}$$

and the normalisation of the initial cancer stock $E(s, s) = 1.0$ imply that we can potentially estimate the impact of age at diagnosis using mortality data right at the point of diagnosis. Hence we use the cancer specific mortality data for 6 months, 1.5 years, 2.5 years to estimate a non-linear function

$$M^j(t-s) = \exp \left\{ -a_1 + a_2 \cdot e^{-a_3(t-s)} \right\}$$

for each age-group $j \in \{1, 2, 3, 4\}$ separately (which provides excellent fit) and use it to derive extrapolated estimations for the mortality rates right after the diagnosis, i.e. $M^j(0)$.

However the data shows that the impact of the age at diagnosis can change for increasing duration of cancer. Therefore estimating ψ_1 and ψ_2 only from data $M^j(0)$, might lead to skewed results, which are not able to replicate the mortality risk of cancer also in the long-run. Hence we evaluated $M^j(t)$ at $t = 0, 1, \dots, 10$ to obtain a relevant data-set \mathcal{M} . While we still do not know the development of E ex-ante, we can estimate ψ_0 , ψ_1 , and ψ_2 as follows:

- (i) We assume a highly general mortality function (with $d = t - s$ describing the duration of the diagnosed cancer disease) in the form of

$$\widetilde{\mu}^m(d, s) = \nu_d \cdot \exp \left\{ \zeta_1 \cdot \left(\frac{s}{T}\right)^{\zeta_2} \right\}$$

- (ii) Now we can use the dataset \mathcal{M} to estimate the 13 parameters $\nu_0, \nu_1 \dots \nu_{10}$, ζ_1 , and ζ_2 using a generalised least-square fit. However to obtain the corresponding mortality rates for the four age-groups, we need to calculate the averaged values within each age-group. Thereby we need to account for the varying incidence of cancer depending on age and use the empirical equivalents to S_1 , Z_1 , and Z_2 , which we can all derive from the estimates of the cancer incidence rate.

- (iii) Finally we can use the estimates for ζ_1 and ζ_2 as estimates for ψ_1 and ψ_1 in μ^m and the estimated value for ν_0 provides an approximation for ψ_0 .

Calibration of the base mortality function $\mu^b(t, b)$

We use a very general form of the mortality function, so we are able to match the health expenditure profile and the survival profile at the same time, i.e. we assume the functional form presented in Section 2.4.

$$\begin{aligned}\mu^b(t, S_1, b_1) &= \mu^b(t, b_1) = g(t)b_1^{\varepsilon(t)} & g(t) &= \exp\{\gamma_0 + \gamma_1 t + \gamma_2 t^2 + \gamma_3 t^3\} \\ \varepsilon(t) &= \alpha_0 + \alpha_1 \cdot t \quad (< 0)\end{aligned}$$

As the data delivers better information about the health expenditure rather than the health investments/measures we will directly connect the mortality rate to the health expenditures and as a result also set the price for health investments $p^b = 1$. To find fitting values for γ_i and α_i we use an iterative procedure:

- (i) We start with a reasonable estimate for the value of health profile in the first stage.
- (ii) Now consider the first-order optimality condition

$$(-\mu_{b_1}^1(t, b_1(t)))\psi_H^1(t) = p^b (= 1)$$

and the fact, that for our functional specification it holds that

$$\mu_{b_1}^1(t, b_1(t)) = \varepsilon(t) \frac{\mu^1(t, b_1(t))}{b_1(t)}$$

- (iii) Consequently we can rewrite the optimality condition as:

$$\varepsilon(t) = \frac{b_1(t)}{-\mu^1(t, b_1(t))\psi_H^1(t)}$$

Inserting the general expenditure data $b_1^{Data}(t)$ obtained from the NTA database, the calculated value of health profile and the non-cancer mortality rates from the data, we get an estimated profile for $\varepsilon(t)$, which can then be used to estimate the parameters α_1 and α_2 .

- (iv) Having found the parameters for $\varepsilon(t)$, we can then simply use the definition of the mortality function to find the parameters $\gamma_i, i \in \{1, 2, 3, 4\}$ by using

$$\ln(g(t)) = \ln\left(\frac{\mu^1(t)}{b(t)^{\varepsilon(t)}}\right)$$

and inserting the mortality and health expenditure data together with the estimate for $\varepsilon(t)$ on the right side of the equation.

- (v) As a next step we undertake several steps in our general optimisation algorithm to get closer to the optimal solution.
- (vi) After a certain number of steps we stop and calculate the a-priori value of health (average value of health in the first and second stage) for the current solution of controls, states and costates and return to step (ii).

As the optimal solution converges during the process, also the calibration for the parameters γ_i and α_i should converge for proper guesses for the initial choices.

Chapter 3

Allocating tests and lockdowns to contain pandemic spread and medical overload within a network

This chapter in its entirety directly corresponds to the paper Freiburger, Grass, Kuhn, Seidl and Wrzaczek (2022a), which is currently under final revision for publication in the Journal of Public Economic Theory.

Abstract

During the COVID-19 pandemic countries invested significant amounts of resources into its containment. In early stages of the pandemic most of the (non-pharmaceutical) interventions can be classified into two groups: (i) testing and identification of infected individuals, (ii) social distancing measures to reduce the transmission probabilities. Furthermore, both groups of measures may, in principle, be targeted at certain sub-groups of a networked population. To study such a problem, we propose an extension of the SIR model with additional compartments for quarantine and different courses of the disease across several network nodes. We develop the structure of the optimal allocation and study a numerical example of three symmetric regions that are subject to an asymmetric progression of the disease (starting from an initial hotspot). Key findings include that (i) for our calibrations policies are chosen in a "flattening-the-curve", avoiding hospital congestion; (ii) policies shift from containing spillovers from the hotspot initially to establishing a symmetric pattern of the disease; and (iii) testing that can be effectively targeted allows to reduce substantially the duration of the disease, hospital congestion and the total cost, both in terms of lives lost and economic costs.

3.1 Introduction

Since the beginning of 2020 the virus SARS-CoV-2 has been spreading around the world at an incredible pace; and by August 2021 COVID-19 (the disease induced by the virus) has cost more than 5.17 million lives.¹ Before an effective and safe vaccination became available in sufficient quantities, which in case of COVID-19 took about a year, mainly two instruments were available to governments: social distancing and testing.²

¹<https://covid19.who.int/>, last access December 1, 2021.

²Note that while our analysis is motivated by the ongoing COVID-19 pandemic (including the calibration of the numerical analysis), our model and the epidemic, economic and demographic structures described as well as the policy mechanisms generalize to many other forms of present-day epidemics. Thus while some of the insights on optimal testing and lockdown policies may come too late in the fight against COVID-19, the insights are worth consideration for future pandemic events.

Social distancing is aimed at reducing the spread of the virus, and for the purpose of our paper, comprises - in a broad sense - measures such as locking down non-essential parts of the economy, travel restrictions, wearing masks in public, requiring distance between people in public spaces, etc. Many of these measures have been used for decades to fight against infectious diseases (e.g., like the plague, leprosy, or the Spanish flu) and can be implemented quickly at potentially high costs to the economy and the individual (including disutility).

Testing and contact tracing, on the other hand, is essential in order to be able to identify and isolate infected people and break infection chains. It is not surprising that many countries invested significantly into the development and deployment of such measures. While these relatively modern instruments are very useful for identifying and quarantining infected individuals and, thereby, allowing to gain control over the epidemic without locking down large parts of the population, tests and effective means of contact tracing have been in short supply, especially during early parts of the epidemic and under circumstances of overwhelming infection rates.

The implementation of these measures varies substantially across countries. While social distancing and lockdown measures are commonly used with different intensity and focus and duration, testing tracing strategies are used very differently. South Korea, e.g., concentrates on testing with a very efficient tracing strategy to detect infected people before they develop symptoms and are spreading around (*testing with efficient tracing*). Other countries, like Austria, try to test large parts of the population regularly and are implementing tests as a precondition for public entities like restaurants, events, cinemas, etc (*unfocused testing*). As a third strategy a couple of countries are testing only people that develop symptoms. Regular testing of people without having symptoms is not intended (*no testing*).

While there is a growing literature that evaluates social distance and lockdown measures (see the literature review later in this section), the question how testing and tracing is implemented optimally and how it interacts with distance and lockdown measures has not been addressed so far. Another aspect that has attracted only little attention is the question whether these instruments should be applied uniformly across the population or should be varied according to the specific circumstances (in terms of localized infection rates and specific characteristics) of subgroups. This is in contrast to the controversial and shifting public and political debate, which keeps coming back to whether or not targeted travel restrictions are appropriate, and whether social distancing should be focused in a way that shields only vulnerable groups but allows free interaction amongst others. This is paralleled by a debate on whether testing capacities should be focused on particular regional hotspots or on certain subgroups of the population (accompanied by sophisticated tracing strategies), at least as long as test-kits are in limited supply.

What matters in such debates is heterogeneity across the population in two dimensions: (i) The state of disease progression, which typically varies across regions - and sometimes across population subgroups. (ii) The characteristics of sub-groups of the population, where some are more vulnerable in a medical sense, i.e., more prone to get infected and/or suffer from a severe course of the disease; and some may be more vulnerable economically, as captured by them facing higher costs of social distancing. In the following we discuss the general properties of these two dimensions of heterogeneity.

Different regions: As a pandemic, COVID-19 has turned into a problem at global scale. This notwithstanding, countries and even regions differ in the starting date of the epidemic as well as in the pattern of epidemic progression. The advent of several waves of COVID-19 for many countries also speaks evidence to the fact of cyclical patterns, where countries who had acquired control over the disease, lost it again due to spill-overs from other countries. Another crucial issue concerning regionality are differences in the availability of intensive care capacities, with large variation even for countries that are close to each other.³ Similarly, intensive care units (ICU) may be regionally centralized, especially in small countries. For all of these reasons, policy measures may need to be regionally differentiated in their start and length, as well as in their intensity.

³E.g., Germany: 33.9, Italy: 8.6 intensive care beds per 100,000 people (year 2020). See de.statista.com for details.

Heterogeneous population: Scientists agree that COVID-19 is a highly contagious disease that can take widely different courses. While it is mostly harmless for young and healthy people (although a number of long-term effects are already recorded⁴), it can get very critical and even lead to death for old people or people with pre-existing illnesses. Moreover, the social behaviour of people of different age-groups, different professions (people working in the health care sector, or as blue or white collar workers) is heterogeneous (social and working contacts are very different for children, blue and white collared workers, retired people). Hence, the population has to be subdivided in (disjoint) heterogeneous groups with respect to their infection risk, their probability to infect others and their susceptibility to experience a heavy course of the disease.

These arguments underline the importance of extending the existing epidemiological models in two dimensions: (a) to allow for the interaction of heterogeneous sub-groups and for the targeting of policy-instruments to certain groups; and (b) to allow for a more thorough analysis of the role of testing within a setting of optimal policy-making. While some attempts have been made in both directions, we will demonstrate in our literature review that these are to some extent patchy and focusing on specific contexts (e.g., analysis of exogenous variations in (non-optimal) policies within decentralized economies). Thus, we believe to make headway with a coherent analysis of optimal policy-making from a social planner perspective.

We propose an extension of the SIR model with additional compartments for quarantine and different courses of the disease (light, heavy and detected). We consider the population to be distributed across a network (representing different regions or social groups, respectively) and model the diffusion of the virus between its nodes. We allow for heterogeneity across the network: Regions can be heterogeneous with respect to size and population density (implying different patterns of population interaction); social groups may differ with respect to parameters such as mortality and the number of interactions with other groups. Policy measures (testing and social distance/contact reduction) are allowed to be targeted in a way that minimizes the social cost of the pandemic. Tests are provided subject to a capacity constraint, and we consider variation in the effectiveness of testing (depending e.g., on whether or not tests can be targeted in the presence of contact tracing), which allows us to study the interaction of test efficacy with the level of contact reduction.

After providing and interpreting the first-order conditions that govern optimal policy-making and deriving properties of the optimal solution, we turn to numerical analysis. Here, we study the optimal control of a Covid-like disease⁵ spreading across three different regions. The regions are assumed to be symmetric and only differ in the starting conditions (i.e., the disease spreads from an initial hot-spot) such that the effects from the network structure are not dominated by regional asymmetries. We compare the optimal allocation and outcomes across four regimes: a regime of uncontrolled disease progression, a regime with lockdowns without any testing, and for two regimes with lockdown and testing, allowing for targeted testing in one case. While the social cost (in terms of economic loss, medical treatment costs and value of lives lost) is reduced significantly by a relatively rigorous lockdown, the introduction of testing only yields significant cost reductions if tests can be targeted. Moreover, targeted testing changes the testing and lockdown strategy substantially. Whereas tests are allocated in a cyclical sequence across different regions/sub-groups of the population, following high infection rates; targeted tests follow this pattern only initially but are shared equally across the sub-groups in the following. Policies are aimed at preventing spillover of the disease from the hotspot, but are then geared to the harmonization of the disease course across groups. For our parametrization, they are of a "flattening-of-the-curve" type.

The remainder of the paper is organized as follows. Before introducing our model in greater detail we provide a brief overview of the recent literature on epidemiological models concerning the COVID-19 pandemic in Section 3.2. Section 3.3 presents an extended SIR-type model over a network. Section 3.4

⁴E.g., higher risk for a heart attack or a stroke, reduced lung capacity, taste disorder.

⁵We chose the epidemiological parameters of the disease such as transmission, hospitalisation, mortality and recovery rates to reflect those of Covid-19 found in the literature (see Section 3.5.1 for information on the calibration). However we do not seek to replicate all aspects of what continues to be an emerging condition, but give insight for optimal epidemiological strategies for potential new infectious diseases in the future, which pose similar risks to Covid-19.

continues with optimality conditions and properties of the optimal solution. A numerical scenario in which the disease spreads across three symmetric regions but starting from an asymmetric distribution of initial infections, is developed and analyzed in Section 3.5. Section 3.6 concludes. In the supplementary material, we present a second numerical scenario, involving a network composed of three population subgroups (blue-collar workers facing high costs of lockdown, white-collar workers facing low costs of lockdown, vulnerable individuals, e.g., the elderly, with poorer health outlook upon infection).

3.2 Literature review

Within this literature review, we restrict ourselves to a selection of papers that consider optimal policy making in a dynamic framework (i.e., in particular compartment models⁶ considered with optimal control theory) involving some heterogeneity across the population.

When dealing with the outbreak of an infectious disease, it is important to understand the impact of different policy measures. Optimal control theory provides tools to determine the optimal application of control instruments over the course of an epidemic to keep the direct and indirect harms of the disease as low as possible. While several papers consider only health related objectives (see e.g., Hansen and Day (2011) or Bliman et al. (2021)), there is also a growing literature that takes economic consequences of policy measures into account too. The optimal timing of a lockdown is studied in Caulkins et al. (2020), who found that the optimal duration of a lockdown decreases in its starting time. They analyze the trade off between economic damage and lost lives and identify conditions under which no lockdown, an intermediate and a permanent lockdown is optimal. Caulkins et al. (2021) extend this research by not only allowing the length of the lockdown to be optimally determined but also its intensity. Furthermore, they incorporate the impact of lockdown fatigue, i.e., a declining adherence to policy measures related to their duration, on the optimal strategy. They find that the optimal solution is history-dependent and that under certain conditions multiple lockdowns can be optimal. Caulkins et al. (2022) adjust the models of Caulkins et al. (2020) and Caulkins et al. (2021) for more contagious virus mutations. Alvarez et al. (2021) analyze how the effectiveness of a lockdown, the fatality rate and the value of a statistical life affects the optimal lockdown policy. Not surprisingly, a low fatality rate or a low valued statistical life lead to a shorter, less intense lockdown, however, a more effective lockdown increases the duration of the lockdown instead of decreasing it. The optimal lockdown policy is also studied in Federico and Ferrari (2021). They use an SIR framework where they assume the transmission rate to evolve according to a stochastic differential equation. They characterize three distinct phases in a pandemic: in the first phase, there should be no lockdown, while in the second it should be vigorous. In the third phase of the pandemic the lockdown should be moderate. The optimal containment measures to reduce the spreading of the disease taking economic damage into account have also been subject to investigation by Aspri et al. (2021). Among other things, they emphasize the importance of testing by which they find it is possibly to almost eliminate mortality at low economic costs in a one-region/group model. Ouardighi et al. (2020) take account of social fatigue and popular discontent when determining the optimal application of non-pharmaceutical interventions (mobility restrictions, isolation and securing social interactions) over time.

Acemoglu et al. (2020) emphasize the importance of distinguishing between different risk groups when deciding about policy measures. Their results suggests a strict and long lockdown for the most vulnerable population group and a less strict lockdown for the other groups. Richard et al. (2021) focus on non-pharmaceutical interventions in an age-structured optimal control framework. To minimize deaths and control costs, they recommend a strategy where control measures vary throughout the pandemic and are more intense for the older population. Bonnans and Gianatti (2020) consider an age-structured population and analyze the optimal confinement strategy. The optimal vaccination strategy

⁶SIR models have been used to study infection dynamics for now almost 100 years starting with the seminal paper by Kermack and Mckendrick (1927). The enormous impact of the COVID-19 pandemic has lead to a large number of recent papers using this type of models to analyze different aspects of the crisis. For a general and broad literature review we refer to Bloom et al. (2020).

for different population groups is considered by Grundel et al. (2021). In their optimal control problem they minimize the amount of social distancing subject to the constraints that the number of people requiring ICU care remains beyond a certain threshold and the availability of vaccines is limited. In principle, they recommend vaccinating people with high contact rates first, however, they also find situations in which vaccinating high risk groups is preferable. Angelov et al. (2021) use a distributed optimal control epidemiological model to study the impact of prioritization of age groups with respect to vaccination. They find that a random distribution of vaccines is efficient in reducing the overall number of infections, however, a lower mortality is obtained through prioritization of elderly and of those groups with the highest contact rates. Fabbri et al. (2021) also take account of age-structure within an SIR framework in a more theoretically oriented paper that provides sufficient optimality conditions.

To analyze whether policies should or should not vary between different regions in Italy, Carli et al. (2020) extend their SIR-based framework to a multi-region model. They focus on non-pharmaceutical interventions such as lockdowns and mobility restrictions between different regions. They propose a model predictive control approach to handle potential deviations of the predictive model from the actual development of a pandemic. They compare the impact of interventions that are uniformly applied to different Italian regions and interventions that are applied in a differentiated manner. The considered objective is to minimize economic and health-related costs. They find that the model predictive control approach leads to substantially lower total costs than different benchmark control strategies and propose it as a suitable tool to mitigate future Covid waves. To determine the optimal vaccine distribution among different priority groups, Gamchi et al. (2021) embed a multi-region SIR model with multiple priority groups into a vehicle routing problem. By means of the weighted augmented epsilon constraint method, they study the trade off between the social costs and the costs of vaccinations and the fixed costs of vehicles. The main focus of the paper is on the solution approach that combines the combinatorial VRP with an optimal control problem in a multi-objective framework. Numerical illustrations, however, confirm the importance of paying attention to high-risk groups.

Different risk-groups, testing and the impact of social distancing measures are also considered in Gollier (2020). In this paper, confinement strategies are compared where the degree of confinement is either constant over time or can assume depending on the number of infected two different levels. Our paper extends this research by optimally determining the degree of testing and contact reducing measures.

Gori et al. (2021) focus on the economic impact of the Covid pandemic and analyse the impact of testing to avoid costly lockdowns in a framework that does not differentiate between different regions or population groups, but they explicitly consider the dynamics of capital. The duration of a lockdown is determined by a feedback rule, which depends, in particular, on the prevalence. The lockdown and testing intensity are determined optimally, however, they are assumed to be constant over time. In our paper, we allow the intensity of social distancing measures and testing to vary over time.

3.3 SIR-type Model

For the epidemic dynamics we extend the well-known SIR model (see Kermack and Mckendrick (1927)) over a network denoted by Ω . The network consists of $|\Omega|$ (denoting the cardinality of Ω) network nodes representing different population groups or a distribution according to space (i.e., geographical location).⁷ At every node $j \in \Omega$ we consider the following seven compartments: susceptible individuals $S_j(t)$, infected individuals without or with light symptoms $L_j(t)$, infected individuals without or with light symptoms that are detected $D_j(t)$ (by a positive test), infected individuals with heavy symptoms $H_j(t)$, recovered individuals having never been detected $R_j^L(t)$, recovered individuals having been detected $R_j^D(t)$, and deceased individuals $M_j(t)$.⁸ Table 1 summarizes the seven compartments of our

⁷To shorten the notation we will often use the expression (network) node $j \in \Omega$ for referring to a specific group our region.

⁸As we are considering testing effort as a control variable, it is necessary to distinguish recovered individuals according to whether or not their infection was detected. Recovered that were detected (through testing or through a heavy course

network model and illustrates the refinements in comparison to the classical SIR model.

Symbol	Description	Abbreviation	Classical SIR model
$S_j(t)$	susceptibles		$S(t)$ susceptibles
$L_j(t)$	infected without or with light symptoms	light cases	$I(t)$ infected
$D_j(t)$	diagnosed light cases detected through testing	diagnosed cases	
$H_j(t)$	infected with heavy symptoms (hospitalized)	heavy cases	
$R_j^L(t)$	recovered light cases not detected during infection	recovered light cases	$R(t)$ recovered
$R_j^D(t)$	recovered heavy and diagnosed light cases	recovered diagnosed cases	
$M_j(t)$	deceased		outflow

Table 3.1: Compartments (state variables) of the extended SIR model. The first three columns show the mathematical symbol, the description and the abbreviation used in the paper. The last column relates the compartments to those of the classical SIR model.

Figure 3.1 illustrates the flows between the different compartments on each node $j \in \Omega$ of the network. The compartments in the grey boxes are grouped into the susceptible population (i.e., as a source of infections), the infected population (red - responsible for new infections), the quarantined population (blue - infected population without contact to the susceptible population) and the immune population (green - unable to infect susceptibles and to acquire an infection, i.e., the sink). The flows generally

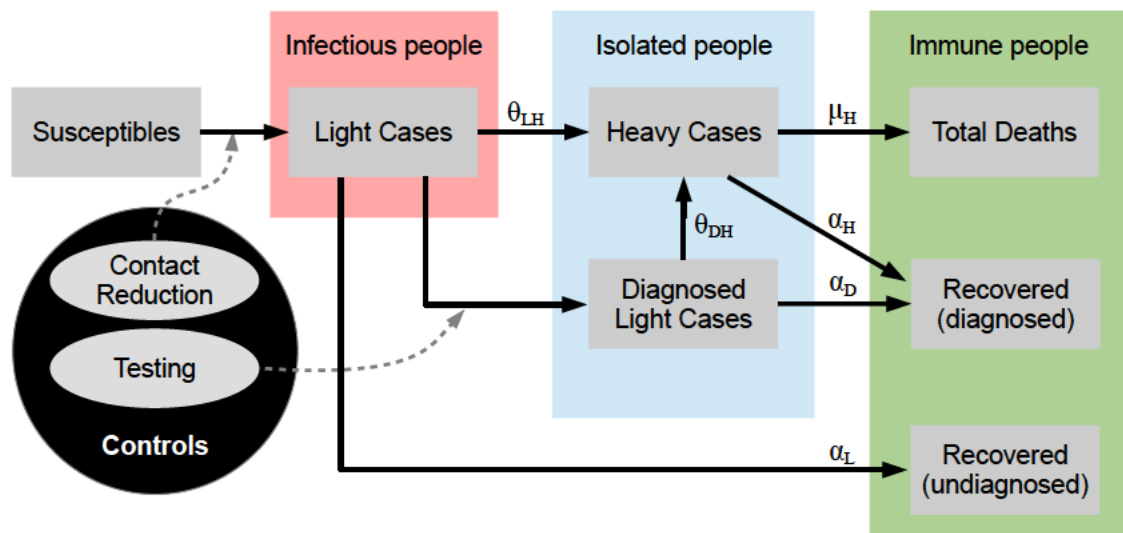


Figure 3.1: Flow chart of the SIR-type model. Grey boxes: compartments, grey ellipses: controls, black arrows: flows between the compartments, dashed arrows: flows that can be changed by controls.

follow the potential different disease progressions of the infection. Susceptibles can get infected and initially only suffer light symptoms (in every scenario), i.e., become a light case. Consequently there a

of disease) are assumed to be immune (at least for several months) implying that they are exempt from further testing. In contrast, recovered individuals that remained undetected cannot be distinguished from susceptibles, implying that they will be tested with the same effort.

three potential paths for the individuals: (i) a light case can recover at rate α_L , (ii) a light case can develop more severe symptoms and escalated to a heavy case, which requires hospital treatment, (iii) a light case can be detected and diagnosed by being tested. In case (i) the individual becomes part of the undiagnosed recovered compartment, which are subsequently assumed to be immunized⁹. A heavy case in (ii) is automatically isolated in the hospital and does not contribute to the spread of the infection any more. Over time the individual has a chance to recover at rate α_H , but also faces a mortality risk at rate μ_H . In the case of fatality they end up the compartment of total deaths, while after recovery they become a diagnosed recovered individual. The diagnosed light cases in (iii) also do not contribute to the infection dynamics as they are isolated (e.g., in home isolation). Being isolated they can either recover or escalate to a heavy case analogously to the undetected light cases. However, their recovery and escalation rates can differ since the diagnosis allows for potential treatment and escalation prevention for these specific cases.

The controls are denoted in the grey ellipses and consist of contact reduction measures and testing efforts. While contact reduction measures impact the flow from susceptibles to light case, testing efforts affect the flow from light cases to diagnosed light cases. Both of these impacts channels are indicated with dashed arrows in Figure 3.1.

The network structure of our framework allows us to model a population, which is heterogeneous with respect to its spatial distribution and its epidemiologically relevant characteristics like profession, infection and mortality risk, system relevance (amongst others). Furthermore these subgroups interact according to their "interconnections" along the arcs¹⁰ (individuals move between the compartments of the same node, i.e., abstracting from flows between nodes, but the size of the flows depends also on the interactions with other nodes), which enables us to describe complex infection dynamics across different subpopulations. For simplicity (and due to a relatively short time horizon of the pandemic being active) we abstract from modelling population inflow (i.e., births or migration) and natural outflow (i.e., non-COVID-19 mortality). Following from the network structure of Ω and the state variables presented in Figure 3.1 at each node, our framework consists of $7 \times |\Omega|$ state variables in total.

In the following sections we will present the dynamics of each state variable, the policy maker's objective and cost functions, and we conclude by presenting the complete problem formulation.

3.3.1 Dynamical system

For the presentation of the state dynamics we follow the flows between the states as in Figure 3.1.

Susceptibles at node $j \in \Omega$ are assumed to be in contact with other susceptibles, light cases or recovered individuals (diagnosed cases are isolated at home and heavy cases are isolated in a hospital) of any node $k \in \Omega$ (i.e., along an undirected arc connecting nodes j and k). Thus susceptibles can only be infected by a contact with an undetected light case (from any node). The probability that a random contact with an individual from node k is infectious corresponds to the share of light cases in the non-isolated population, i.e., $\frac{L_k}{S_k + L_k + R_k^L + R_k^D}$. Note that D_k and H_k do not appear neither in the nominator nor in the denominator of this probability, since diagnosed and heavy cases are isolated. The combination of contact rate of susceptibles in node j with individuals of node k and the rate, at which susceptibles of node j get infected by light cases of node k (in case of contact), at any point in time t is denoted by $u_{k,j}(t)$. From now on we will refer to it as transmission rate from k to j . Consequently,

⁹The ongoing development of the Covid-19 pandemic has shown that loss of immunity (obtained through infection or vaccination) over time significantly contributes to the infection dynamics and is a highly relevant for effective management of a pandemic. As this loss of immunity can have multiple reasons like diminishing efficacy of vaccinations or the emergence of new mutations of the virus, significant modelling effort is required to adequately reflect this aspect of the Covid-19 pandemic and is the goal of the authors in upcoming work. In this paper we want to focus on the network effects within our framework and thereby rely on the more standard SIR base model without possibilities for reinfection.

¹⁰Note that in general we can assume Ω to be a continuous space. The formulation of the dynamics and the derivation of the optimality conditions are quite similar. Then the resulting model is a distributed parameter optimal control model.

the dynamics for the number of susceptibles at any node j can be described by equation (3.1).

$$\dot{S}_j = -S_j \cdot \sum_{k \in \Omega} \left(u_{k,j} \cdot \frac{L_k}{S_k + L_k + R_k^L + R_k^D} \right), \quad S_j(0) = S_{j0}. \quad (3.1)$$

Within our framework we propose that the (non-negative) transmission rate $u_{k,j}(t)$ is a control variable of the decision maker, which has to always remain below the natural transmission rate denoted by $\beta_{k,j}$. Specifically $u_{k,j}(t)$ can be directly chosen as a "target transmission rate" (linearly) in the dynamics, but with non-linear control costs in the objective function (which is discussed after the dynamical system).¹¹ While this is equivalent to the (perhaps more common) modelling of the decision-maker choosing an effort (at diminishing returns) to control transmission, the direct choice of a transmission rate is more directly aligned with policy-makers setting targets in terms of reducing the effective reproduction number (which depends on the transmission rate) below a certain threshold.¹²

The number of **light cases** in node $j \in \Omega$ increases with the inflow of newly infected susceptibles (outflow of (3.1)) and decreases at the (exogenous) rate $\theta_{LH}(j)$ at which light cases turn into heavy cases, at the (exogenous) rate $\alpha_L(j)$ at which light cases recover without treatment, and at the rate at which light cases are detected and quarantined. The detection rate $\frac{v_j}{L_j + \kappa(S_j + R_j^L)}$ sets the number of performed tests $v_j(t)$ (from now on referred to as testing effort at node $j \in \Omega$) in relation to the number of individuals which are potentially infected and still undetected $L_j + \kappa(S_j + R_j^L)$. Although the social planner knows/estimates the number of light cases, he cannot identify who is infected without testing. The parameter $\kappa \in [0, 1]$, which is assumed to be an exogenous parameter, corresponds to different testing capabilities/efficacies. For κ close to zero, the decision maker is able to almost perfectly identify the light cases a priori, i.e., already before testing. This scenario corresponds to the assumption of all infected showing specific symptoms such that testing can be restricted to these individuals. For κ close to one, the decision maker lacks such information and most infections develop asymptomatic. Without any form of contact tracing, the planner can only test close to randomly across the relevant population, which consists of light cases L_j , of susceptibles S_j , and of recovered individuals R_j^L , who were not diagnosed in the past.¹³ However even in case of high rates of asymptomatic infections, implementing efficient contact tracing of positive cases makes the identification of potential infections possible and enables better targeting of tests.¹⁴ In our framework this corresponds to smaller values of κ . Interior values of $\kappa \in (0, 1)$ then reflect scenarios of imperfect testing. The dynamics for the number of light cases then read

$$\begin{aligned} \dot{L}_j &= S_j \cdot \sum_{k \in \Omega} \left(u_{k,j} \cdot \frac{L_k}{S_k + L_k + R_k^L + R_k^D} \right) - \theta_{LH}(j)L_j - \\ &\quad - \alpha_L(j)L_j - \frac{v_j}{L_j + \kappa(S_j + R_j^L)}L_j, \\ L_j(0) &= L_{j0}. \end{aligned} \quad (3.2)$$

Diagnosed light cases at node $j \in \Omega$ are placed in isolation and do not further contribute to the spread of the disease. These individuals can either recover at (exogenous) rate $\alpha_D(j)$ or turn

¹¹We are not specifying whether a given target for the transmission rate is achieved through changes in the contact or infection rate, however we assume, that the more cost-effective measures are applied first, what also motivates the non-linear costs.

¹²This is reflected in many policies being justified by curbing the reproductive number to a value below one, and/or being conditioned on levels of new infections.

¹³Note that the other groups at each node D_j , H_j , R_j^D , M_j were isolated and diagnosed at some point during their infection and are hence not part of the population relevant for testing.

¹⁴Note that $v_j(t)$ must be bounded, as otherwise the state $L_j(t)$ could turn negative. We assure that this is ruled out by assuming a constraint on the total number of performed tests (see (3.9)), reflecting that test kits including protective gear and, especially, laboratory capacity have been a scarce resource in particular at the beginning of the pandemic. This is also in line with the fact that in all countries the number of performed tests per day are below a maximum number of 5% of the total population (see [de.statista.com](https://www.de.statista.com)).

into heavy cases at an (exogenous) rate $\theta_{DH}(j)$. We allow these rates to (possibly) differ from the corresponding transition rates for undetected light cases, reflecting the effects of possible treatment and better monitoring of diagnosed cases. The number of diagnosed cases increases at the rate at which tests are identifying light cases (as explained above). Putting things together, the dynamics of the diagnosed cases at node $j \in \Omega$ read

$$\dot{D}_j = \frac{v_j}{L_j + \kappa(S_j + R_j^L)} L_j - \alpha_D(j) D_j - \theta_{DH}(j) D_j, \quad D_j(0) = D_{j0}. \quad (3.3)$$

Heavy cases are assumed to be diagnosed and treated in a hospital. Therefore, they do not contribute to the spread of the virus any more. Individuals suffering a heavy course enter through the escalation of light infections (detected or undetected) and leave the hospital either through death at rate $\mu_H(\overline{H}_j, j)$ or recovering at rate $\alpha_H(\overline{H}_j, j)$. As we will explain further on below, both rates may depend on hospital congestion, as measured by \overline{H}_j (see equation (3.5)) at node $j \in \Omega$. The dynamics for heavy cases are then given by

$$\begin{aligned} \dot{H}_j &= \theta_{LH}(j) L_j + \theta_{DH}(j) D_j - \alpha_H(\overline{H}_j, j) H_j - \mu_H(\overline{H}_j, j) H_j, \\ H_j(0) &= H_{j0}. \end{aligned} \quad (3.4)$$

To account for potential congestion effects in the medical sector at node $j \in \Omega$ the recovery rate $\alpha_H(\overline{H}_j, j)$ and the mortality rate $\mu_H(\overline{H}_j, j)$ are assumed to depend on the aggregated variable \overline{H}_j , as defined by

$$\overline{H}_j = \sum_{k \in \Omega} H_k f_H(j, k). \quad (3.5)$$

Here, the function $f_H(j, k)$ relates heavy cases from node k to hospital (or, indeed, intensive care) utilization at node j . If, for instance, the network describes social strata of a population all of which are treated within the same hospital, then we have $f_H(j, k) = 1$ for $j, k \in \Omega$ and thus $\overline{H}_j = \overline{H} = \sum_{k \in \Omega} H_k$ for $j \in \Omega$. If, in contrast, the network describes different regions with segregated hospitals, then $f_H(j, k) = 1$ for $k = j$, and zero otherwise, implying that $\overline{H}_j = H_j$. Of course, other forms of $f_H(j, k)$ can be assumed, reflecting situations in which (some) patients from one region may be treated in another region.

Recovered light and diagnosed cases, as well as the **deceased** at node j collect the outflows from the corresponding compartments. Thus, the dynamics read

$$\dot{R}_j^L = \alpha_L(j) L_j, \quad R_j^L(0) = R_{j0}^L \quad (3.6)$$

$$\dot{R}_j^D = \alpha_D(j) D_j + \alpha_H(\overline{H}_j, j) H_j, \quad R_j^D(0) = R_{j0}^D \quad (3.7)$$

$$\dot{M}_j = \mu_H(\overline{H}_j, j) H_j, \quad M_j(0) = M_{j0}. \quad (3.8)$$

3.3.2 Costs and objective function

The decision maker has two control variables: the transmission rate $u_{k,j}(t)$ between light cases at node $k \in \Omega$ and susceptibles at node $j \in \Omega$ (equation (3.1)) and the testing effort $v_j(t)$ at node $j \in \Omega$ (equation (3.2)). Both controls aim at containing the epidemic and lowering the number of COVID-19 deaths (summed in M_j at node $j \in \Omega$). Testing and tracing is a scarce resource. In spring 2020, testing kits (protective gear) and laboratory capacities were scarcely available and had to be rationed. To account for limited availability of testing, we include a constraint on the corresponding control, i.e.,

$$\sum_{j \in \Omega} v_j(t) \leq \overline{V}(t), \quad (3.9)$$

where $\bar{V}(t)$ is the maximum number of available tests for the entire network. The dependency on time t reflects that supply may grow over time.¹⁵

Let $C_U(u_{k,j}(t), t, k, j)$ denote the costs of curtailing the transmission rate $u_{k,j}(t)$ from node k to j (at t) to a level below its unrestricted (natural) level $\beta_{k,j}$. Thus, we assume $C_U(u_{k,j}(t), t, k, j)$ to be convex decreasing (i.e., higher costs for a lower transmission rate) for all $0 \leq u_{k,j}(t) \leq \beta_{k,j}$. Without any intervention, i.e., for $u_{k,j}(t) = \beta_{k,j}$ there are zero costs. The costs for testing efforts at node $j \in \Omega$ are denoted by $C_V(v_j(t), t, j)$ and are convex increasing. Again, we assume that $v_j = 0$ generates no costs. We can then summarize the properties of the cost functions related to the control variables ($j, k \in \Omega$) as follows:

$$\frac{\partial C_U(\cdot)}{\partial u_{k,j}} < 0, \quad \frac{\partial^2 C_U(\cdot)}{\partial u_{k,j}^2} \geq 0, \quad C_U(\beta_{k,j}, t, k, j) = 0 \quad (3.10a)$$

$$\frac{\partial C_V(\cdot)}{\partial v_j} > 0, \quad \frac{\partial^2 C_V(\cdot)}{\partial v_j^2} \geq 0, \quad C_V(0, t, j) = 0. \quad (3.10b)$$

The occurrence of heavy cases (subsumed in the H_j compartment, $j \in \Omega$) leads to two types of (monetary) costs. First, (by definition) all heavy cases have to be treated in a hospital, some of them even in an ICU, implying medical treatment costs $C_M(H_j(t), t, j)$ associated with type j patients at time t . Again we assume these costs to be convex increasing in $H_j(t)$. Second, some of the heavy cases do not survive. The lost lives from group j are weighted by the statistical value of life Ψ , which we assume to be equal across all nodes.

The policy-maker aims at minimizing total costs, composed of the costs for controlling the transmission rates, the costs for testing, the treatment costs and the value of lost lives, over the course of the pandemic from its onset at $t = 0$ to its end at $t = T$ which for the purpose of fixing ideas we identify with the point in time at which an effective and safe vaccine is universally available.¹⁶ Thus the objective function can be formulated as

$$\min_{\substack{v_j(t) \geq 0 \\ 0 \leq u_{k,j}(t) \leq \beta_{k,j}}} \int_0^T e^{-\rho t} \left[\sum_{j \in \Omega} \left(C_M(H_j(t), t, j) + \mu_H(\bar{H}_j(t), j) H_j(t) \cdot \Psi + C_V(v_j(t), t, j) \right) + \sum_{(k,j) \in \Omega^2} C_U(u_{k,j}(t), t, k, j) \right] dt + \quad (3.11)$$

$$+ e^{-\rho T} \mathcal{S}(\mathcal{X}(T), T), \quad (3.12)$$

where $\rho \geq 0$ is the discount rate and where the salvage value function $\mathcal{S}(\mathcal{X}(T), T)$ contains all follow-up costs of the pandemic. Here $\mathcal{X}(T)$ is used as an abbreviation and denotes a vector of all seven compartments across the network. For the salvage value we assume the same costs as for the planning horizon, but since the pandemic is over after T , there is no need for further controls (implying zero control costs). Moreover, our assumption of unlimited availability of an effective and safe vaccine implies that susceptibles cannot get infected any more.^{17,18} Hence, the relevant compartments can be reduced to

¹⁵The control constraints can be altered to account for different properties of the network. If, for instance, the network describes different regions, then each region may face an individual constraint, i.e., $v_j(t) \leq \bar{V}_j(t)$ for $j \in \Omega$.

¹⁶Note that a-priori this terminal time is typically unknown to the policy-maker. However, there is no loss in generality as T can always be chosen large enough for the policy-maker being certain beyond a margin of reasonable doubt that the pandemic would have ended through the availability of a vaccine or otherwise. In the case of COVID-19 for instance, which in most industrialized economies emerged from March 2020 onwards, policy-makers were expecting as early as spring 2020 that vaccines would be available by spring 2021.

¹⁷We assume that all susceptibles in the network are vaccinated within a very short spell of time. The real situation (at beginning of 2021), however, teaches us that vaccination is not an easy task and that its rollout may take considerable time. This lies outside of the scope and focus of our model and is investigated in, e.g., Buratto et al. (2021). See also footnote 9.

¹⁸We assume for our purposes, that the vaccination delivers sterile immunity for all individuals, i.e., the remaining undetected light cases cannot infect anybody after the vaccine becomes available. Without further information this

the light, the diagnosed and the heavy cases, as only these individuals cause costs until there are no infected left. The salvage value function can then be written as

$$\mathcal{S}(\mathcal{X}(T), T) := \int_T^\infty e^{-\rho(t-T)} \left[\sum_{j \in \Omega} C_M(H_j(t), t, j) + \mu_H(\overline{H}_j(t), j) H_j(t) \cdot \Psi \right] dt, \quad (3.13)$$

where the states evolve (without further control) according to ($j \in \Omega$)

$$\dot{L}_j = -(\theta_{LH}(j) + \alpha_L(j)) L_j \quad (3.14a)$$

$$\dot{D}_j = -(\theta_{DH}(j) + \alpha_D(j)) D_j \quad (3.14b)$$

$$\dot{H}_j = \theta_{LH}(j) L_j + \theta_{DH}(j) D_j - (\alpha_H(\overline{H}_j, j) + \mu_H(\overline{H}_j, j)) H_j \quad (3.14c)$$

$$\overline{H}_j = \sum_{k \in \Omega} H_k f_H(j, k). \quad (3.14d)$$

3.3.3 Complete problem

Putting things together the policy-maker faces the following finite time optimal control problem:

$$\begin{aligned} \min_{\substack{v_j(t) \geq 0 \\ 0 \leq u_{k,j}(t) \leq \beta_{k,j}}} \int_0^T e^{-\rho t} \left[\sum_{j \in \Omega} (C_M(H_j, t, j) + \mu_H(\overline{H}_j, j) H_j(t) \cdot \Psi + C_V(v_j, t, j)) \right. \\ \left. + \sum_{(k,j) \in \Omega^2} C_U(u_{k,j}, t, k, j) \right] dt + e^{-\rho T} \mathcal{S}(\mathcal{X}(T), T) \end{aligned} \quad (3.15a)$$

$$\dot{S}_j = -S_j \cdot \sum_{k \in \Omega} \left(u_{k,j} \cdot \frac{L_k}{S_k + L_k + R_k^L + R_k^D} \right) \quad (3.15b)$$

$$\begin{aligned} \dot{L}_j = S_j \cdot \sum_{k \in \Omega} \left(u_{k,j} \cdot \frac{L_k}{S_k + L_k + R_k^L + R_k^D} \right) - \theta_{LH}(j) L_j - \\ - \alpha_L(j) L_j - \frac{v_j}{L_j + \kappa(S_j + R_j^L)} L_j \end{aligned} \quad (3.15c)$$

$$\dot{D}_j = \frac{v_j}{L_j + \kappa(S_j + R_j^L)} L_j - \alpha_D(j) D_j - \theta_{DH}(j) D_j \quad (3.15d)$$

$$\dot{H}_j = \theta_{LH}(j) L_j + \theta_{DH}(j) D_j - \alpha_H(\overline{H}_j, j) H_j - \mu_H(\overline{H}_j, j) H_j \quad (3.15e)$$

$$\dot{R}_j^L = \alpha_L(j) L_j \quad (3.15f)$$

$$\dot{R}_j^D = \alpha_D(j) D_j + \alpha_H(\overline{H}_j, j) H_j \quad (3.15g)$$

$$\dot{M}_j = \mu_H(\overline{H}_j, j) H_j \quad (3.15h)$$

is also the best educated guess the social planner can make at the beginning of the time horizon, when his decisions are determined. The Covid-19 pandemic has shown, that vaccinations are not always able to eliminate the risk of infection completely, while they are still very effective in reducing the number of hospital cases and case mortality. As we will see later in the numerical results, the treatment costs for infected individuals are miniscule compared to the costs for lockdowns and lives lost. Hence our assumption sterile immunity does not significantly effect the salvage function, since the significantly reduced numbers of hospitalisations are highly unlikely to threaten the ICU-capacities and reduced mortality implies very low values in lives lost after the vaccine becomes available.

$$\sum_{j \in \Omega} v_j(t) \leq \bar{V}(t), \quad \bar{H}_j = \sum_{k \in \Omega} H_k f_H(j, k), \quad (3.15i)$$

where $\mathcal{S}(\mathcal{X}(T), T)$ is defined by (defined by (3.13) and (3.14)), and where $\mathcal{X}(0) = \mathcal{X}_0$.

3.4 Properties of the optimal solution

Within this section we first formulate the first order conditions for the control variables (i.e., target transmission rates and number of performed tests) and provide an intuitive understanding (Subsection 3.4.1). Second, we are deriving general properties that help to understand the optimal solution (Subsection 3.4.2).

3.4.1 Optimal allocation

From the Maximum Principle we can formulate the Hamiltonian and derive first order optimality conditions and adjoint equations for the costate variables (see Grass et al. (2008)). The whole set of derivations is relegated to Appendix 3.7.1. Within this section we provide some economic intuition by discussing the first order conditions and some theoretical results that enhance the understanding of the optimal solution. A thorough discussion of the shadow prices relating to the different state variables, which are frequently part of the FOCs, can be found in Appendix 3.7.2.

The first order conditions for the control variables (in the interior of their respective feasible region) can be formulated as follows ($j, k \in \Omega$)

$$\frac{\partial C_U(u_{k,j}, t, k, j)}{\partial u_{k,j}} = (\lambda_{L_j} - \lambda_{S_j}) S_j \cdot \frac{L_k}{S_k + L_k + R_k^L + R_k^D} \quad (3.16a)$$

$$\frac{\partial C_V(v_j, t, j)}{\partial v_j} = (\lambda_{D_j} - \lambda_{L_j}) \cdot \frac{L_j}{L_j + \kappa(S_j + R_j^L)} + \Lambda, \quad (3.16b)$$

where λ_x denotes the shadow price of state variable x ; and where Λ denotes the Lagrangian multiplier relating to the constraint on testing capacity (3.9). The first set of equations (3.16a) determines the optimal target transmission rates (for infections from node k to node j), the second set (3.16b) refers to optimal testing efforts at node j .¹⁹

Equation (3.16a) equates the cost of a marginal reduction in the target transition rate with the benefit of lowering the probability of an additional infection at node j .²⁰ The latter consists of the product of (i) the value of avoiding an infection at node j , as measured by the difference between λ_{L_j} and λ_{S_j} ; (ii) the number of susceptibles at node j ; and (iii) the probability that the contact of a susceptible of node j at node k is infectious (i.e., $\frac{L_k}{S_k + L_k + R_k^L + R_k^D}$).

Analogously, equation (3.16b) equates the cost of a marginal increase in testing effort with the marginal benefit of raising the probability of identifying and isolating a light case at node j . The latter consists of the product of (i) the value of identifying a light case at node j , as given by the difference between λ_{D_j} and λ_{L_j} ; and (ii) the probability that a light case is detected by the test (i.e., $\frac{L_j}{L_j + \kappa(S_j + R_j^L)}$). The Lagrangian multiplier Λ is zero as long as the testing capacity is not exhausted (see complementary slackness condition as formulated in Appendix 3.7.1). In case the testing capacity is in full usage, the multiplier turns negative, implying that the marginal benefit exceeds the marginal cost of testing at all nodes. Under such a situation of rationing, tests are allocated such that the gap between marginal benefits and costs is equilibrated across nodes. This implies, in particular, that all else equal more tests tend to be allocated to those nodes that either exhibit a larger number of light cases L_j , as

¹⁹The presented optimality conditions hold for interior solutions. We omit the explicit presentation of boundary solutions, which is straightforward. The general formulation including boundary controls can be found in Appendix 3.7.1.

²⁰Recall that an increase in the target transition rate lowers costs, i.e., that $\frac{\partial C_U(u_{k,j}(t), t, k, j)}{\partial u_{k,j}(t)} < 0$.

this raises the probability of detecting a light case and, thus, the efficacy of testing; or exhibit a higher value of detection, as is true, e.g., for nodes at which infections are very harmful.

Sufficiency conditions are difficult to deal with in our model. The Mangasarian and Arrow sufficiency conditions (see Grass et al. (2008)) are not promising for our model due to non-convexities in the constraints; a feature which is well-known for models with SIR dynamics (see Gersovitz and Hammer (2004)). Another possibility is presented in Goenka et al. (2021) (see also the discussion in Boucekkine et al. (2021)), who prove local optimality by adopting the Leitmann-Stalford sufficiency conditions (see Leitmann and Stalford (1971)). However, the additional compartments as well as the network structure of our model imply considerable complications so that the proposed method is not applicable.

For the numerical solution we adopted a gradient-based optimisation method developed by Veliov (2003) for age-structured optimal control problems. Following gradual improvements along the direction of the (negative) gradient assures that the numerical solution does not correspond to a maximum of the objective function, which can be the case for techniques using the first order conditions directly. We are aware that this approach still poses the risk of termination of the algorithm in a local optimum. However, we checked our numerical solution for a variety of initial guesses of the control profiles to combat this problem.

3.4.2 Properties

The first order conditions (3.16a) and (3.16b) define the values of the control variables, target transmission rates and testing effort, in the optimal solution in each node j . However, it is not directly clear how they are related. The following Proposition investigates how the optimal target transmission rates are connected to each other and highlights the network aspect of our framework.

Proposition 3

Consider the full finite time optimal control model (3.15) and assume that an optimal solution exists. If the transmission rate from any two nodes k_1 and k_2 to nodes j_1 and j_2 respectively does not lie on the boundary, the relation of the marginal costs of the transmission rates is independent of any shadow price and depends only on the share of infectious people among the non-isolated population in the corresponding node, i.e.,

$$\frac{\frac{\partial C_U(u_{k_1,j_1},t,k_1,j_1)}{\partial u_{k_1,j_1}}}{\frac{\partial C_U(u_{k_2,j_1},t,k_2,j_1)}{\partial u_{k_2,j_1}}} = \frac{\frac{\partial C_U(u_{k_1,j_2},t,k_1,j_2)}{\partial u_{k_1,j_2}}}{\frac{\partial C_U(u_{k_2,j_2},t,k_2,j_2)}{\partial u_{k_2,j_2}}} = \frac{\frac{L_{k_1}}{S_{k_1}+L_{k_1}+R_{k_1}^L+R_{k_1}^D}}{\frac{L_{k_2}}{S_{k_2}+L_{k_2}+R_{k_2}^L+R_{k_2}^D}}, \quad \forall k_1, k_2, j_1, j_2 \in \Omega. \quad (3.17)$$

Proof. Since the transmission rates do not lie on the boundary the corresponding first order conditions for $u_{k_1,j_1}(t)$, $u_{k_2,j_1}(t)$, $u_{k_1,j_2}(t)$ and $u_{k_2,j_2}(t)$ are defined by (3.16a). Dividing one by the other (for fixed j_1, j_2 respectively) proves the above assertion. ■

Therefore, the relation of the marginal costs of the target transmission rates does not depend on the shadow prices, but only on the share of infectious people among the non-isolated population at the different nodes. Interestingly, this means that the optimal target transmission rate at every node (in relation to the other nodes) is set only upon their current status of the pandemic course without considering the dynamics. Consequently, it is enough to set only one target transmission rate optimally with respect to the dynamic effects included in the adjoint variables (according to (3.16a)). Any other target transmission rate being responsible for infections in the same node can be set without considering the dynamic effects.

Especially the relation of marginal costs of the optimal transmission rates from two different node k_1 and k_2 to a common neighbouring node j does not depend on the epidemic situation at node j . On the

other hand a small transformation of equation (3.17), i.e.,

$$\frac{\frac{\partial C_U(u_{k_1,j_1,t},k_1,j_1)}{\partial u_{k_1,j_1}}}{\frac{\partial C_U(u_{k_1,j_2,t},k_1,j_2)}{\partial u_{k_1,j_2}}} = \frac{\frac{\partial C_U(u_{k_2,j_1,t},k_2,j_1)}{\partial u_{k_2,j_1}}}{\frac{\partial C_U(u_{k_2,j_2,t},k_2,j_2)}{\partial u_{k_2,j_2}}},$$

shows that the relation of marginal costs of the optimal transmission rates from any given node k to some (fixed) neighbouring nodes j_1 and j_2 is identical for all possible origin nodes k . Thus the marginal rate of transformation (in terms of the lockdown measures for infection protection of groups j_1 and j_2) is equalised across all nodes of infection origin k .

For an analytical relationship between the optimal testing strategy and the optimal target transmission rates, we will assume an interior solution for testing, where additionally the testing capacity constraint is not binding. We can derive the following equation for testing in node j and the optimal target transmission rate within node j by dividing equations (3.16a) and (3.16b).

$$\frac{\left(\frac{\partial C_U(u_{j,j,t},j,j)}{\partial u_{j,j}}\right) / S_j}{\frac{\partial C_V(v_{j,t},j)}{\partial v_j}} = \frac{(\lambda_{L_j} - \lambda_{S_j})}{(\lambda_{D_j} - \lambda_{L_j})} \cdot \frac{L_j + \kappa(S_j + R_j^L)}{L_j + S_j + R_j^L + R_j^D} \quad (3.18)$$

On the left hand side we obtain the ratio of the marginal costs of a change in the transmission rate per susceptible, i.e., the population which is directly affected by the transmission, and the marginal costs of additional tests. On the right side we first have the ratio of the shadow price of a susceptible getting infected and the shadow price of a light case getting detected via testing. This ratio is multiplied with the share of the potential testing pool with respect to the active (non-isolated) population. Note that the number of infections L_j does not appear as an isolated term in equation (3.18), but only as part of the active population (denominator) and the testing pool (nominator).

This ratio of testing pool and active population plays a crucial rule for the relative allocation of resources towards lockdown measures or testing, but also varies over time. However, for Covid-like diseases (see the numerical solution in Section 3.5) the variations of nominator and denominator are relatively small. The active population at each point in time is equal to the total population without the total number of fatalities M_j and isolated cases ($D_j + H_j$). The latter two terms are orders of magnitude smaller than the total population and as a result the active population is not subject to large variations over time. A similar argumentation also holds for the testing pool, but we have to account for the number of recovered diagnosed cases. In case of high numbers of detections of light cases (through testing efficiency or large testing volumes) we would expect the size of the testing pool to potentially decrease over time. Nevertheless, a more substantial impact on the ratio $\frac{L_j + \kappa(S_j + R_j^L)}{L_j + S_j + R_j^L + R_j^D}$ results from the testing efficiency κ . For highly effective testing (i.e., $\kappa \approx 0$) this ratio would be close to zero (again the number of light cases is at least one order of magnitude smaller than the active population for Covid-like diseases) and consequently imply more tests and less lockdown measures (in case of a convex costs for both measures). For ineffective testing ($\kappa \approx 1$), the ratio is close to one and (with the ratio of shadow price differences becoming the sole defining factor for allocations of tests and lockdowns) conversely suggests higher lockdown measures and less testing efforts.

3.5 Numerical Analysis

In this section we apply the model to study the disease dynamics and the impact of optimal policy-making within a network composed of three heterogeneous groups. We propose a model of three different regions (i.e., a geographic network) that interact with one another and share resources with respect to tests and hospital capacities with an initial infection hot spot in one of them. In the online supplementary material we present an alternative application of our framework, where we assess the potential benefits of targeting social distancing, lockdown and testing measures at specific subgroups of a population distinguished by their economic and demographic make-up.

We assume that the pandemic is terminated by a vaccine that becomes available after one year (i.e., $T = 360$ days) and that no transmission takes place afterwards. We present the dynamic development until day 400, as this fully covers the dynamics underlying the salvage value function in equations (3.13) and (3.14). As we try to focus on the identification of network effects in our model, we assume the three regions to be identical regarding their characteristics and their interactions to be symmetric. Introducing specific heterogeneity in some aspects could obscure effects resulting purely from the network structure. Using this approach we can attribute all heterogeneous measures and developments to the different initial conditions of the different regions. Specifically, we impose:

- (G1) The regions are identical in population size. With the total population being normalized to 1, we then have $N_i(t) = 1/3$ for all t and $i \in \Omega$.
- (G2) Epidemiological and cost parameters are equal for all regions (i.e., regeneration, escalation and mortality rates; costs for treatment, testing, implementing the target transmission rate, and lost lives), see Table 3.2.²¹
- (G3) Hospital capacities are shared between the regions, which means that heavy cases can in general be admitted to hospitals in other regions if the local capacities are exhausted, i.e., $f_H(j, k) \equiv 1$.
- (G4) Testing capacities are shared between the regions.

We do not allow for migration between the nodes of the network, and individuals remain assigned to their initial region at all times. Transmissions across regions occur through temporary travel (and consequent interactions) of some individuals between regions. However, we wish to stress that individuals continue to be assigned to their initial (home) region, as this assignation is crucial for the definition and description of the target transmission rates.

3.5.1 Calibration

We assume, that within and across regions, there is an uncontrolled (baseline) transmission rate of $\beta_{i,j}$. Reducing the transmission rate through social distancing and lockdown measures is assumed to imply quadratic costs. The planer has the possibility to choose target values for the all transmission channels $u_{i,j}$ separately. This means that not only can the transmission targets be chosen differently for interactions within and across different regions, but we explicitly allow for a distinction between the target rates $u_{j,k}$ and $u_{k,j}$. This enables us to model heterogeneous policy schemes, as were enacted by many countries during the pandemic. Notably, mandatory quarantine policies for incoming travellers have been frequently and repeatedly enacted by many countries over the course of the pandemic as a measure to curb the import of infections from other regions. While such measures have often been implemented to varying degrees, applying, e.g., only to high-infection regions, they have also been asymmetric in the following sense: Implementing preventive quarantine in region j (to return to our framework) for people from region k (and also people returning to j from k) reduces the transmission of infections of region k on region j . However, region k does not necessarily need to implement the mirror strategy, i.e., to quarantine individuals travelling from j to k . Hence, the two transmission rates $u_{j,k}$ and $u_{k,j}$ can actually be targeted distinctly.

Trying to reduce transmission rates to zero would lead to a complete halt of all economic behaviour, and the cost would then amount to the full GDP produced over this time period. The quadratic loss function then describes the increasing difficulty of transferring economic activity into contexts (e.g., home office) with lower risks of disease transmission and of closing down traffic across regions.

$$C_U(u, t, i, j) = \left(1 - \frac{u}{\beta_{i,j}}\right)^2 \cdot \overline{C_u^{i,j}} \quad (3.19)$$

²¹World Bank data shows that income within western and northern Europe mostly lies within 40000-65000\$ per capita. The statistical value of life is chosen to be 20 times the GDP per capita (per year) as in Alvarez et al. (2021).

The parameter $\overline{C_u^{i,j}}$ measures the GDP corresponding to activities that generate transmission of the virus from region i to region j . The GDP of all regions is normalized in units per day and person. Again we normalize the total GDP per day and person, which is the sum over all three regions, to 1, i.e., $\sum_{i,j} \overline{C_u^{i,j}} = 1$. Each region is assumed to produce one third of the total GDP (corresponding to its population size, see assumption (G1)), where 50% of each regional share results from interactions within the home region and 25% from interactions with each of the two other regions, i.e.,

$$\overline{C_u^{i,j}} = \begin{cases} N_i(0) \cdot 0.5 & \text{for } i = j \\ N_i(0) \cdot 0.25 & \text{for } i \neq j. \end{cases} \quad (3.20)$$

The region-specific uncontrolled transmission rate is defined as

$$\beta_{i,j} = \begin{cases} 0.5 \cdot \mathcal{I}_0 \cdot \alpha_L & \text{if } i = j \\ 0.25 \cdot \mathcal{I}_0 \cdot \alpha_L & \text{if } i \neq j. \end{cases} \quad (3.21)$$

Assuming that each light case (that does not escalate to a heavy case) infects an average number \mathcal{I}_0 of people, we can define the uncontrolled transmission rates by dividing through by the average time in the infectious stage $1/\alpha_L$. Furthermore, transmissions have to be assigned to contacts with different groups. As we do for the economic costs, we assume that 50% of contacts take place within each region and 25% across the boundary with each of the two neighbouring regions.²² A complete lockdown and elimination of all interaction is not plausible due to necessary activity, e.g., in the medical sector or in the provision of essential goods and services. Thus, we assume that all transmission rates have a lower bound $\underline{u}_{i,j}$ which we set to 10% of the uncontrolled rates in our illustrations, i.e., $\underline{u}_{i,j} = 0.1 \cdot \beta_{i,j}$.

The average time spent with light or no symptoms while being infectious is assumed to be 15 days for both undetected and detected cases, such that $\alpha_L = \alpha_D = 1/15$. We can infer the probability of escalation of a light case, i.e., the risk of hospitalisation, from the general infection fatality rate (IFR) and the mortality risk of patients in hospital treatment. Assuming an IFR of 0.0065 and a mortality risk of 15% for hospitalized cases, we obtain an escalation probability of $0.0065/0.15 = 0.043$. Spreading this probability over the average dwell time in the light case state, we obtain a per-diem escalation rate of $\theta_{LH} = 1 - (1 - 0.043)^{\alpha_L} = 0.002948$ (and an analogous definition for θ_{DH}).²³

Finally we need to specify the cost functions for hospital treatments and for tests, respectively, which we assume both to be linear and identical for all regions,

$$C_M(H_j, t, j) = \overline{C_M^j} \cdot H_j$$

$$C_V(v_j, t, j) = \overline{C_V^j} \cdot v_j.$$

Since we normalised the total population to 1, the cost parameters $\overline{C_M^j}$ and $\overline{C_V^j}$ correspond to the treatment costs per day and person, as measured in units of per capita GDP per day. Heavy cases in our framework are not only ICU-patients, but also patients admitted to the hospital with less intense treatment needs. Recent studies have found a wide range of costs per patient per day receiving ICU and general treatment. For our analysis we take the values from Edeka et al. (2021) who found costs of 190\$ (general ward) and 1330\$ (ICU) per patient day. According to the US Center for Disease Control (CDC), about one third of hospitalized cases require intensive care. Consequently, we take the expected costs for a heavy case to be 443\$ per patient day. The costs of testing vary strongly with the type of

²²We are aware that we are using an heuristic parameter \mathcal{I}_0 for the calibration of the uncontrolled transmission rates instead of the basic reproduction number \mathcal{R}_0 . We chose this strategy for two reasons. First, the derivation of \mathcal{R}_0 in a multigroup extended SIR-framework is quite involved (see Van den Driessche and Watmough (2002) for details). Second, actual estimates of \mathcal{R}_0 for Covid-19 cover a relatively broad range, so the \mathcal{R}_0 in our model resulting from the heuristic definition of the transmission rates is likely to be within this range. At the same time, the calibration of $\beta_{j,k}$ is significantly easier.

²³We adopted these parameter values from Caulkins et al. (2020) and Caulkins et al. (2021).

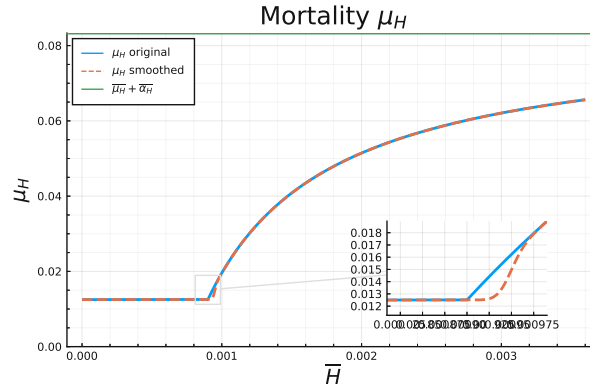


Figure 3.2: Original and final smoothed version of the mortality function describing congestion in the health sector.

test. However, generally the price is relatively low compared to all other costs (we set it to 5\$). It is thus capacity that acts as a potentially limiting factor.

ICU capacity is assumed to be fixed at 30 beds per 100.000 individuals.²⁴ Testing capacity is very low at the beginning of a pandemic, but increases over time. In our case, we assume that testing capacity increases from 10 up to 1000 tests per 100.000 individuals per day.

$$\bar{V}(t) = \bar{N} \cdot \left(0.0001 + 0.01 \cdot \frac{t}{\bar{T}} \right)$$

The mortality and regeneration rates of hospitalized individuals are assumed to depend on the number of heavy cases. We assume that the sum of the two rates is constant and describes an average hospitalisation time of 12 days. For an ICU with spare capacity the share of hospitalised people dying is assumed to be 45%. Translated to the total number of hospitalised patients, the base mortality is 15%, and thus related to the average time spent in the hospital according to $\overline{\mu_{H,j}} = 0.15/12$ and $\overline{\alpha_{H,j}} = 0.85/12$. As soon as the ICU-capacity is exhausted (the expected number of ICU patients is $\bar{H}/3$), we assume that any further patient with intensive care needs has a 100% mortality risk. Following the calculations of Caulkins et al. (2020), this implies that the average mortality for all heavy cases is increasing concave with the number of heavy cases approaching $1/12$ in the limit. For computational reasons we propose the following functional forms, which smooth the non-differentiable point of $\mu_H(\bar{H}, t, j)$, where $\bar{H}/3$ reaches the ICU cap. Smoothing ranges from this point up to a 10% overload of the ICU capacities:

$$\mu_H(\bar{H}, t, j) = \begin{cases} \overline{\mu_{H,j}} & \text{if } \frac{\bar{H}}{3} < ICU \\ \overline{\mu_{H,j}} + \overline{\alpha_{H,j}} \cdot \left(1 - \frac{ICU}{\bar{H}/3} \right) \cdot f(\bar{H}) & \text{if } ICU \leq \frac{\bar{H}}{3} \leq 1.1 \cdot ICU \\ \overline{\mu_{H,j}} + \overline{\alpha_{H,j}} \cdot \left(1 - \frac{ICU}{\bar{H}/3} \right) & \text{if } 1.1 \cdot ICU \leq \frac{\bar{H}}{3} \end{cases}$$

$$\alpha_H(\bar{H}, t, j) = \overline{\mu_{H,j}} + \overline{\alpha_{H,j}} - \mu_H(\bar{H}, t, j).$$

The function $f(\cdot)$ is a sigmoid function of the form $f(x) = \frac{x^3}{x^3 + (1-x)^3}$ which is sufficient for a smooth final function. Figure 3.2 shows the general form of μ_H and highlights the marginal differences between the original function (following Caulkins et al. (2020)) and our smoothed version.

Finally, we assume that the three regions differ (only) in the number of initial infections, reflecting the realistic setting in which the pandemic breaks out and spreads within a single region of origin before spreading into other regions. We thus propose that the pandemic has already progressed within the first region (group), where 10% of the population are already infected at the starting time of the model.

²⁴Countries that are equipped approximately with that amount of ICU capacity are Germany 33.9 (year 2017), Austria 28.9 (year 2018) and US 25.8 (year 2018). See [de.statista.com](https://www.de.statista.com) for details.

For the second group, we assume an initial prevalence of infections at 1% of the population, while the third region enters with a zero prevalence at the beginning of the time horizon.

Economic Parameters		Epidemiological Parameters	
n	3	\mathcal{I}_0	2.5
ρ	0.0	α_L	1/15
\bar{N}	$\sum_i N_i(0) = 1.0$	α_D	1/15
GDP	$50000[\$/c]/365[d] = 137.0[\$/c/d]$	θ_{LH}	$1 - (1 - 0.0065/0.15)^{\alpha_L}$ = 0.002948
Ψ	$10^6[\$/GDP] = 7300[GDP]$	θ_{DH}	$1 - (1 - 0.0065/0.15)^{\alpha_D}$ = 0.002948
ICU - Cap	$\bar{N} \cdot 0.0003$	$\bar{\mu}_{H,j}$	0.15/12 = 0.0125
\bar{C}_M^i	$(2/3 \cdot 190\$ + 1/3 \cdot 1330\$)/GDP$ = 4.161[GDP]	$\bar{\alpha}_{H,j}$	0.85/12 = 0.07083
\bar{C}_V^i	$5\$/GDP = 0.0365[GDP]$		

Table 3.2: Parameters for all numerical scenarios.

3.5.2 Results

In following subsections we study four cases, which represent a sequence of increasing capabilities of the social planner: (i) "Uncontrolled", i.e., the development of the epidemiological states if no intervening measures are taken. (ii) "No Testing", where the social planner is (only) able to introduce lockdown measures and social distancing orders. (iii) "Ineffective testing", where the social planner also allocates tests (subject to the capacity constraint) for the identification of light cases, but where in the absence of effective contact tracing, tests have to be allocated randomly across the non-diagnosed population, i.e., where $\kappa = 1$. (iv) "Perfect testing", as the optimal solution for a scenario where effective contact tracing allows tests to be (almost) perfectly targeted at light cases, i.e., where $\kappa = 0.1$. Table 3.3 shortly summarises the set-up for the different scenarios analysed.

Case	Transmissions	Testing	κ
Uncontrolled	uncontrolled	none	-
No Testing	optimally controlled	none	-
Ineffective Testing	optimally controlled	optimal	1
Perfect Testing	optimally controlled	optimal	0.1

Table 3.3: Qualitative comparison of the four different cases analysed

Figures 3.4, 3.5, 3.6 and 3.7 will allow us to compare the development of the pandemic, the optimal target transmission rates, the testing allocations and the composition of the costs across the "no testing", "ineffective testing" and "perfect testing" scenarios. Before, we discuss the uncontrolled development of the pandemic as a benchmark.

Uncontrolled development

Figure 3.3 presents the most important epidemiological states, as described by equations (3.1)–(3.4) for an uncontrolled pandemic: susceptibles, light cases, and heavy cases. The blue solid line denotes the corresponding state for region 1, the red one that for region 2 and the yellow one that for region 3. The

dashed black line denotes the total number of light and heavy cases, respectively. In the absence of any controls the different initial conditions across the regions make hardly any difference for the course and outcome of the pandemic. The spike in light and heavy cases is only delayed by a few days for the second and third region. Furthermore, an equilibrium is reached after 150 days where there are practically no susceptibles nor light cases left which could trigger new infections (recall that diagnosed and heavy cases are assumed to be isolated). Hence, the pandemic is short but severe, as follows from the fact that ICU capacity, denoted by the dotted green line in the panel is exceeded by the total number of heavy cases (dashed black line) over almost the full course of the pandemic. This causes a high number of deaths, which, as we will see, could have been prevented by measures aimed at controlling the disease.

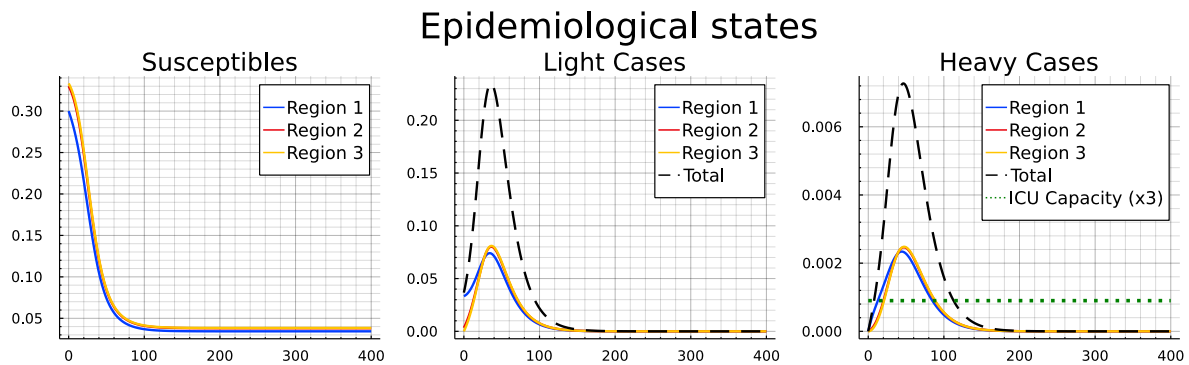


Figure 3.3: Pandemic development for the "Uncontrolled" case.

Table 3.4 includes a summary of the terminal outcomes of the pandemic after 400 days for each region as well as for all regions in total (as percentages of each group resp. the total initial population size).²⁵ The first four columns show the percentage of people, who have not been infected (S), recovered from the disease (R_L and R_D), and died (stage M), respectively. The final four columns show (from the left) the cumulated costs of the pandemic up until day 100, 200, 300 and 400 (i.e., end time), respectively.²⁶ As we know from Figure 3.3 the pandemic is raging mainly over the first 100 days, and has "burnt out" by day 200. Thus, costs do not change afterwards. The Table also mirrors that there is little difference across regions. Without any control measures the disease can quickly spill over from region 1 to region 2 and 3 and take the same course afterwards.

Interestingly, in region 1 slightly fewer people die as compared to regions 2 and 3, which is partly due to a first-mover advantage in the ICU. The fact that the early-coming heavy cases from region 1 are hospitalized in a situation without congestion yet gives them the advantage that all (or many) of them get the required critical care. A short time later, the ICU is congested and the available capacity is divided across the three regions, leaving many unserved in all regions.²⁷

Turning to infections, we see that in each region only slightly more than 10% of the population remain susceptible and more than 85% are either recovered light or heavy cases. Nearly 3% of the population have died during the pandemic, predominantly as a result of an overwhelmed health sector (recall the right panel in figure 3.3).

²⁵An extended version of this table can be found in Appendix 3.7.3.

²⁶A decomposition of the costs with respect to medical costs, lost lives can be found in Appendix 3.7.3.

²⁷We have verified the first-mover advantage by an additional numerical run without ICU capacity, where it turns out that the number of deaths is highest in region 1.

Uncontr. Case	Epidemiological states at $t = 400$				Costs until t (in GDPPCPD)			
	S	R_L	R_D	M	t=100	t=200	t=300	t=400
Region 1	10.29%	85.91%	0.87%	2.93%	70.23	71.91	71.92	71.92
Region 2	11.39%	84.86%	0.81%	2.94%	70.31	72.23	72.24	72.24
Region 3	11.51%	84.74%	0.8%	2.95%	70.32	72.27	72.28	72.28
Total	11.06%	85.17%	0.83%	2.94%	210.86	216.41	216.44	216.44
No Testing	Epidemiological states at $t = 400$				Costs until t (in GDPPCPD)			
	S	R_L	R_D	M	t=100	t=200	t=300	t=400
Region 1	27.58%	69.29%	2.6%	0.46%	19.92	27.22	30.26	31.52
Region 2	40.03%	57.34%	2.14%	0.38%	7.18	15.74	20.6	22.58
Region 3	41.5%	55.93%	2.09%	0.37%	5.34	13.69	18.76	20.83
Total	36.37%	60.85%	2.28%	0.41%	32.45	56.65	69.63	74.93
Ineffect. Testing	Epidemiological states at $t = 400$				Costs until t (in GDPPCPD)			
	S	R_L	R_D	M	t=100	t=200	t=300	t=400
Region 1	31.31%	63.99%	4.22%	0.44%	19.25	25.39	28.17	29.12
Region 2	44.14%	48.92%	6.53%	0.36%	7.06	14.92	18.93	20.3
Region 3	45.57%	47.36%	6.66%	0.35%	5.35	13.17	17.32	18.74
Total	40.34%	53.42%	5.80%	0.38%	31.66	53.49	64.42	68.16
Perfect Testing	Epidemiological states at $t = 400$				Costs until t (in GDPPCPD)			
	S	R_L	R_D	M	t=100	t=200	t=300	t=400
Region 1	54.02%	35.62%	10.07%	0.29%	15.07	18.25	18.8	18.82
Region 2	68.06%	21.8%	9.94%	0.2%	6.18	10.12	10.81	10.84
Region 3	69.24%	20.98%	9.59%	0.2%	5.31	9.32	10.02	10.05
Total	63.77%	26.13%	9.86%	0.23%	26.56	37.69	39.63	39.7

Table 3.4: Summary and comparison of the endstates of the pandemic and the costs of the pandemic across the three different regions for the four scenarios analysed. Endstates are given in percentage of the initial population size of each region, resp. the total population. Costs are given in units of GDP per capita per day (GDPPCPD).

Controlled development

In this section we compare the development across a sequence of three scenarios in which the social planner is assumed to have increasing capabilities of controlling the disease. Figure 3.4 allows for comparison of the epidemiological development of the pandemic over time between the no-testing, ineffective-testing and perfect-testing case.²⁸ The first column illustrates the development for each of the regions (colored lines) and in total (black dashed line) when tests are not available. The following three conclusions are evident. Compared to the uncontrolled development, first, the duration of the pandemic is much longer and basically stretches until a vaccine becomes universally available at the end of the time horizon. Second, the main initial focus of the measures lies on pushing down infections in region 1 to a level at which the number of light cases is as low as in the other regions. After that (i.e., after ≈ 130 days) infections in region 1 start to lie below those in the other regions, which is due to the smaller number of susceptibles, following the initial period. As one would expect, heavy cases follow the course of the light cases with a delay. Third, the ICU capacity constraint is surpassed only slightly²⁹, but again for a long time (between days 20 – 270). These features correspond to "classical" curve-flattening as an optimal policy to contain the number of deaths by avoiding excessive ICU congestion while at the same time containing the costs of shutting down the economy. Such a policy remains in place up until either herd immunity is reached (which lies well beyond the time horizon for the COVID-19 like disease underlying our calibration) or up until the vaccine is universally available. The second part of Table 3.4 summarizes the outcomes for this scenario after 400 days. It can be seen that the impact of the pandemic on fatalities is significantly reduced. The death rate drops from nearly 3% of the population to values between 0.37% and 0.46%. Furthermore, we see that the direct lockdown measures lead to a decrease in the number of infected individuals in the first region, while in the second and third one the spreading of the disease is significantly slowed down. After the end of the pandemic, around 36% of the total population across all regions have not been infected.

By slowing down transmissions and thereby reducing the number of light cases, the lockdown measures bring down the number of heavy cases from around 3.8% of the population to around 2.7%.³⁰ As it turns out, this is tantamount to containing the number of heavy cases to a level below or slightly above the capacity limit. This confirms that the optimal strategy for lockdown measures is to keep the number of heavy cases at a level, that does not overburden the ICU facilities, but at the same time contains the (economic) costs of lockdown.

We can now use the second column of Figure 3.4 to identify the impact of the availability of tests, which are subject to a capacity constraint. For this scenario we assume there is no infrastructure for contact tracing and the disease hardly shows any specific symptoms for light developments of the infection. Therefore the social planner is not able to target tests and consequently tests need to be allocated randomly across all members of the population without a diagnosis. The main epidemiological development in the case without testing looks qualitatively very similar. However the number of light cases continuously decreases between days 20 and 250 instead of reaching a plateau over a long time. While this mostly corresponds to the number of diagnosed cases turning positive, the number of never-infected individuals (i.e., the susceptibles at $t = 400$) is also slightly higher in all three regions. Furthermore the ICU capacities are used to the full extent about 20 days less as the total number of heavy cases drops slightly earlier.

Finally, we present the optimally controlled epidemiological development when testing is "nearly" perfect in the sense of being targeted mainly at the group of light cases, a setting that is broadly consistent with highly effective contact tracing. In this scenario we obtain significant differences to the

²⁸To keep the scales for each state variable the same for all scenarios, we decided to present the uncontrolled development of the pandemic separately, since the peak number of light and heavy cases are tenfold the number when any controls are available.

²⁹The fact that ICU capacity is surpassed at all might come as a surprise, however recalling the smoothed mortality function in Figure 3.2, it becomes apparent, that the first few cases that surpass the ICU capacity are subject to only a marginal increase in their mortality risk.

³⁰To obtain these figures add the (total) shares of R_D and M from Table 3.4, while noting that all heavy cases and only heavy cases are either diagnosed survivors or deceased.

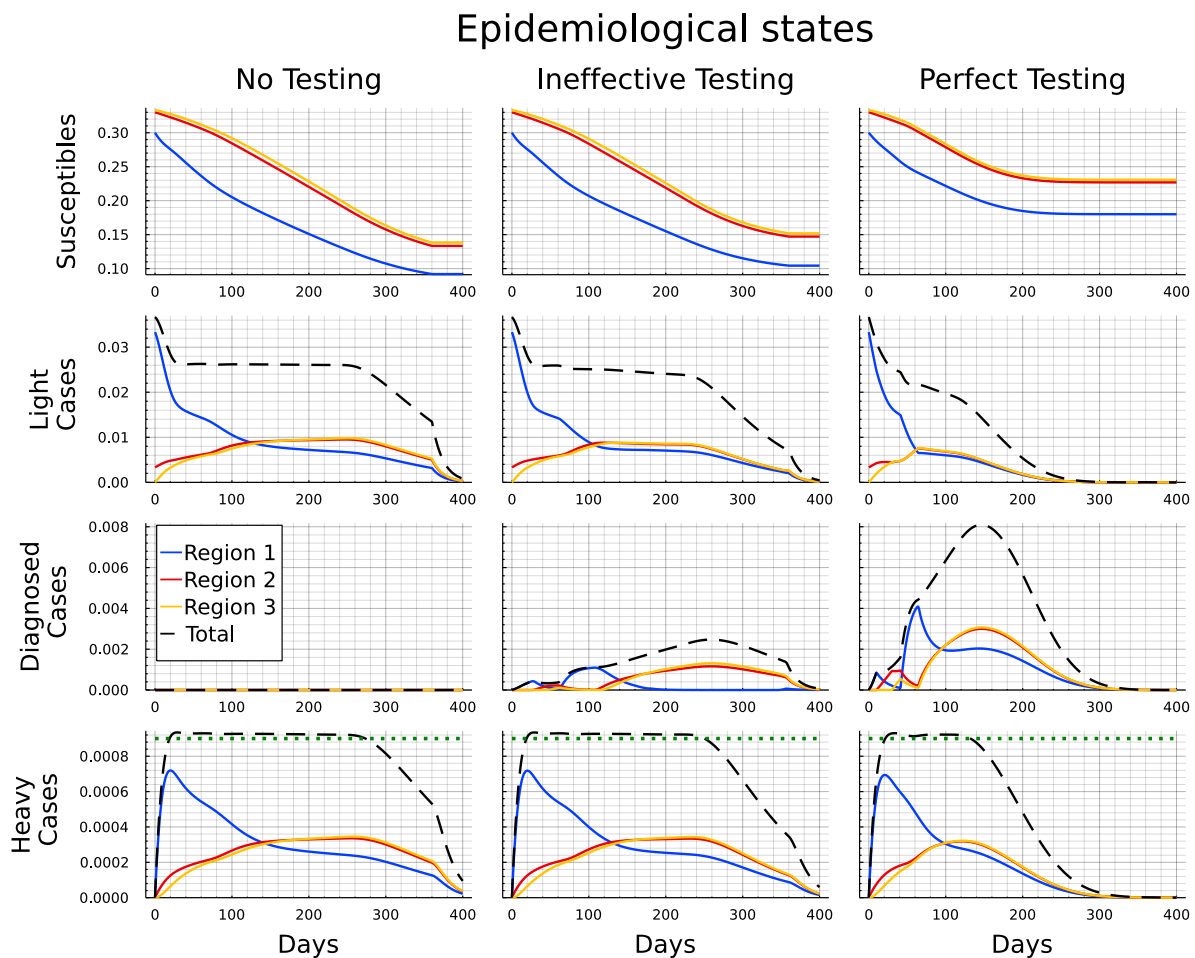


Figure 3.4: Comparison of the pandemic development for the "No testing", "Ineffective Testing" and "Perfect testing" case.

two previous cases of "no testing" and "ineffective testing", with the availability of testing now showing noticeable impacts.

In the third column in Figure 3.4 we observe that the pandemic ends after approximately 250 days and, thus, before the arrival of universal vaccination after 360 days. This is indicated by the number of light cases being close to zero and the number of susceptibles showing no significant change any more. During the first 60 days the development appears to not significantly differ from the development with ineffective testing³¹. Subsequently, however, we observe a steeper decline in the number of light cases in region 1. And although the light cases reach a similar peak in regions 2 and 3 regardless of the effectiveness of testing, in the case of perfect testing the numbers immediately drop from day 60 onwards, while they stay on a plateau for roughly 150 days if testing is ineffective.

The decreasing number of light cases in case of perfect testing also corresponds to the strongly increasing number of detected cases. The peak is earlier and nearly threefold the peak under ineffective testing. Note that the number of diagnosed individuals is higher throughout in the case of perfect testing even though the number of light cases is significantly lower. Hence the increased effectiveness clearly overcompensates the smaller pool of undetected light cases. The ICU capacities are also in full usage in case of perfect testing. However, as the infection dynamics have been slowed down, the number of heavy cases drop below this threshold significantly earlier compared to all other controlled cases.

Finally, note that after the apparent end of the pandemic after 250 days, with no further new infections,

³¹It appears that the low testing capacity is dominating the positive impact from higher testing effectiveness and the epidemiological development is only marginally different.

there are still active cases in the population. However they are mostly either heavy cases or diagnosed light cases who are isolated and cannot infect any remaining susceptible individuals. After all, the effective containment of the pandemic within the original hotspot allows to compress its overall duration (at much lower levels of infections) to only 250 days.

Comparing the end results of the epidemiological development after 400 days as shown in the second half of Table 3.4 we see that the impacts of testing strongly vary with respect to its effectiveness. For ineffective testing an additional 4% of the total population were protected from getting infected during the pandemic (increase from 36.37% to 40.34%). However the total number of deaths only dropped from 0.41% to 0.38%. In the case of nearly perfect testing, however, only slightly more than a third of the population got infected until a vaccine became available. Also the number of deaths was reduced to only 0.23% of the population and, thus, almost halved in comparison to the scenario without testing. This is largely owed to ICUs operating at capacity limit over a much shortened time span between roughly days 20 and 150.

Optimal transmission rates and tests

In this section we discuss in more detail the optimal strategies of the social planner for transmission rates and tests, which stand behind the previously discussed epidemiological development. In Figure 3.5 we present the optimal target transmission rates for the different scenarios. The upper boundary (black dashed line) corresponds to the transmission rates in the case of an uncontrolled pandemic. The red solid lines describe the optimal transmission rates if tests are not available, the blue dash-dotted lines for the case of ineffective testing and the green dotted lines for perfect testing. We first focus on the target transmission rates when testing is not an option for the social planner.

Let us begin with the within-region measures, corresponding to the panels along the diagonal from top-left to bottom right. Within region 1, the relatively high initial level of infections is curbed by the introduction of an initial short lockdown (a type of wave breaker). After some relaxation, measures are tightened again in order to control infections that now tend to be imported from regions 2 and 3. From day 100 onwards restrictions are gradually lifted until vaccination is available after 360 days. For region 2, the pattern of internal lockdown is qualitatively similar. Here, the initial tightening of the lockdown is in line with transmissions from imported cases. Notably, while the lockdown is softer than the one in region 1 initially, this reverses from day 50 onwards, where restrictions within region 2 are relaxed more cautiously and tend to remain in place over a longer period. The pattern of the internal lockdown within region 3 is similar to region 2 with the only difference being its more gradual introduction at the beginning of the planning horizon, following from the absence of internal infections at the starting point. Overall, the pattern suggests that the planner has an incentive (i) to equalize the pattern of the pandemic and its control across the three regions (which follows from the minimization of convex control costs), and (ii) to smooth out the lockdown over time, following its initial introduction. While this is relatively successful within the "follower" regions 2 and 3, the initial presence of the disease in region 1 makes this objective more difficult to achieve within this region.

Turning to cross-region measures, an initial observation is that for all regions the restrictions are considerably stronger than the internal measures. This suggests that the planner is much more ready to impose travel bans, which is particular true in respect to region 1, where transmissions into the other two regions are initially reduced to their lowest possible level, implying the strongest feasible measures. Similar yet more modest restrictions are implemented to avoid the spillover of transmissions from region 2 into region 3, which initially features no infections.

Conversely, and assuming the possibility of asymmetric restrictions, measures taken at constraining transmissions from regions 2 and 3 into region 1 and from region 3 into region 2, respectively, are more modest and implemented only gradually in line with the scope for "reimporting" infections. This asymmetric behaviour corresponds with the implementation of a mandatory quarantine (or similar measures to ensure non infectiousness) for individuals of region 1 trying to travel to regions 2 or 3, while travel in the reverse direction can be done "without" precautionary measures. The same implementation



Figure 3.5: Optimal target transmission rates for the "No testing", "Ineffective Testing" and "Perfect testing" case.

of mandatory quarantine would hold for individuals returning to regions 2 or 3 after having travelled to region 1.

Referring back to Proposition 3 this asymmetric behaviour becomes less surprising. Considering a special case of Proposition 3 with $k_1 = j_1 = k$ and $k_2 = j_2 = j$ and a small rearrangement of the terms leads to the following equation:

$$\frac{\partial C_U(u_{k,k}, t, k, k)}{\partial u_{k,k}} \cdot \frac{\partial C_U(u_{j,j}, t, j, j)}{\partial u_{j,j}} = \frac{\partial C_U(u_{k,j}, t, k, j)}{\partial u_{k,j}} \cdot \frac{\partial C_U(u_{j,k}, t, j, k)}{\partial u_{j,k}}. \quad (3.22)$$

Equation (3.22) connects all potential interactions between two nodes and illustrates that the product of marginal costs of transmission reduction *within* the two nodes has to be equal the product of marginal costs *between* the two nodes. In our numerical example, given the similar strategies within each region and significantly stronger transmission reductions from region 1 on regions 2 and 3, equation (3.22) directly implies reduced measures regarding transmissions in the opposite direction.

Given the substantially different target transmission rates from region 1 to region 2 (resp. from 1 to 3) and vice versa, questions about the extent, at which asymmetric measures improve the efficiency of disease control, might arise. To explore the implication of this aspect we additionally looked at a setting with symmetric target transmission rates, i.e., where the planner is constrained to set the same transmission targets $u_{j,j}$ within regions $j = 1, 2, 3$ and $u_{j,k} = u_{k,j}$ for all k and j across regions. The results, which are not presented here, indicate that (i) again the measures are chosen in a flattening-the-curve fashion, leading (ii) to a very similar level and distribution of mortality (and, thus, loss of lives). However comparing costs (which we will discuss for the four main scenarios in more detail in

Section 3.5.2) reveals the benefits of asymmetric measures: Economic costs of lockdowns are about 14% higher under the restriction of symmetric measures, what also leads to an increase in total cost by about 9%.

The introduction of ineffective testing now enables the social planner to modestly reduce the lockdown measures and allow slightly higher transmission rates at certain points in time (blue dash-dotted lines). While the optimal target transmission profile within region 1 and the transmissions of regions 2 and 3 do not appear to be significantly different, we can observe slightly reduced lockdown measures within regions 2 and 3 especially later on, when testing capacities are higher compared to the beginning. Furthermore the social planner allows for higher transmission rates between regions 2 and 3 (in both directions) as well as from region 1 to regions 2 and 3.

To assess the potential impact of the testing capacity we conducted a robustness check allowing for the maximum capacity of 1,000 tests per 100,000 individuals right from the start. The results show that the additional tests in the beginning are used exclusively in the first group and allow for a faster reduction of unidentified light cases. This "advance" against the disease is mainly used to relax lockdown measures and thereby to reduce the economic costs over the whole time-horizon by about a quarter. The epidemiological development remains largely unaffected, and there is hardly any change in the usage of ICU capacity or total lives lost.

The impact of testing becomes remarkably more pronounced if testing becomes more effective (green dotted line). While the patterns are qualitative similar to the "ineffective testing" case the measures are more lenient and confined to a shorter time period. After 100 days all transmission rates within the network are significantly higher and in particular, there is no need for policy interventions any more after 250 days, as is evident from the lack of any significant difference between the target and the uncontrolled transmission rates.

Interestingly, however, between days 40 and 80 we can identify increased efforts of protecting individuals in region 1 (from transmissions within group 1 and from the other two regions) compared to the other two controlled scenarios. Turning back to Figure 3.4 these efforts coincide with the peak in the light case numbers in region 2 and 3 and furthermore also a change in the test allocation as we will discuss further down below. Overall we can still conclude that despite the introduction of either ineffective or perfect testing, a "flattening-the-curve policy" continues to be the preferred choice for most of the pandemic.

Figure 3.6 illustrates the optimal allocation of the available tests at each point in time between the three regions. The left panel shows the strategy for ineffective testing, the right for perfect testing. First of all we can see that the testing capacities are used to their full extent (i.e., (3.9) is binding for

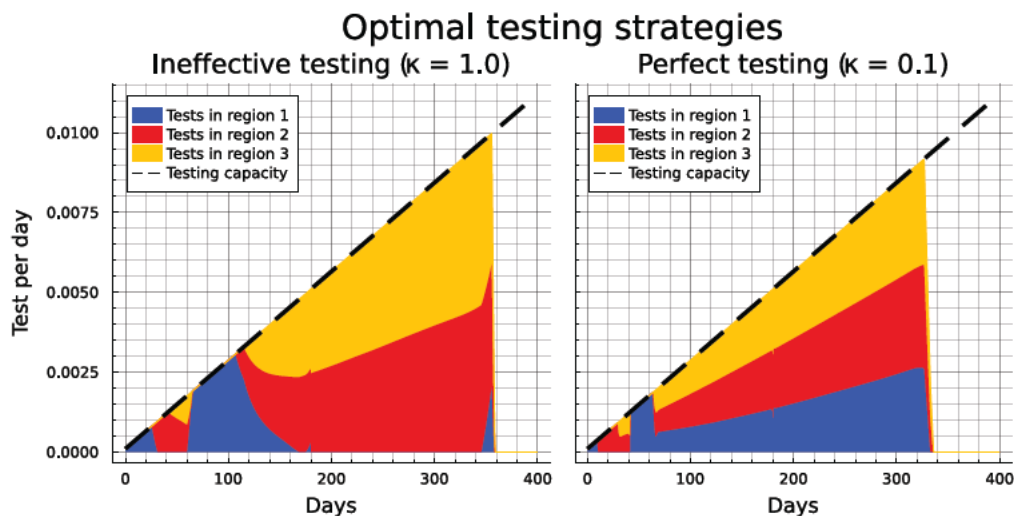


Figure 3.6: Optimal testing strategy in the "ineffective" and "perfect" testing case.

all t) at almost all times, since testing is relatively cheap, especially compared to all other measures and potential costs. Second, Figure 3.6 shows that a certain pattern regarding the allocation between the two regions is actually present for both levels of testing effectiveness. Testing starts exclusively in region 1 for a little more than 20 days (about a week for perfect testing, respectively), which is intuitive, as it is this region which is the hot spot of the pandemic initially. The focus subsequently shifts to region 2, followed by a mixed allocation across regions 2 and 3, before between days 75 and 100 (respectively days 60 and 80 for perfect testing) testing effort is again concentrated on region 1. This circular pattern is also visible in the pattern of diagnosed cases (see Figure 3.4) and broken only from day 100 (respectively day 80 for perfect testing) onward, where the number of infected cases has about equalised across the three regions. Despite the qualitative similarities, highly effective testing results in the shortening of the time span of each part of the pattern.

Subsequently, the patterns start to differ between ineffective and perfect testing. In the ineffective case testing is gradually shifted away from region 1 and shared between regions 2 and 3 which feature a higher number of light cases (due to the later epidemiological peak in these regions). Between days 180 and 340 tests are shared to roughly equal amounts between regions 2 and 3, while there is a small spike in tests in region 1 shortly before a vaccine becomes available at the end of the time horizon. If testing is highly effective, after the brief period of focused region 1 testing, there is an immediate switch towards an equal distribution of tests between all three regions over the remaining duration of the pandemic (compared to the gradual shift in the ineffective case). Especially interesting, however, is the complete lack of testing after day 340, i.e., 20 days before a vaccine becomes available. This elimination of any testing effort happens abruptly and despite the comparatively very low costs. Hence we can deduce, that effective testing and optimal lockdown measures allow the social planner to reduce the infections to such a high degree, that neither testing nor significant reductions in transmission rates are necessary to keep the pandemic under control for the last 3 weeks leading up to the availability of a vaccine.³² Lastly recall, that in the case of perfect testing after day 250 the infection numbers (especially the number of undetected light cases) remain on very low level. Hence the positive testing efforts between days 250 and 340 reflect a strategy where highly effective tests, if available at relatively low cost, are used in a precautionary way even after the apparent end of the pandemic in order to suppress any potential resurgence.

Cost composition

In a last step we assess impacts of the different scenarios on the objective function, i.e., the aggregated social costs of the pandemic. First we highlight, that the costs arising under the uncontrolled development of the pandemic are significantly higher than in any of the controlled scenarios. The right part of Table 3.4 summarises the costs arising up to different points in time (100, 200, 300 and 400 days respectively) and how the costs are distributed across the different regions. In the uncontrolled case the costs are nearly equally distributed between the three regions, but they almost completely arise during the first 100 days of the pandemic. This corresponds to the time frame with the number of heavy cases drastically overburdening the ICU capacities leading to a high number of fatalities. As there are no lockdown measures or testing efforts, all costs result from either hospital treatment expenditures or value of lives lost. Thereby the treatment costs play only a marginal role at less than 1% of total costs.³³ To contextualize the total costs of 216 units of GDP per capita per day, recall that the population size is normalized to one. Hence letting the pandemic spread uncontrolled through the population implies the value of lives lost during the first 100 days equals roughly two thirds of the total yearly GDP of the economy.

³²This certainly is an anticipatory effect of the finite time horizon. Without reaching the herd immunity level of the disease (the optimal outcome of our controlled regimes for the relatively short time horizon), a resurgence of infections is inevitable, if the target transmission rates return to their uncontrolled levels. However just comparing the outcomes of the two testing regime shows, that perfect testing allows for improved management of the pandemic to a degree, where infection numbers are so low, that even highly effective testing at a low price is not worth applying for the last 20 days.

³³For details see Table 3.6.

The remaining results in Table 3.4 show drastic improvements regarding the total health and economic costs of the pandemic. Allowing for reductions in the transmission rates brings the costs down by some 65% to 74.93 GDP per capita per day. Another 3% cost reduction compared to the uncontrolled case is possible through the introduction of ineffective testing. Similar to all previous results more significant improvements can be achieved if testing is effective. The total costs of 39.7 GDP per capita per day are only slightly above half the costs of the "no testing" case and represent a mere 18% of the costs resulting from the uncontrolled development.

The distribution of costs across time and the regions has also changed in relation to the uncontrolled case, but is qualitatively similar for all controlled scenarios. Region 1, the original hot spot of the pandemic, is responsible for a significantly higher share of costs compared to regions 2 and 3, which end up with more similar levels of aggregated costs. For all regions the costs are spread more smoothly over time, clearly reflecting the curve-flattening strategy.

A more distinct analysis of the composition of health and economic costs can be deduced from Figure 3.7. This figure presents the costs profiles of the three controlled scenarios and allows for a comparison between them. Furthermore the costs are composed into their different sources and attributed to the regions, where they arise. Noting that the costs for testing (yellow) and medical treatments (green) are insignificant in all scenarios, we will focus on the two main driving factors: lockdown measures and value of lives lost. In all three controlled cases it is optimal to incur high

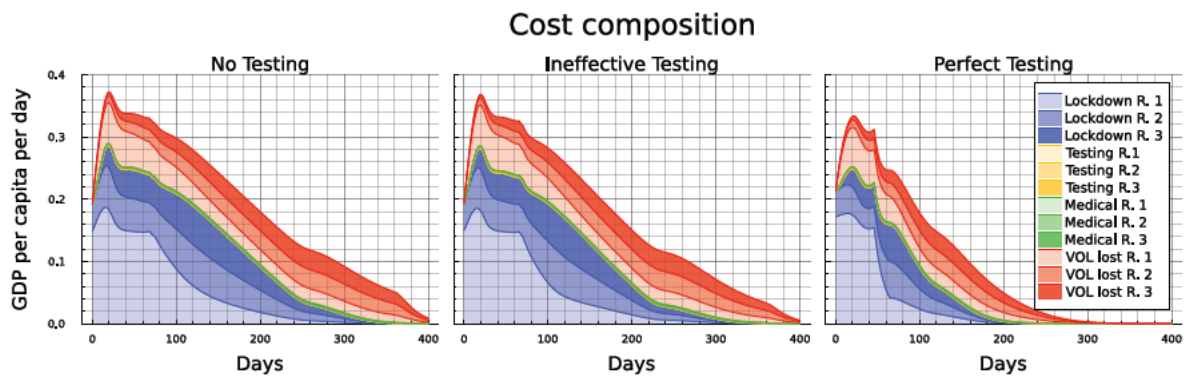


Figure 3.7: Origins of costs for the "No testing" (left), "Ineffective testing" (middle) and "Perfect testing" case (right).

lockdown costs to reduce the number of fatalities (and therefore the losses in value of life) especially within the first 40 days. The most significant part of lockdown costs results from transmission reductions in relation to region 1. As the pandemic develops over time, the shares of lockdown measures and fatalities become more balanced. Towards the end of the pandemic, when transmission rates return close to their uncontrolled values, the value of lives lost constitute the majority of costs. While the shares of value of lives lost between the three regions over time shift from region 1 towards regions 2 and 3, the aggregated absolute value remains fairly constant (again reflecting the "flattening the curve"-strategy).

The availability of imperfect testing lowers the cost of lives lost by some 6%, the economic cost of lockdown by some 11%, and total costs by some 9% relative to the "No testing case", but affects only marginally the overall cost profile. Testing becoming highly effective not only leads to the cost of lost lives being cut by some 51% and the economic cost by some 44%, but also implies a significant shift in the cost profile. While the cost structure is fairly similar to the other two controlled cases within the first 60 days and also reaches a comparable peak, thereafter the costs drop substantially and the distribution of both cost types between the three regions is roughly equal at every point in time. Notably, while the economic costs associated with lockdown measures are eliminated after roughly 200 instead of 300 days, the costs associated with the loss of lives have also been reduced to a negligible level by day 300 in the "perfect testing" case. This is in contrast to the "ineffective testing" case, where deaths accrue up to

and even beyond day 360, where universal vaccination eliminates new infections but where a significant number of the infected still develop a heavy course and die.

Summary comparisons

We conclude the analysis by comparing the four (main) scenarios on the basis of a number of key outcomes, as summarized in Table 3.5.

		Uncontrolled	No Testing	Ineffective Testing	Perfect Testing
Endstates ($t=400$)	S	11.06%	36.37%	40.34%	63.77%
	R_L	85.17%	60.85%	53.42%	26.13%
	R_D	0.83%	2.28%	5.80%	9.86%
	M	2.94%	0.41%	0.38%	0.23%
Agg. Costs in GDPPCPD ($t=400$)	Total	216.44	74.93	68.16	39.7
	VOL lost	214.56	29.6	27.79	16.85
	Lockdown	0.0	44.0	39.04	22.03
End of spread (days)		161.0	360.0	360.0	286.0
End of full ICU-usage (days)		113.0	274.0	250.0	132.0
End of lockdown measures (days)		0.0	360.0	360.0	249.0

Table 3.5: Comparison of the 4 (main) scenarios.

The first part of the table allows for an immediate comparison across the pandemic endstates in terms of the (remaining) susceptible population, the diagnosed and undiagnosed recovered cases, as well as the deceased at time $t = 400$. The second part summarizes the total costs after $t = 400$ and the value of lost lives and lockdown costs as the two main contributing factors (aggregated over all regions). We leave these entries for ease of comparison but without further comment.

The last three rows record a number of important dates in the progress of the pandemic. "End of Spread" defines the point in time at which the total number of susceptibles changes by less than 1 in 10,000, which is tantamount to a standstill of the pandemic. Thereby, we recall that in the "Uncontrolled" case the pandemic is over very quickly after 161 days but at enormous costs in terms of lost lives. Meanwhile, both for the "No testing" and "Ineffective testing" cases, the end of the spread does not occur before a vaccine is available. In contrast, "perfect testing" combined with an optimal lockdown policy allows to shorten the pandemic by 74 days.

The second row shows the first day at which the ICU is no longer at (or beyond) its capacity limit. Again in the uncontrolled case this happens early on, as the disease rages heavily within the first 3 months. This hides, of course, the number of heavy cases not receiving treatment and dying due to overload. More notably, we see that testing, regardless of its effectiveness, allows for an earlier relief of the strain on the ICU. In the limit, "perfect testing" more than halves the time span of congestion, whereas in case of "ineffective testing" the reduction is more minor.

The last row shows the date at which all lockdown measures have been lifted, as defined by the time at which all target transmission rates are within 1% of the uncontrolled rates. Again "perfect testing" is highly effective in bringing down to 249 the number of days over which lockdown measures of some intensity are implemented. "Ineffective testing" leads to no reduction in the duration of the lockdown (of some intensity) below the full 360 days up until the availability of universal vaccination.

3.6 Conclusions

This paper considers an extended SIR model across a network. It combines several important new features that are relevant especially for the COVID-19 pandemic but generalize to other infectious diseases that involve severe patterns of illness. The model distinguishes light and heavy courses of the disease, where the latter are characterized by the need for hospital or ICU treatment. Whereas a heavy course is per definition known and excluded from infectious contacts, light cases need to be detected by testing efforts. We allow for parametric variations in the extent to which tests can be targeted effectively, e.g., due to high capacity of contact tracing. This influences the success rate of detecting people with a light course by testing efforts. Detected people suffering a light course are quarantined and also do not contribute further to new infections. We consider that ICU capacity is constrained. If the capacity constraint is exceeded people with a heavy course cannot be treated in the ICU and die.

We consider this model on a network, where the population at each node is subdivided in the relevant compartments. The network structure of the model can represent different geographical regions, an example we consider in the main body of the article, or different subgroups of the population, or, indeed, both. Disease transmission occurs both within and across network nodes. By the optimal choice of a set of target transmission rates (within certain bounds) the decision maker can reduce the spread of the disease. Depending on the model interpretation it is possible that the ICU capacity corresponds to a single node or is shared between (some of the) nodes.

In general we were able to derive, that for the optimal strategy the ratio of marginal costs of the target transmission rates from two different nodes/regions onto another one only depends on the current shares of infectious individuals in the two origin regions/nodes. Furthermore we found that the relative size of the testing pool (which then again is strongly affected by the testing effectiveness) is one decisive factor for the relative allocations of testing and lockdown measures.

From a numerical example in which the pandemic spreads from one region into two otherwise identical regions, we can summarize the following key messages.

1. The analysis demonstrates that lockdown is effective in terms of saving lives and reducing total costs. While this depends, of course, on our assumptions about the value of a life saved, the cost reduction is so large as to make our finding robust.
2. For our calibration, the optimal lockdown policy is geared towards "flattening-the-curve" and thereby containing the number of heavy cases to a level that ICU capacities are (just) able to cope without significant congestion.
3. Lockdown within regions is typically weaker than measures undertaken to contain the transmission of the pandemic across regions. Initially, strong measures are taken to contain the disease within the region of the initial outbreak but these are adjusted over time to establish a course of the disease that is symmetric across the three regions. The common practice followed by many countries during the pandemic of imposing travel bans provoked by spikes in infections in foreign countries finds some support in our model.
4. Testing, on which countries have relied to very different extent during the COVID-19 pandemic, proves to be effective by relaxing the trade-off between saving lives and containing economic costs. However, the extent to which this is true hinges on (i) the effectiveness of contact tracing as a necessary requirement for tests that are targeted at (likely) light infections; and (ii) on limits to the testing capacity. If tests are in limited supply (especially during the early stages of the pandemic) only well targeted testing has the potential to speed up the termination of the disease and, thereby, to cut the time over which lockdown measures are required. In policy terms, this requires the availability of effective tracing capacities. The massive cost reductions over the pandemic afforded by the introduction of effective (rather than ineffective) testing indicates that policy makers should be willing to invest significantly into improvements in testing efficacy.

5. Both lockdown and testing are initially concentrated on the hot spot(s) with high numbers of current infections with the twin aim of flattening-the-curve and containing spill-over effects into other regions. Policies are subsequently adjusted in a way that they become more similar over time.
6. When the policy objective is the minimization of the total cost of the pandemic across all regions alike, then both lockdown and testing policies are chosen in a way that aims at equalizing the pandemic across regions. While this leads to the assimilation of policies across regions, such an endeavour is the more successful the more targeted testing can be used.

Our model features a number of limitations, most notably that it is set out from a social planner perspective. This provides a characterization of the first-best allocation as a yardstick and - in its regional interpretation - reflects the policy that should be implemented by a strong centralized government. But it does not yet capture well the policy outcomes in a decentralized setting where (i) local decision-makers at each node of the network - in the regional context best thought of as regional authorities - optimize a local (regional) objective but not social welfare; and/or (ii) where individuals are taking voluntary decision in respect to self-protection. While the modelling in (ii) is difficult to integrate with the planner model, one natural extension to the present model would be the consideration of a game between local decision-makers. Comparison against the first-best allocation would allow to identify inefficiency related to externalities across network nodes (regions) as well as possible coordinating measures on the part of the central government aimed at improving the allocation. Finally, we believe our framework to be flexible enough to be applied to a broad range of pandemic, population and economic settings, including consideration of social rather than regional networks and a variety of of settings that involve asymmetric networks. We relegate these issues to future work.

3.7 Appendix

3.7.1 Optimality conditions

(3.15) is a finite time optimal control model, for which the standard form of the Maximum Principle can be applied (see e.g., Grass et al. (2008)). First we formulate the Hamiltonian, which is

$$\begin{aligned}
\mathcal{H}(\cdot) = & - \sum_{j \in \Omega} \left[C_M(H_j, t, j) + \mu_H(\overline{H}_j, j) H_j \cdot \Psi + C_V(v_j, t, j) \right] - \\
& - \sum_{(k,j) \in \Omega^2} C_U(u_{k,j}, t, k, j) + \sum_{j \in \Omega} \left[\lambda_{S_j} \left(-S_j \cdot \sum_{k \in \Omega} \left[u_{k,j} \cdot \frac{L_k}{S_k + L_k + R_k^L + R_k^D} \right] \right) \right] + \\
& + \sum_{j \in \Omega} \left[\lambda_{L_j} \left(S_j \cdot \sum_{k \in \Omega} \left[u_{k,j} \cdot \frac{L_k}{S_k + L_k + R_k^L + R_k^D} \right] - \theta_{LH}(j) L_j - \alpha_L(j) L_j - \right. \right. \\
& \quad \left. \left. - \frac{v_j}{L_j + \kappa(S_j + R_j^L)} L_j \right) \right] + \\
& + \sum_{j \in \Omega} \left[\lambda_{D_j} \left(\frac{v_j}{L_j + \kappa(S_j + R_j^L)} L_j - \alpha_D(j) D_j - \theta_{DH}(j) D_j \right) \right] + \\
& + \sum_{j \in \Omega} \left[\lambda_{H_j} \left(\theta_{LH}(j) L_j + \theta_{DH}(j) D_j - \alpha_H(\overline{H}_j, j) H_j - \mu_H(\overline{H}_j, j) H_j \right) \right] + \\
& + \sum_{j \in \Omega} \left[\lambda_{R_j^L} \cdot \alpha_L(j) L_j + \lambda_{R_j^D} \cdot \left(\alpha_D(j) D_j + \alpha_H(\overline{H}_j, j) H_j \right) + \lambda_{M_j} \cdot \mu_H(\overline{H}_j, j) H_j \right] +
\end{aligned}$$

$$\begin{aligned}
& + \sum_{j \in \Omega} \lambda_{\overline{H_j}} \left(\sum_{k \in \Omega} H_k f_H(j, k) \right) + \Lambda \left(\sum_{j \in \Omega} v_j - \overline{V} \right) + \\
& + \sum_{(k,j) \in \Omega^2} \left[\overline{\xi_{k,j}}(u_{k,j} - \beta_{k,j}) + \underline{\xi_{k,j}}(u_{k,j} - u_{k,j}) \right] + \sum_{j \in \Omega} \left[\overline{\nu_j}(v_j - \overline{v_j}) + \underline{\nu_j}(v_j - v_j) \right]. \quad (3.23)
\end{aligned}$$

λ_{x_j} denotes the adjoint variable of state variable $x_j \in \{S_j, L_j, D_j, H_j, R_j^L, R_j^D, M_j\}$ ($j \in \Omega$) and Λ denotes the Lagrangian multiplier of the control constraint (3.9). $\overline{\xi_{k,j}}, \underline{\xi_{k,j}}, \overline{\nu_j}, \underline{\nu_j}$ ($j, k \in \Omega$) are the Lagrange multipliers for the upper and lower boundaries of the controls. The lower bound of the transmission rates denoted by $\underline{u_{k,j}}$ equals zero in the full model (3.15).

The first order conditions for the controls read ($j, k \in \Omega$)

$$\frac{\partial \mathcal{H}(\cdot)}{\partial u_{k,j}} = -C'_U(k, j) + \frac{(\lambda_{L_j} - \lambda_{S_j}) S_j L_k}{S_k + L_k + R_k^L + R_k^D} + \overline{\xi_{k,j}} - \underline{\xi_{k,j}} = 0 \quad (3.24a)$$

$$\frac{\partial \mathcal{H}(\cdot)}{\partial v_j} = -C'_V(j) - (\lambda_{L_j} - \lambda_{D_j}) \frac{L_j}{L_j + \kappa(S_j + R_j^L)} + \Lambda + \overline{\nu_j} - \underline{\nu_j} = 0, \quad (3.24b)$$

with complementary slackness conditions

$$\Lambda \left(\sum_{j \in \Omega} v_j - \overline{V} \right) = 0 \quad (3.25a)$$

$$\overline{\xi_{k,j}}(u_{k,j} - \overline{u_{k,j}}) = 0 \quad , \quad (k, j) \in \Omega^2 \quad (3.25b)$$

$$\underline{\xi_{k,j}}(\overline{u_{k,j}} - u_{k,j}) = 0 \quad , \quad (k, j) \in \Omega^2 \quad (3.25c)$$

$$\overline{\nu_j}(v_j - \overline{v_j}) = 0 \quad , \quad j \in \Omega \quad (3.25d)$$

$$\underline{\nu_j}(v_j - v_j) = 0 \quad , \quad j \in \Omega. \quad (3.25e)$$

From the Hamiltonian the adjoint equations can straightforwardly be obtained ($j \in \Omega$)

$$\begin{aligned}
\dot{\lambda}_{S_j} &= \rho \lambda_{S_j} + (\lambda_{S_j} - \lambda_{L_j}) \sum_{k \in \Omega} u_{k,j} \frac{L_k}{S_k + L_k + R_k^L + R_k^D} - \\
& - \sum_{k \in \Omega} \frac{(\lambda_{S_k} - \lambda_{L_k}) S_k u_{j,k} L_j}{(S_j + L_j + R_j^L + R_j^D)^2} - \frac{\kappa(\lambda_{L_j} - \lambda_{D_j}) L_j v_j}{(L_j + \kappa(S_j + R_j^L))^2} \quad (3.26a)
\end{aligned}$$

$$\begin{aligned}
\dot{\lambda}_{L_j} &= \rho \lambda_{L_j} + \sum_{k \in \Omega} (\lambda_{S_k} - \lambda_{L_k}) S_k u_{j,k} \frac{S_j + R_j^L + R_j^D}{(S_j + L_j + R_j^L + R_j^D)^2} + \\
& + (\lambda_{L_j} - \lambda_{D_j}) \frac{\kappa v_j (S_j + R_j^L)}{(L_j + \kappa(S_j + R_j^L))^2} - (\lambda_{H_j} - \lambda_{L_j}) \theta_{LH}(j) - \\
& - (\lambda_{R_j^L} - \lambda_{L_j}) \alpha_L(j) \quad (3.26b)
\end{aligned}$$

$$\dot{\lambda}_{D_j} = \rho \lambda_{D_j} + (\lambda_{D_j} - \lambda_{H_j}) \theta_{DH}(j) + (\lambda_{D_j} - \lambda_{R_j^D}) \alpha_D(j) \quad (3.26c)$$

$$\begin{aligned}
\dot{\lambda}_{H_j} &= \rho \lambda_{H_j} + C'_M(j) + \mu_H \cdot \Psi - (\lambda_{R_j^D} - \lambda_{H_j}) \alpha_H - (\lambda_{M_j} - \lambda_{H_j}) \mu_H - \\
& - \sum_{k \in \Omega} \lambda_{\overline{H_k}} f_H(k, j) \quad (3.26d)
\end{aligned}$$

$$\begin{aligned} \dot{\lambda}_{R_j^L} &= \rho \lambda_{R_j^L} - \sum_{k \in \Omega} (\lambda_{S_k} - \lambda_{L_j}) S_k u_{j,k} \frac{L_j}{(S_j + L_j + R_j^L + R_j^D)^2} + \\ &\quad + (\lambda_{L_j} - \lambda_{D_j}) \frac{\kappa v_j L_j}{(L_j + \kappa(S_j + R_j^L))^2} \end{aligned} \quad (3.26e)$$

$$\dot{\lambda}_{R_j^D} = \rho \lambda_{R_j^D} - \sum_{k \in \Omega} (\lambda_{S_k} - \lambda_{L_j}) S_k u_{j,k} \frac{L_j}{(S_j + L_j + R_j^L + R_j^D)^2} \quad (3.26f)$$

$$\dot{\lambda}_{M_j} = \rho \lambda_{M_j} \quad (3.26g)$$

and

$$\lambda_{H_j} = -\frac{\partial \mu_H}{\partial H_j} H_j(t) \cdot \Psi + (\lambda_{R_j^D} - \lambda_{H_j}) \frac{\partial \alpha_H}{\partial H_j} H_j - (\lambda_{H_j} - \lambda_{M_j}) \frac{\partial \mu_H}{\partial H_j} H_j \quad (3.27)$$

with the following transversality conditions

$$\lambda_{S_j}(T) = 0 \quad (3.28a)$$

$$\lambda_{L_j}(T) = \frac{\partial \mathcal{S}(\mathcal{X}(T), T)}{\partial L_j(T)} \quad (3.28b)$$

$$\lambda_{D_j}(T) = \frac{\partial \mathcal{S}(\mathcal{X}(T), T)}{\partial D_j(T)} \quad (3.28c)$$

$$\lambda_{H_j}(T) = \frac{\partial \mathcal{S}(\mathcal{X}(T), T)}{\partial H_j(T)} \quad (3.28d)$$

$$\lambda_{R_j^L}(T) = 0 \quad (3.28e)$$

$$\lambda_{R_j^D}(T) = 0 \quad (3.28f)$$

$$\lambda_{M_j}(T) = 0. \quad (3.28g)$$

3.7.2 Interpretation of shadow prices

Solving for the shadow prices (i.e., the adjoint variables) allows us to identify the interconnected channels that determine the "values" (from an epidemiological perspective) of individuals within the different compartments. The shadow price of a susceptible at node j reads

$$\begin{aligned} \lambda_{S_j}(t) &= \int_t^T e^{-\rho(s-t)} \left(\underbrace{(\lambda_{L_j} - \lambda_{S_j}) \sum_{k \in \Omega} u_{k,j} \frac{L_k}{S_k + L_k + R_k^L + R_k^D}}_{=(i)} - \right. \\ &\quad \left. - \underbrace{\frac{L_j}{(S_j + L_j + R_j^L + R_j^D)^2} \sum_{k \in \Omega} u_{j,k} (\lambda_{L_k} - \lambda_{S_k}) S_k}_{=(ii)} - \right. \\ &\quad \left. - \underbrace{\frac{\kappa v_j}{(L_j + \kappa(S_j + R_j^L))^2} (\lambda_{D_j} - \lambda_{L_j}) L_j}_{=(iii)} \right) dt. \end{aligned} \quad (3.29)$$

The value of a susceptible, as given by λ_S , aggregates the discounted value of three effects. (i) captures the expected loss of an additional (marginal) susceptible at node j becoming infected and turning into

a light case. The expected loss consist of the vale of a susceptible getting infected and turning into a light case, $\lambda_{L_j} - \lambda_{S_j} < 0$,³⁴ weighted with the (aggregate) probability of getting infected through a contact from any node in the network. The other two effects capture second-order effects. Specifically, (ii) describes the expected value of a marginal susceptible in lowering the share of light cases in the population at node j and, thus, the aggregate risk of a light case at j infected a susceptible at any other node k . (iii) measures that an additional susceptible at node j makes it harder to identify a light case via testing. The change in probability of finding a light case in the relevant population via testing is then weighted by the value of a light case being detected, $\lambda_{D_j} - \lambda_{L_j} > 0$.³⁵

The shadow price of light cases at node j is structured similarly and reads

$$\begin{aligned} \lambda_{L_j}(t) = & \int_t^T e^{-\rho(s-t)} \left[(\lambda_{H_j} - \lambda_{L_j}) \theta_{LH}(j) + (\lambda_{R_j^L} - \lambda_{L_j}) \alpha_L(j) + \right. \\ & + \frac{S_j + R_j^L + R_j^D}{(S_j + L_j + R_j^L + R_j^D)^2} \sum_{k \in \Omega} (\lambda_{L_k} - \lambda_{S_k}) S_k u_{j,k} + \\ & \left. + \frac{\kappa v_j (S_j + R_j^L)}{(L_j + \kappa(S_j + R_j^L))^2} (\lambda_{D_j} - \lambda_{L_j}) \right] dt + e^{-\rho(T-t)} \frac{\partial \mathcal{S}(\mathcal{X}(T), T)}{\partial L_j(T)}. \end{aligned} \quad (3.30)$$

The first two terms in the integral are direct effects. They contain the value of a light case escalating to a heavy case (moving to the H_j compartment) and the value of a light case recovering from the infection (moving to the R_j^L compartment). Both terms are multiplied by their respective rate of occurrence. The third term captures the effect of light cases at node j on overall infections within node j and into other nodes. The fourth term captures the effect on the detection of light cases. The term outside the integral is the discounted effect of light cases on the costs that arise after the planning horizon.

Moving on to the shadow price of detected cases at node j , we can derive the following expression:

$$\begin{aligned} \lambda_{D_j}(t) = & \int_t^T e^{-\rho(s-t)} \left[(\lambda_{H_j} - \lambda_{D_j}) \theta_{DH}(j) + (\lambda_{R_j^D} - \lambda_{D_j}) \alpha_D(j) \right] dt + \\ & + e^{-\rho(T-t)} \frac{\partial \mathcal{S}(\mathcal{X}(T), T)}{\partial D_j(T)}. \end{aligned} \quad (3.31)$$

The two terms within the integral capture once again the values of a detected individual developing heavy symptoms or recovering from the disease (in this case moving to the R_j^D compartment, however). Again, the values of the two transitions are multiplied by their respective arrival rates. Note that for detected cases these may differ from the respective rates in case of light cases. Specifically, the scope for monitoring and early on treatment in case of detection will lead to a higher (lower) rate of recovery (disease escalation), implying in turn a higher value of a detected case. As detected cases are quarantined, there are no knock-on effects on the process of infection, nor on the success of testing. The term outside the integral again corresponds to the discounted effect on costs after the planning horizon.

The shadow price of heavy cases at node j reads

$$\begin{aligned} \lambda_{H_j}(t) = & \int_t^T e^{-\rho(s-t)} \left[\underbrace{-C'_M(j) - \mu_H \cdot \Psi}_{=(i)} + \underbrace{(\lambda_{R_j^D} - \lambda_{H_j}) \alpha_H + (\lambda_{M_j} - \lambda_{H_j}) \mu_H}_{=(ii)} + \right. \\ & \left. + \underbrace{\sum_{k \in \Omega} \lambda_{H_k} f_H(k, j)}_{=(iii)} \right] dt + e^{-\rho(T-t)} \frac{\partial \mathcal{S}(\mathcal{X}(T), T)}{\partial H_j(T)}. \end{aligned} \quad (3.32)$$

³⁴The negative sign is readily verified for an interior optimum from the FOC (3.16a).

³⁵The positive sign is readily verified for an interior optimum from the FOC (3.16b).

The integral aggregates the discounted value of three effects: (i) counts the costs of an additional heavy case at node j , consisting of medical treatment costs and the value of a lost life (weighted by the death rate). (ii) values the transitions to recovery (in this case into the $\lambda_{R_j^D}$ compartment) and death. Note that while (i) values mortality by the value of a lost life, (ii) focuses on the value change due to a heavy case moving to the M_j compartment. Note that these two effects typically differ in their sign. (iii) assigns a value to heavy cases at node j based on their utilisation of hospital that is no longer available at other nodes. The term outside the integral captures the discounted effect on the costs after the planning horizon.

The shadow prices of recovered cases (light or diagnosed) at node j

$$\lambda_{R_j^L}(t) = \int_t^T e^{-\rho(s-t)} \left(-\frac{L_j}{(S_j + L_j + R_j^L + R_j^D)^2} \sum_{k \in \Omega} (\lambda_{L_k} - \lambda_{S_k}) S_k u_{j,k} + (\lambda_{D_j} - \lambda_{L_j}) \frac{\kappa v_j L_j}{(L_j + \kappa(S_j + R_j^L))^2} \right) dt, \quad (3.33)$$

$$\lambda_{R_j^D}(t) = \int_t^T e^{-\rho(s-t)} \left(-\frac{L_j}{(S_j + L_j + R_j^L + R_j^D)^2} \sum_{k \in \Omega} (\lambda_{L_k} - \lambda_{S_k}) S_k u_{j,k} \right) ds \quad (3.34)$$

consist of indirect effects, similar to the ones for susceptibles. Recovered cases at node j lower the probability of infection from j to any other node. Light recovered cases also tend to lower the probability of detection at this node, an effect that is absent for recovered detected cases, implying they tend to be of a higher value.

Finally, the shadow price of the aggregated variable \overline{H}_j can be derived as

$$\lambda_{\overline{H}_j}(t) = -\frac{\partial \mu_H}{\partial \overline{H}_j} H_j(t) \cdot \Psi + (\lambda_{R_j^D} - \lambda_{H_j}) \frac{\partial \alpha_H}{\partial \overline{H}_j} H_j + (\lambda_{M_j} - \lambda_{H_j}) \frac{\partial \mu_H}{\partial \overline{H}_j} H_j. \quad (3.35)$$

The first part counts the value of additional lives lost due to the increase in mortality within congested hospitals; the second term values the decline in the recovery rate for heavy cases within congested hospitals; and the third term values the increase in the mortality rate from perspective of the epidemiological system (and disregarding the value of lives lost).

3.7.3 Summary tables for numerical results

	S	R_L	R_D	M	Agg. Costs	t=100	t=200	t=300	t=400
Region 1	0.03429	0.28638	0.0029	0.00977	Total	70.23	71.91	71.92	71.92
	(10.29%)	(85.91%)	(0.87%)	(2.93%)	VOL lost	69.64	71.28	71.29	71.29
					Medical	0.59	0.63	0.63	0.63
Region 2	0.03796	0.28286	0.0027	0.00981	Total	70.31	72.23	72.24	72.24
	(11.39%)	(84.86%)	(0.81%)	(2.94%)	VOL lost	69.74	71.61	71.62	71.62
					Medical	0.57	0.62	0.62	0.62
Region 3	0.03837	0.28247	0.00268	0.00982	Total	70.32	72.27	72.28	72.28
	(11.51%)	(84.74%)	(0.8%)	(2.95%)	VOL lost	69.74	71.64	71.65	71.65
					Medical	0.57	0.62	0.62	0.62
Total	0.11062	0.85171	0.00828	0.0294	Total	210.86	216.41	216.44	216.44
					VOL lost	209.13	214.53	214.56	214.56
					Medical	1.73	1.88	1.88	1.88

Table 3.6: Terminal states at $T = 400$ of the population (susceptibles S , recovered light R_L cases, diagnosed/heavy cases R_D , and deceased M) and cumulative costs in units of GDP per capita per day at days 100, 200, 300 and 400 in the "Uncontrolled" case.

	S	R_L	R_D	M	Agg. Costs	t=100	t=200	t=300	t=400
Region 1	0.09195 (27.58%)	0.23097 (69.29%)	0.00865 (2.6%)	0.00154 (0.46%)	Total	19.92	27.22	30.26	31.52
					VOL lost	5.07	7.94	10.12	11.27
					Medical	0.23	0.36	0.46	0.51
					Lockdown	14.63	18.92	19.69	19.74
Region 2	0.13344 (40.03%)	0.19112 (57.34%)	0.00715 (2.14%)	0.00127 (0.38%)	Total	7.18	15.74	20.6	22.58
					VOL lost	1.62	4.5	7.51	9.28
					Medical	0.07	0.2	0.34	0.42
					Lockdown	5.49	11.04	12.75	12.88
Region 3	0.13833 (41.5%)	0.18643 (55.93%)	0.00697 (2.09%)	0.00124 (0.37%)	Total	5.34	13.69	18.76	20.83
					VOL lost	1.32	4.11	7.2	9.05
					Medical	0.06	0.19	0.33	0.41
					Lockdown	3.96	9.4	11.23	11.37
Total	0.36372	0.60852	0.02277	0.00405	Total	32.45	56.65	69.63	74.93
					VOL lost	8.01	16.55	24.83	29.6
					Medical	0.36	0.75	1.12	1.34
					Lockdown	24.07	39.36	43.67	44.0

Table 3.7: Terminal states of the population (susceptibles S , recovered light R_L cases, diagnosed/heavy cases R_D , and deceased M) and cumulative costs in units of GDP per capita per day at days 100, 200, 300 and 400 in the "No testing" case.

	S	R_L	R_D	M	Agg. Costs	t=100	t=200	t=300	t=400
Region 1	0.10436 (31.31%)	0.21329 (63.99%)	0.01407 (4.22%)	0.00147 (0.44%)	Total	19.25	25.39	28.17	29.12
					VOL lost	5.02	7.76	9.83	10.7
					Medical	0.23	0.35	0.44	0.48
					Lockdown	14.0	17.28	17.89	17.93
					Testing	0.0	0.01	0.01	0.01
Region 2	0.14713 (44.14%)	0.16305 (48.92%)	0.02176 (6.53%)	0.00119 (0.36%)	Total	7.06	14.92	18.93	20.3
					VOL lost	1.63	4.55	7.41	8.66
					Medical	0.07	0.21	0.34	0.39
					Lockdown	5.36	10.15	11.17	11.22
					Testing	0.0	0.01	0.02	0.03
Region 3	0.15189 (45.57%)	0.15788 (47.36%)	0.02219 (6.66%)	0.00116 (0.35%)	Total	5.35	13.17	17.32	18.74
					VOL lost	1.34	4.21	7.14	8.43
					Medical	0.06	0.19	0.32	0.38
					Lockdown	3.95	8.77	9.83	9.89
					Testing	0.0	0.01	0.02	0.03
Total	0.40338	0.53422	0.05802	0.00382	Total	31.66	53.49	64.42	68.16
					VOL lost	7.98	16.52	24.38	27.79
					Medical	0.36	0.74	1.1	1.26
					Lockdown	23.31	36.21	38.89	39.04
					Testing	0.01	0.02	0.05	0.07

Table 3.8: Terminal states of the population (susceptibles S , recovered light R_L cases, diagnosed/heavy cases R_D , and deceased M) and cumulative costs in units of GDP per capita per day at days 100, 200, 300 and 400 in the "Ineffective testing" case.

	S	R_L	R_D	M	Agg. Costs	t=100	t=200	t=300	t=400
Region 1	0.18006 (54.02%)	0.11874 (35.62%)	0.03356 (10.07%)	0.00098 (0.29%)	Total	15.07	18.25	18.8	18.82
					VOL lost	4.49	6.62	7.12	7.14
					Medical	0.2	0.3	0.32	0.32
					Lockdown	10.37	11.33	11.34	11.34
					Testing	0.0	0.01	0.01	0.02
Region 2	0.22686 (68.06%)	0.07266 (21.8%)	0.03313 (9.94%)	0.00068 (0.2%)	Total	6.18	10.12	10.81	10.84
					VOL lost	1.78	4.29	4.93	4.95
					Medical	0.08	0.2	0.22	0.23
					Lockdown	4.32	5.62	5.64	5.64
					Testing	0.0	0.01	0.02	0.02
Region 3	0.23081 (69.24%)	0.06992 (20.98%)	0.03195 (9.59%)	0.00065 (0.2%)	Total	5.31	9.32	10.02	10.05
					VOL lost	1.55	4.1	4.74	4.77
					Medical	0.07	0.19	0.22	0.22
					Lockdown	3.68	5.03	5.04	5.04
					Testing	0.0	0.01	0.02	0.02
Total	0.63773	0.26132	0.09864	0.00231	Total	26.56	37.69	39.63	39.7
					VOL lost	7.83	15.01	16.79	16.85
					Medical	0.36	0.68	0.76	0.77
					Lockdown	18.38	21.98	22.03	22.03
					Testing	0.01	0.02	0.05	0.06

Table 3.9: Terminal states of the population (susceptibles S , recovered light R_L cases, diagnosed/heavy cases R_D , and deceased M) and cumulative costs in units of GDP per capita per day at days 100, 200, 300 and 400 in the "Perfect testing" case.

Chapter 4

Modelling Disaster Risks: A Dynamic Household Model

This chapter in its entirety directly corresponds to the paper Freiberger, Hoffmann and Prskawetz (2022c) in final preparation for submission. This work received financial support from the Vienna Doctoral Programme on Water Resource Systems (DK-plus W1219-N22) based at the TU Wien.

Abstract

In the last decades, many parts of the world faced an increase in the number of extreme weather events and worsening climatic conditions with negative impacts for local populations and their livelihoods. While various empirical studies have identified key factors of disaster preparedness and vulnerability, we still lack a conceptual understanding of how these forces interact and how they impact household decision making. This study develops a dynamic household model, in which households face stochastic environmental hazards, which can lead to a loss of their wealth. To respond to the risk, households can either relocate to a safer area or undertake preventive measures to protect their physical assets. Both actions require material and immaterial resources, which constrain the household's decision. Households are assumed to be heterogeneous with respect to key empirically identified factors for individual disaster risk: education, income, risk awareness, time preference and their access to preventive measures.

Theoretical insights on the optimal household strategies are derived from the FOCs and the HJB-equations. Furthermore a numerical solution for the optimal policy functions is presented calibrated to data from Thailand and Vietnam. Using Monte-Carlo-Simulations the corresponding stationary distributions are derived and compared to their empirical counterparts and the impacts of different household characteristics with a special focus on education are systematically assessed.

4.1 Introduction

Disaster risks are increasing on a global scale. Since the 1980s, the number of persons affected by weather- and climate-related hazards has increased by 43%. In 2021 alone, more than 100 million people were directly affected by disasters that were causing damages of \$230 billion (Centre for Research on the Epidemiology of Disasters (CRED), 2022). Climate change will lead to a further increase in the frequency and intensity of events, while at the same time undermining communities' capabilities to adequately prepare against hazards and cope with their consequences, leading to an increase in vulnerability in many regions (IPCC, 2022; Hoffmann and Mutarak, 2017; Black et al., 2011).

Although the human component has long been recognized as central for disaster risk management and planning (Aerts et al., 2018), hazard models often do not fully account for the influence of the social environment and behavioral drivers (Kuhlicke et al., 2020; Lechowska, 2018). Especially, the behavior of households can play an important role in mitigating disaster risks (Hoffmann and Blecha, 2020; Kohn

et al., 2012). For example, undertaking precautionary measures, such as stockpiling of food, enhancing structures, buying an insurance policy, or having savings, can be critical (Wisner et al., 2014). Also, the relocation from hazardous areas can represent an effective strategy to reduce risks, if the circumstances allow and if households are able and willing to move (Siders et al., 2019).

Households in low- and middle-income countries are particularly vulnerable as they often lack resources and capacities to adapt to and cope with environmental hazards and shocks. In line with recent efforts of the international community to reduce disaster risks and vulnerabilities, our aim is to analyze the role of human capital, specifically formal education, in influencing household vulnerability, which refers to the household's ability to adequately prepare against, respond to and cope with hazardous events. While various studies both from high as well as low and middle-income countries have reported a positive effect of education on individual aspects of disaster preparedness and vulnerability (Muttarak and Pothisiri, 2013; Meyer, 2015; Muttarak and Lutz, 2014; Hoffmann and Muttarak, 2017; Adger et al., 2005), we still lack an holistic understanding of the mechanisms through which education affects household decision making with respect to natural hazards and how these different channels interact with each other.

To this end, we develop a dynamic household model populated by households that differ in their human capital/education. Households face environmental risks and hazards, which can lead to a potentially existential loss of their wealth. To respond to the risk, households can either relocate to a safer area or undertake preventive measures to protect their physical assets. Both actions require material and immaterial resources, which constrain the household's decision.

Based on empirical findings we model four different channels through which education can influence the vulnerability of households: (i) education is positively correlated with income levels and hence financial resources rendering costly precautionary measures possible; (ii) education can provide households with access to cost-efficient prevention measures, for example through social capital/networks; (iii) information, knowledge and awareness of disaster risks are positively correlated with education; and (iv) education is related to time preferences fostering far-sighted decisions (Paton and Johnston, 2001; Drabo and Mbaye, 2015; Nawrotzki et al., 2015; Lutz et al., 2014).

Data from the Thailand and Vietnam Socio-Economic Panel (TVSEP)¹, that cover a broad range of disaster risk behaviour of households, is used to parameterize the model and assess the models predictive quality. With their diverse socio-economic background and high exposure to disaster risks, the two emerging lower-middle income countries represent well-suited empirical testing grounds for our proposed model.

Solving the dynamic household model leads to a set of optimal policy functions, which contain a prediction on household behaviour across different settings of wealth and exposure (among others). We use these policy functions to employ Monte-Carlo-Simulations for the simulation of the long-run behaviour and outcomes of a synthetic population. This synthetic population thereby reflects the compositions of the population found in the TVSEP with respect to four key household characteristics (education, prevention access, awareness and income).

The policy functions together with the simulated synthetic population allow us to gain insights on the qualitative and quantitative effects of different household characteristics on behaviour in face of natural hazards. While the policy functions allow us to identify impacts in the short-term behaviour across varying settings, the simulated households are able to illustrate the long-term effects of the different channels of influence of education. Furthermore, we can identify types of households most prone to disaster risk and establish whether households are persistently trapped in disadvantageous situation or are just temporarily in an unfortunate position along their transition paths.

By allowing evaluation of short and long term impacts of different policy interventions our framework is capable of informing public policy and resilience building efforts. We expect the predictions to be unbiased, since they are not founded on empirical correlations (which could change in the face of policy interventions), but on the behavioural decisions by households based on their intrinsic motivation to

¹See www.tvsep.de

maximize welfare.

Our model also provides interesting insights in related fields of the literature, such as on environmental migration (Hunter et al., 2015; Obokata et al., 2014; Abel et al., 2019), environmentally induced poverty traps (Sachs et al., 2004; Ikefuji and Horii, 2007; Dasgupta, 1998), and environmental management (Selin and Chevez, 1995).

The remainder of the paper is structured as follows. Section 4.2 provides an overview of the literature on education and disaster vulnerability and risks. Section 4.3 presents the conceptual framework on which the mathematical model introduced in Section 4.4 is based on. Section 4.4 includes analytical insights on the decision making process of the household. Section 4.4.5 gives a quick overview of the numerical solution strategy and Section 4.5 introduces the TVSEP as the main data source for the calibration in the numerical analysis, which is also used for the assessment of the predictive quality of the model. Finally Section 4.6 presents the numerous results of the numerical calibration exercise and compares them to the empirical evidence from the TVSEP. Section 4.7 concludes.

4.2 Education and Disaster Risk Reduction: The Empirical Evidence

There is a growing empirical literature on the relationship between education and disaster risks and vulnerability. Commonly, households with a lower socio-economic status and an on average lower education level are found to be more likely to reside in areas with higher disaster risk, which makes them more exposed to natural disasters in the first place (Adger et al., 2005; Fothergill and Peek, 2004). Given an elevated risk level, preparing against disasters and the undertaking of preventive measures is crucial. Numerous studies report that education, be it formal or informal, increases preparedness at the individual and household level, including preparedness for earthquakes (Russell et al., 1995), hurricanes (Baker et al., 2011; Norris et al., 1999; Reiningger et al., 2013), floods (Lave and Lave, 1991; Thieken et al., 2007), tsunami (Muttarak and Pothisiri, 2013), as well as general emergency preparedness (Al-Rousan et al., 2014; Smith and Notaro, 2009).

Similar findings are reported not only at the individual and household level but also at the aggregate level in country comparisons (Pichler and Striessnig, 2013). At the same time, better educated households are found to respond faster and more effectively, once a disaster strikes, e.g. by taking warnings more seriously and by evacuating faster (Sharma et al., 2013; Wamsler et al., 2012; Muttarak and Lutz, 2014). Also in the aftermath of a disaster, education has been shown to positively influence the ability to cope with and adapt to shock. Case studies include among others Indonesia (Frankenberg et al., 2013; Irmansyah et al., 2010) and Thailand (Garbero and Muttarak, 2013).

While there is convincing evidence that education positively affects preparedness, prevention, and the ability to cope with hazards, the exact mechanisms explaining its positive effects on disaster vulnerability are not fully understood. Education effects can be distinguished in direct and indirect effects. Direct effects concern any immediate effects education has on an individual, such as improving her knowledge, awareness and beliefs about natural hazards. Indirect effects, on the other hand, refer to positive influences on (material, informational, and social) individual and household resources, which allow a better preparation against and adaptation to natural hazards and harmful environmental conditions. Our theoretical model takes both direct and indirect channels into consideration, which play a role in explaining education effects on disaster vulnerability.

With regard to direct channels of influence, studies show that education equips one with knowledge, cognitive abilities, and skills that are useful when it comes to preparing for the possibility of a disaster (Blair et al., 2005; Ceci, 1991; Lee, 2010; Eslinger et al., 2009; Quartz and Sejnowski, 1997). These can be particularly helpful with understanding disaster warnings and making informed decisions about how to react. At the same time, education has been found to raise the level of awareness, helping the better educated to assess risks related to disaster threats and to find adequate responses (Bruine de Bruin et al., 2007; Peters et al., 2006). Time preferences are another channel through which education could affect

disaster preparedness, which has received less attention in the literature. Recent evidence suggests that education can change time preferences as well as the capacity to plan for the future, allowing the more educated to act more goal-oriented and to better allocate resources and make investments in financial, health or education for their future (Chew et al., 2010; Oreopoulos and Salvanes, 2011; Grossman, 2006). This could influence the adoption of such precautionary measures which require long term investments as purchasing disaster insurance.

Indirectly, education can provide households with access to different forms of resources, which enable them to better prepare against or avoid natural hazards. On average, individuals with higher formal education earn higher incomes resulting in higher wealth levels, which enables them to invest in more costly preparedness actions or the relocation from risk areas (Card, 1999; Heckman et al., 2018). Thanks to their educational background, they often also have better possibilities to diversify their income sources and have more money at their disposal to buffer negative shocks. Moreover, there is evidence showing that education improves access to informational and social resources, which can reduce vulnerability by providing households access to cost-efficient means of disaster prevention and adaptation. For example, studies have shown that education improves access to information and communication technologies (Xiao and McCright, 2007). At the same time, it was found that better educated households can build on broader and more resourceful social networks that can support them in the preparation and aftermath of disasters (Kirschenbaum, 2006; Solberg et al., 2010; Witvorapong et al., 2015).

4.3 Conceptual framework

The disaster risk for a household can be split into three components: hazard, exposure and vulnerability. The hazard of a natural disaster captures the likelihood of a disaster occurring. While increasing hazards of natural disasters shape the development of disaster risk all around the world due to changing climate conditions, for the households this development seems to be unconnected to their individual decisions and the hazard is treated as an exogenous parameter in our model. We therefore focus our analysis on exposure and vulnerability and consider the role of household's characteristics and decisions in shaping these variables.

A household's exposure captures the probability that a household is hit by a natural disaster once a disaster occurs. The level of exposure is mainly determined by the location of the household's settlement as different areas (e.g. close to a river or in a region with higher seismic activity) exhibit varying degrees of likelihood of being affected by a disaster (e.g. flood or earthquake). To assess the influence of different household characteristics on exposure and thereby replicate the differential exposure levels within countries we introduce the exposure level as a state variable of the household's decision problem. Vulnerability on the other hand captures the impact of a natural disaster on the household's well-being in a more holistic way. As a disaster affects several different aspects of a household's life (lost assets, lost income, need for restoring efforts, emotional, psychological and health impacts), assessing their resilience against a disaster requires modelling these household characteristics and their underlying decisions.

Figure 4.1 shows a conceptual flow diagram that depicts the framework of the household's decision model. First, the household faces three main dynamic constraints represented by the state variables settlement location, financial assets and physical (non-financial) assets (i.e. their durable consumption goods like housing, cars, appliances, etc.). Furthermore households can be distinguished by four key characteristics: their income level, their awareness with regard to natural disasters, their time preference rate and their access to disaster prevention measures. In each time period the household decides on (i) the new settlement location, (ii) final good consumption, (iii) financial savings, (iv) investment in physical assets and (v) prevention efforts.² These decisions depend on the previous period's state

²As described in Courbage et al. (2013) the term prevention generally describes two types of efforts in the literature of risk behaviour: (i) actions to reduce the probability of conceding losses (loss prevention) and (ii) actions to reduce the size of the losses (self-insurance or loss reduction). However since we consider preventive efforts of the first kind explicitly through settlement location (resp. exposure level), we explicitly only consider loss reduction efforts, when we refer to "prevention measures" or "prevention" within this paper.

variables and are shaped by the characteristics of the household (as captured in the right panel of Figure 4.1). Note, that we assume these characteristics to be different across different educational groups. While the settlement decision directly determines the households exposure level, the latter four

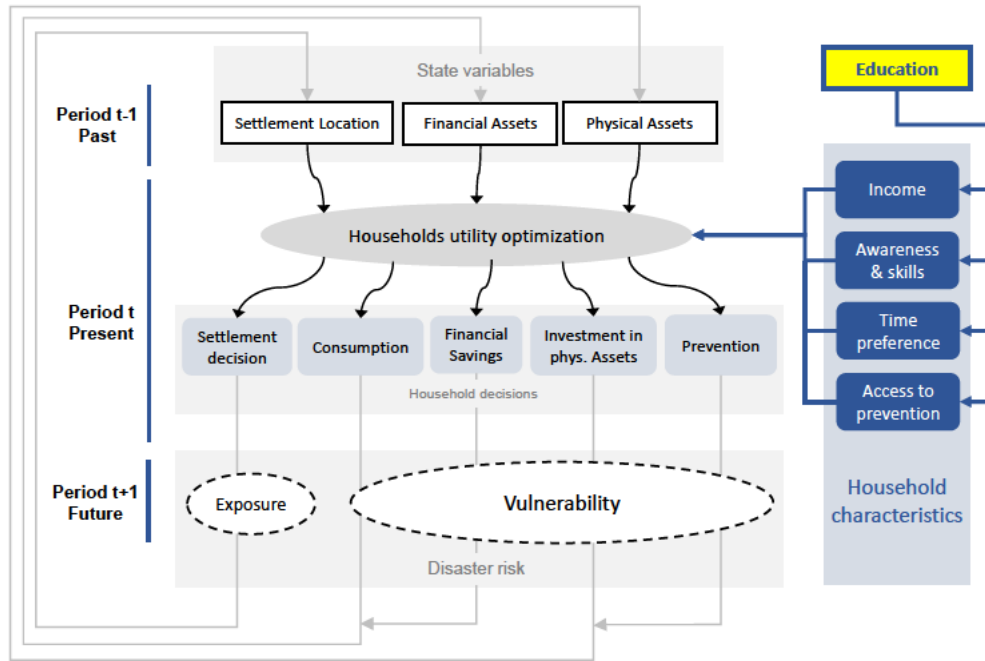


Figure 4.1: Conceptual framework

decision variables (consumption, financial savings, investment in physical assets and prevention) define the household's vulnerability. The household's decisions made in period t will determine the constraints (as represented by the state variables) in the next period $t + 1$ (as indicated by arrows going from the bottom back to the top).

In the following section we will introduce the specifications of our framework more precisely and present the mathematical expressions. This will then allow us to assess how the potential risk of a natural disaster affects the behaviour of households.

4.4 The household model

The risk of a household being hit by a natural disaster depends on two stochastic processes: $\mathcal{N} = \{N_t | t = 0, 1, 2, \dots\} \in \{0, 1\}^{\mathbb{N}}$ and $\mathcal{D} = \{D_t | t = 0, 1, 2, \dots\} \in \{0, 1\}^{\mathbb{N}}$. N_t captures the occurrence ($N_t = 1$) or absence ($N_t = 0$) of a natural disaster between time periods $t - 1$ and t . D_t on the other hand indicates whether a household is affected ($D_t = 1$) by a natural disaster or not ($D_t = 0$) between time period $t - 1$ and t conditional on a natural disaster occurring ($N_t = 1$). While the stochastic process \mathcal{N} is completely exogenous to the household, the probability of being hit by a disaster \mathcal{D} (i.e. the probabilities of $D_t = 0$ or $D_t = 1$) is shaped by the individual household behaviour.

- (i) By relocating their settlement households can alter the probability of being hit by a natural disaster in the future, i.e. they can adjust their exposure $E_{t+1} = \mathbb{P}[D_{t+1} = 1 | N_{t+1} = 1]$ at each point in time t by changing their household location.³

³In this model we assume that the exposure level is the distinctive trait of different settlement locations. In general we could still allow households to choose the place to relocate by its position L_t out of a general geographic set \mathcal{L} . However, as the location enters the household problem only through the induced exposure to disaster risk, it significantly simplifies the model to use the exposure level E_t directly as the decision variable. This allows to ignore

- (ii) A natural disaster threatens the household to loose all its physical assets W_{t+1} .⁴ To prevent such a damage, the household can undertake protective measures like investments in insurance or physical disaster protection. These efforts will determine the share P_{t+1} of physical assets that is protected from a disaster while the remaining amount, $(1 - P_{t+1})W_{t+1}$, is assumed to be destroyed in case of a disaster.
- (iii) Alternatively households can build up a stock of financial assets S_{t+1} , which are in no danger of being destroyed and furthermore generate interest at rate r_t . These assets not only allow the households to smooth the impact of uncertain labor income, but also enable them to reinvest into physical assets W especially after losses during a natural disaster. As a result financial assets enhance the households abilities to mitigate potential damages ex-post, i.e. after a disaster has occurred. Despite (in contrast to physical assets W) not contributing directly to the period utility, savings are important to smooth consumption. Labor activity and hence labor income may be reduced after a disaster when households have to invest time to deal with damages caused by natural disasters. Hence we assume that the working income is reduced by a share Δ^y , if the household was hit by a disaster in the previous time period.

These three mechanisms (how households are affected and react to a disaster) will shape the evolution of the dynamic state constraints at the household level. The decision on the settlement relocation is represented by an indicator variable I_t ($I_t = 1$ if the households relocates and respectively $I_t = 0$ if the household remains at the same settlement location). The location decision determines the level of exposure in the next period as represented by equation (4.1).

$$\begin{cases} E_{t+1} = E_t & \text{if } I_t = 0 \\ E_{t+1} \in [0, 1] & \text{if } I_t = 1 \end{cases} \iff 0 = (E_{t+1} - E_t) \cdot (1 - I_t) \quad (4.1)$$

Note, that in (4.1) we have applied the fact that the case distinction can also be written as a single constraint on the two variables E_{t+1} and I_t .

Nevertheless, relocating to a new settlement location is costly by itself and not all physical assets can be transferred to the new settlement location.⁵ We implement these losses by assuming that only a share $1 - \Delta^W$ of the households physical assets remain in their possession in case of a settlement relocation. Assets accumulate according to equation (4.2).

$$W_{t+1} = (1 - \delta)(1 - \Delta^W I_t) \widetilde{W}_t + w_t \quad (4.2)$$

Additional to potential losses due to the moving decision, the household also has to account for depreciation of physical assets obtained from the last period at rate δ . On the other hand, through sufficient investments w_t physical assets can be accumulated regardless of potential losses. We also allow w_t to be negative modeling the option for households to dissolve some of their physical into financial assets. Thereby we account for costs of dissolving physical assets and potential sales at a lower value as can be seen in equation (4.6).⁶ The physical assets \widetilde{W}_t inherited from the last time period depend on whether the household was affected by a natural disaster or not.

$$\widetilde{W}_t = \begin{cases} W_t & \text{if } D_t = 0 \\ W_t P_t & \text{if } D_t = 1 \end{cases} \quad (4.3)$$

$$= W_t \cdot (1 - (1 - P_t)D_t) \quad (4.4)$$

a functional transformation $L_t \rightarrow E_t$ and hence implies a simpler feasible domain for the relocation decision, since $E_t \in [0, 1]$.

⁴Physical/Non-financial assets W_{t+1} cover housing, cars, appliances and other durable consumption goods.

⁵Also social ties and emotional connections to locations can be seen as a part of the non-financial assets of a household. When relocating they have to be broken up and cause immaterial losses to the household.

⁶This specification is not only intuitive but also generates a disincentive for the households to constantly buy and resell physical assets, what would be a quite counter-intuitive behaviour.

The dynamics of financial assets S_t are determined by the difference of total income and total expenditures summarized in equation (4.5).

$$S_{t+1} = y_t \cdot (1 - \Delta^y D_t) + (1 + r_t)S_t - c_t - p^w(w_t) - p^P(E_{t+1}, W_{t+1}, P_{t+1}) - p^E(E_{t+1}) \quad (4.5)$$

$$p^w(w_t) = \begin{cases} w_t & \text{if } w_t \geq 0 \\ \kappa w_t & \text{if } w_t < 0 \end{cases} \quad (4.6)$$

Total income is composed of labour income y_t , which is diminished by the share Δ^y in case of a disaster, and the gross interest generated by financial assets $(1 + r_t)S_t$. Labour income is a stochastic process itself, representing the idiosyncratic income risk, summarised in $\mathcal{Y} = \{y_t | t = 1, 2, \dots\}$. However the stochastic properties (e.g. the range of income realisations and transition probabilities between them) depend on the educational level of the household as we will also present in our numerical calibration.

Total expenditure can be split up into (i) consumption expenditure c_t , (ii) investment in physical assets $p^w(w_t)$ (which could also be negative), (iii) expenditures on prevention effort $p^P(E_{t+1}, W_{t+1}, P_{t+1})$ and (iv) living costs $p^E(E_{t+1})$ depending on the settlement location/exposure level.

The costs for prevention efforts depend on the amount of physical assets and the level of protection the household aims for, but also on the exposure level of the settlement location.⁷ We assume living costs at different locations to depend on the exposure level that characterizes each location. The functional form $p^E(E_t)$ should not only reflect, that safer/less exposed areas are often characterised by higher (induced) rents, but might also imply higher opportunity costs.⁸

To determine the optimal level of expenditures, households maximize their expected utility \mathcal{U} as represented by the sum of all discounted (at rate ρ) expected future period utilities $u(c_t, W_{t+1})$ that depend on consumption c_t and physical wealth⁹ W_{t+1} . We propose an additive separable dynastic utility function.¹⁰ As indicated in equation (4.7) the expected value is built with regard to the stochastic processes of the occurrence of a disaster \mathcal{N} , the probability that the household is hit by the disaster \mathcal{D} and the stochastic process underlying the labour income \mathcal{Y} .

$$\mathcal{U} = \mathbb{E}_{\mathcal{N}, \mathcal{D}, \mathcal{Y}} \left[\sum_{t=1}^{\infty} \left(\frac{1}{1 + \rho} \right)^t u(c_t, W_{t+1}) \right] \quad (4.7)$$

We next discuss how the household characteristics affect their objectives and constraints.

4.4.1 Impact of household characteristics

As presented in Figure 4.1 four different household characteristics (which are all correlated to education) have an impact on the decision making in the utility maximisation process. We indicate the dependence of parameters and functions on education by a subscript or superscript h on parameters or alternatively as a parameter to the relevant functional forms.

Income

As we will show, using the empirical data from the TVSEP, the level and variability of income are closely correlated to education. We therefore assume that the level of education of the household determines the:

⁷This dependency on the exposure level is in accordance to actuarially fair insurance premiums, which are based on the expected costs for the insurer.

⁸Cities are often built along bodies of water. Settling in location which are less prone for flooding coincides with settling further away from the city center, which in general offers the most economic opportunities. Consequently settling further away makes either time investments for travelling to the city center necessary or implies opportunity costs of missing out on economic possibilities.

⁹Which as already previously described captures durable consumption goods like housing, cars, appliances, etc.

¹⁰Households can still be characterised by the four characteristics presented in figure 4.1 as education is empirically shown to be transferred to the next generation with a high probability. Hence, due to the correlation of education with the other key characteristics, we can also assume, that they are likely to be passed on to the next generation.

- mean/expected income level of a household.
- variance/standard deviation of income.
- persistence/correlation of income over time.

More specifically, we assume that the mean-adjusted log-income follows an AR(1)-process (with education specific variation σ_h^2 and persistence ζ_h over time) and is proportional to the mean education specific income \bar{y}_h :

$$y_t^h = \bar{y}_h \exp(\tilde{y}_t) \quad (4.8)$$

$$\tilde{y}_t = \zeta_h \tilde{y}_{t-1} + \varepsilon_t \quad \text{with} \quad \varepsilon_t \sim N(0, \sigma_h^2) \quad (4.9)$$

A similar setting for the modelling of education specific income process can be found in Krueger and Ludwig (2016).

Awareness

Although households decide about their exposure level by adjusting their settlement location, assessing the true exposure level of a given settlement location is far from a trivial task at the individual level. Frequent previous disaster experiences can lead to better judgements of future potential exposure, but longer periods of remaining unharmed might decrease the awareness of the true risk again. As presented in Section 4.2 empirical studies have also shown that awareness is positively correlated to the education of households. Changes in the climatic conditions can increase the hazard of natural disasters (i.e. $\mathbb{P}[N_t = 1] =: H_t$) unexpectedly and pose another potential element of incomplete information for the households.

To capture these various arguments how awareness is built up and develops over time we introduce an awareness parameter $a_t \in [0, 1]$ for each household, which captures levels between full awareness of the risk of different settlement locations and natural hazards ($a_t = 1$), and being completely ignorant about any disaster risk ($a_t = 0$). Note, that we ignore overestimation of disaster risk since underestimation of disaster risk seems to constitute the more relevant situation concerning the identification of drivers of high vulnerability and exposure. To indicate the impact of education on awareness, we add a superscript h to the awareness parameter: a_t^h .

The awareness affects the households utility maximisation problem through the expectation operator $\mathbb{E}_{\mathcal{N}, \mathcal{D}, \mathcal{Y}}$ in the households aggregated utility presented in equation (4.7). While the true objective probability that a household is affected by a natural disaster corresponds to the exposure level E_{t+1} of the location of its settlement times the probability H_{t+1} of a natural disaster occurring, the household underestimates the risk by building its expectations based on the adjusted subjective probability $a_t^h E_{t+1} H_{t+1}$.

Time preference

Another important factor that characterizes household risk behaviour is the time preference rate ρ , i.e. how much weight households put on the impacts of their behaviour on their well-being in the future compared to the present. Empirical evidence such as e.g. Viscusi and Moore (1989) and Jung et al. (2021) indicates that higher educated individuals are more likely to be forward looking and put relatively more weight on their future. We will indicate the education dependence of the time preference rate by adding a subscript to the time preference rate, i.e. ρ_h .

Access to prevention

Multiple studies have empirically shown, that households or individuals with higher education levels can access disaster prevention measures more easily either through their enhanced knowledge or their more

pronounced social networks. To implement this aspect into our framework, we argue that better access to prevention is equivalent to lower costs of protection. We propose, that the efforts needed to ensure the same level of prevention P_t are significantly lower for households with better access to prevention measures. We therefore adjust the prevention cost function for an educational specific parameter: $p^P(E, W, P, h)$. The specific functional forms we apply for our numerical simulations are summarized in Section 4.5.

4.4.2 Problem summary

The complete household optimisation problem is summarized in the following set of equations:

$$c_t, w_t, I_t, E_{t+1}, P_{t+1}, \mathbb{E}_{\mathcal{N}, \mathcal{D}, \mathcal{Y}} \left[\sum_{t=1}^{\infty} \left(\frac{1}{1 + \rho_h} \right)^t u(c_t, W_{t+1}) \right] \quad (4.10)$$

$$S_{t+1} = y_t^h \cdot (1 - \Delta^y D_t) + (1 + r_t)S_t - c_t - p^w(w_t) - p^P(E_{t+1}, W_{t+1}, P_{t+1}, h) - p^E(E_{t+1}) \quad (4.11)$$

$$W_{t+1} = (1 - \delta)(1 - \Delta^W I_t)(1 - (1 - P_t)D_t)W_t + w_t \quad (4.12)$$

$$(E_{t+1} - E_t)(1 - I_t) = 0 \quad (4.13)$$

$$\mathcal{N} \sim \begin{cases} \mathbb{P}[N_t = 1] = H_t \\ \mathbb{P}[N_t = 0] = 1 - H_t \end{cases} \quad (4.14)$$

$$\mathcal{D} \sim \begin{cases} \mathbb{P}[D_t = 1] = a_{t-1}^h E_t H_t \\ \mathbb{P}[D_t = 0] = 1 - a_{t-1}^h E_t H_t \end{cases} \quad (4.15)$$

Note that the optimisation in equation (4.10) is conducted with respect to the decision variables $(c_t, w_t, I_t, E_{t+1}, P_{t+1})$. S_{t+1} and W_{t+1} are omitted here, since the dynamic equations (4.11) and (4.12) completely define those two values given the other five decision variables.¹¹

4.4.3 Analytical results

We can use the Lagrange-approach for stochastic constrained optimisation problems to obtain insights into the inter- and intratemporal trade-offs in the households decision making process.¹²

Proposition 4 (First order conditions (I))

Assume an optimal solution

$$(c_t^*, w_t^*, I_t^*, E_{t+1}^*, W_{t+1}^*, S_{t+1}^*, P_{t+1}^*)_{t=1,2,\dots}$$

for problem (4.10)-(4.15) exists. For points in time t , where the solutions $(c_t^*, W_{t+1}^*, S_{t+1}^*, P_{t+1}^*)$ are in the interior of the feasible region, the optimal solution fulfils the following first order optimality conditions. (To abbreviate, we will not list all variables within each functional form for the longer equations, but indicate at which point in time their decision is made, e.g. $u_c(t) := u_c(c_t, W_{t+1})$.)

$$u_c(c_t, W_{t+1}) = \frac{1 + r_t}{1 + \rho} \mathbb{E}_t \{ u_c(c_{t+1}, W_{t+2}) \} \quad (4.16)$$

$$u_c(t) p_P^P(t+1) = \frac{1 - \delta}{1 + \rho} \mathbb{E}_t \{ u_c(t+1) (p^w)'(w_{t+1}) (1 - \Delta^W I_{t+1}) W_{t+1} D_{t+1} \} \quad (4.17)$$

¹¹Alternatively, we could also omit c_t and w_t as decision variables instead and define their values through equations (4.11) and (4.12) using the other decision variables. We will use this equivalence when solving the household problem numerically.

¹²We will omit the sub-/superscript h explicitly indicating the impact of the education in the following to limit the convolution of the notation. However we encourage the reader to keep the impacts of education in mind.

$$\begin{aligned}
u_c(t) [p_W^P(t+1) + (p^w)'(t)] &= \\
&= u_W(t) + \frac{1-\delta}{1+\rho} \mathbb{E}_t \{ u_c(t+1)(p^w)'(t+1)(1 - \Delta^W I_{t+1})(1 - (1 - P_{t+1})D_{t+1}) \}
\end{aligned} \tag{4.18}$$

$$\begin{aligned}
u_c(t) [p_W^P(t+1) + (p^w)'(t)] &= u_W(t) + \\
&+ \mathbb{E}_t \left\{ \sum_{i=1}^{\infty} \left(\frac{1-\delta}{1+\rho} \right)^i u_W(t+i) \cdot \prod_{j=1}^i \frac{(p^w)'(t+j)}{(p^w)'(t+j) + p_W^P(t+1+j)} (1 - \Delta^W I_{t+j}) (1 - (1 - P_{t+j})D_{t+j}) \right\}
\end{aligned} \tag{4.19}$$

For the proof we refer to Appendix 4.8.1. In the following we provide the economic intuition behind these equations.

Equation (4.16) builds the intertemporal Euler equation for consumption. The marginal utility gain from consumption in the current period t has to be equal the expected marginal utility from consumption in the next period $t+1$ adjusted for gains through interest rate on financial assets and the utility discount rate.

Equation (4.17) illustrates the trade-off between prevention efforts and consumption. On the left hand side the marginal costs of additional prevention are measured in units of marginal utility from consumption. On the right hand side we have the expected benefits of prevention measured in discounted marginal utility from consumption in the next time-period ($\frac{u_c(t+1)}{1+\rho}$). The benefits of prevention materialize in form of additional units of physical assets being protected and preserved in case of a natural disaster ($W_{t+1}D_{t+1}$). However, before the physical assets can be liquidized ($(p^w)'(t)$) into financial assets and be used for consumption, we need to account for losses through depreciation ($1-\delta$) and potential relocation of the household settlement in the next period ($1 - \Delta^W I_{t+1}$). Note that the benefits of prevention are only present in case the household is affected by a natural disaster and there is no positive impact of prevention otherwise. This affect will become more apparent in equation (4.48) later on.

For the discussion of the trade-off between consumption and investments into physical assets we present two different equations (4.18) and (4.19), which are equivalent, but allow for different interpretations. Equation (4.18) takes a similar approach to the prevention/consumption-trade-off as in (4.17). The additional costs of a marginal increase in wealth investment measured in units of marginal utility from consumption constitutes the left hand side. These costs now consist of two parts: The first term covers the additional expenditures for prevention whereas the second term contains the direct investment costs. If the additional wealth investment implies an increase in already positive wealth investment, $(p^w)'(t) = 1$ holds. On the other hand, if reduced liquidation is implied, $(p^w)'(t) = \kappa$ holds. The right hand side of (4.18) also consists of two separate effects. $u_W(t)$ captures the direct impact of the additional wealth investment in terms of utility gained in the present. The latter part in (4.18) allows for an analogue interpretation as Equation (4.17). The additional physical assets are changed to $(1 - (1 - P_{t+1})D_{t+1})$, while the other parts of the terms are identical. In case of a natural disaster the additional physical assets equate to P_{t+1} as only the protected share of the additional marginal physical assets is transferred to the next period. Otherwise this term equals one and all assets are transferred.

Equation (4.19) illustrates the same trade-off, while avoiding the transformation back into units of consumption on the right hand side. The left hand side and the first part of the right hand side are identical to (4.18). The second part now contains the expected impact of the additional marginal physical assets over the future infinite time-horizon. For each time-period the household has to adjust the gains in marginal utility $u_W(t+i)$ for depreciation of physical assets and the discount factor of utility $\left(\frac{1-\delta}{1+\rho}\right)^i$. Furthermore the gains have to be adjusted for (i) potential losses through disaster experience ($(1 - (1 - P_{t+j})D_{t+j})$), (ii) potential losses resulting from settlement relocation ($1 - \Delta^W I_{t+j}$), and (iii) additional costs resulting from additional prevention costs $\left(\frac{(p^w)'(t+j)}{(p^w)'(t+j) + p_W^P(t+1+j)}\right)$.

In the next section we discuss the Bellman formulation of problem (4.10)-(4.15). In the analysis of

the current section, the effects of the exposure/settlement location decision of the household are only indirectly covered through the expectation operator \mathbb{E}_t . The Bellman formulation on the other hand allows us to consider these effects more explicitly.

4.4.4 Bellman formulation

The Bellman equation uses the value function $V(E, S, W, Y, D)$ to describe the optimal objective value depending on the initial value of the state variables and the stochastic processes.¹³ Due to the infinite time horizon and the time invariance of the complete problem (4.10)-(4.15), the value function fulfils the system of equations (4.20)-(4.23).

$$V(E_t, S_t, W_t, y_t, D_t) = \max_{\substack{E_{t+1}, S_{t+1}, \\ W_{t+1}, P_{t+1}, \\ I_t, c_t, w_t}} \left\{ u(c_t, W_{t+1}) + \frac{1}{1+\rho} \left[a_t E_{t+1} H_{t+1} \cdot \mathbb{E}_y V(E_{t+1}, S_{t+1}, W_{t+1} \cdot P_{t+1}, \mathcal{Y}, D_{t+1} = 1) + \right. \right. \\ \left. \left. + (1 - a_t E_{t+1} H_{t+1}) \cdot \mathbb{E}_y V(E_{t+1}, S_{t+1}, W_{t+1}, \mathcal{Y}, D_{t+1} = 0) \right] \right\} \quad (4.20)$$

$$S_{t+1} = y_t \cdot (1 - \Delta^y D_t) + (1 + r_t) S_t - c_t - p^w(w_t) - p^P(E_{t+1}, W_{t+1}, P_{t+1}) - p^E(E_{t+1}) \quad (4.21)$$

$$W_{t+1} = (1 - \delta)(1 - \Delta^W I_t) W_t + w_t \quad (4.22)$$

$$0 = (E_{t+1} - E_t) \cdot (1 - I_t) \quad (4.23)$$

The Bellman-Equation (4.20) illustrates that the optimized objective value in the present combination of state variables $(E_t, S_t, W_t, y_t, D_t)$ is equal to the optimal trade-off between utility derived from consumption and physical assets in the current period $u(c_t, W_{t+1})$ and the discounted expected value from the value function in the next time period. The latter part by itself consists of two different terms. The first part contains the scenario with the household being affected by a disaster, which consequently is weighted with the probability $a_t^h E_{t+1} H_{t+1}$. In this case the households enter the next time period with their physical assets being equal to $W_{t+1} \cdot P_{t+1}$ and this term enters the value function.¹⁴ The second part covers the case of households not being hit by a natural disaster. The prevention measures have no effect here and the household enters the next time period with the same level of physical assets they ended up with in the current period, i.e. W_{t+1} .

Solving the original stochastic problem over an infinite time horizon (4.10)-(4.15) is equivalent to finding a value function $V(\cdot)$, which solves the problem in Bellman formulation (4.20)-(4.23). Furthermore solving for the value function directly implies the optimal decision rules for the household for each possible combination of (E, S, W, y, D) . These decision rules at the same time maximize the expected utility over the infinite time horizon presented in equation (4.10) and they are denoted as the policy functions.

Calculating the FOCs for the optimisation problem in the right hand side of (4.20) leads to equations allowing for similar interpretations regarding the optimal decisions with respect to financial assets, physical assets and prevention. Due to their similarity to the results of Proposition 4, we relegate them to Proposition 5 in the Appendix.

These necessary optimality conditions for consumption, financial and physical assets and prevention hold regardless of the relocation and exposure decisions of the household. Considering the remaining two parts of the overall optimal strategy gets slightly more complex. We cannot derive a first order optimality condition for the moving decision, because it is by definition a boundary solution on $[0, 1]$.

¹³Therefore it would hold that $V(E_0, S_0, W_0, y_0, D_0) = \mathcal{U}$.

¹⁴Note that we slightly adapted the dynamic equation of the physical assets accordingly to $W_{t+1} = (1 - \delta)(1 - \Delta^W I_t) W_t + w_t$. This adjustment allows us to omit the prevention level of the previous time period as an additional state variable and formulate the value function depending on the 5 state variables exposure, financial and physical assets, income and disaster experience.

However distinguishing between the case of relocation or not enables us to gain some insights into the decision process on the settlement location. Assuming an interior solution for the exposure level decision we obtain the first-order-optimality condition (4.24).

$$u_c(c, W) \cdot \left[\frac{\partial p^E}{\partial E}(E) + \frac{\partial p^P}{\partial E}(E, W, P) \right] - \lambda_I(1 - I) = \frac{1}{1 + \rho} \left[aEH \cdot \mathbb{E}_y \frac{\partial V}{\partial E}(E, S, W \cdot P, \mathcal{Y}, 1) + \right. \\ \left. + (1 - aEH) \cdot \mathbb{E}_y \frac{\partial V}{\partial E}(E, S, W, \mathcal{Y}, 0) + aH \cdot \mathbb{E}_y \left(V(E, S, W \cdot P, \mathcal{Y}, 1) - V(E, S, W, \mathcal{Y}, 0) \right) \right] \quad (4.24)$$

The right hand of (4.24) contains the marginal changes in utility through the value function in the next period (similarly to the conditions in Proposition 5). The first two terms contain the marginal changes in the value function for marginal changes in E for the disaster and no-disaster scenario respectively. Both are weighted with the probability at which they occur. However changes in the exposure level not only affect the value function, but also the probabilities at which the disaster occurs or does not occur. Hence, there is an additional term in (4.24), which contains the change in expected utility for higher probability of disaster occurrence. Again this effect is weighted with the awareness a , so it has an increasing impact on the decisions of households in case of higher awareness.

On the left hand side we find the marginal change in utility from consumption through changes in the costs with respect to the exposure level. However, there is the additional term $-\lambda_I(1 - I)$ and for its analysis we need to distinguish two cases:

- In case it is optimal for the household to relocate (i.e. $I = 1$), the new exposure level for the settlement location is chosen optimally and the marginal benefits and marginal costs (in terms of consumption) have to be equal, since $\lambda_I(1 - I) = 0$ holds.
- In case it is optimal for the household to remain in the same location (i.e. $I = 0$), λ_I contains the difference between marginal benefits and marginal costs related to the exposure level of the current settlement location.

4.4.5 Numerical solution strategy

As the previous section has shown, the analytical insights into the solution of problem (4.10)-(4.15) (resp. problem (4.20)-(4.23)) are limited. While the results of Proposition 4 depend on complex expected values of future behaviour, the results of Proposition 5 (in the Appendix) require derivatives of the in general unknown value function $V(\cdot)$. Both aspects are hard to overcome so we will focus on the numerical solution of the model for the remainder of the paper. As finding a numerical approximation of the unknown value function $V(\cdot)$ and the corresponding optimal policy functions is far from trivial itself and requires significant computational efforts, we provide a description of our approach in Appendix 4.8.2.

4.5 Parametrisation and calibration

The Thailand-Vietnam-Socioeconomic-Panel (TVSEP) is a long term project collecting household panel data already over multiple waves. Each survey consist of a sample size of almost 4400 households in 440 villages in rural areas of six provinces in Thailand and Vietnam. Since the questionnaire contains multiple questions on the risk behaviour and perception of households, the TVSEP is an appropriate database for the parametrisation and calibration of our model. In the next section we will discuss the estimation of the parameters of our model and present the functional specification necessary. In Section 4.5.2 we will present the strategy for the construction of distribution of state variables according to our model.

4.5.1 Parameter estimates

First we want to stress, that some parameters of our framework are rather abstract measurements and therefore hard to quantify in the empirical data. Hence for certain parameters plausible guesses were the only feasible strategy moving forward. At the end of this section we summarise all parameters and functional forms in Table 4.1.

We assume one period of time within our model to be two years long, so all parameters are adjusted accordingly. We assume the interest rate to be at moderate 4% per year, resulting in $r = 0.0816$. For the depreciation of physical capital we assume a value of 2% per year, what leads to $\delta = 0.0396$. The shares of physical assets lost during relocation or in the transformation process into financial assets are assumed to be 20% each, what corresponds to values of $\Delta^W = 0.2$ and $\kappa_Z = 0.8$ respectively.

For the share of income lost after suffering from a natural disaster, we can consult the TVSEP data, as it contains specific information on the income households lost through natural disasters. Setting these losses in relation to the household income, we obtain a mean value of $\Delta^y := 0.1499$ as an estimate.

Utility function

For the period utility function we assume an additive separable function consisting of two CRRA-type terms in consumption and physical assets, i.e.

$$u(c, W) = \frac{(c - \underline{c})^{1-\gamma} - 1}{1-\gamma} + \theta \frac{(W - \underline{W})^{1-\beta} - 1}{1-\beta}. \quad (4.25)$$

Utility from physical assets thereby is weighted with the parameter θ compared to utility derived from consumption. We found the parameter choice of $\gamma = 1.5$, $\beta = 2.0$ and $\theta = 1.0$ to be reasonable and in line with other literature generally finding CRRA-parameters between 1.0 and 2.0.

Income process

As already presented Section 4.4.1 we assume, that the demeaned log-income follows an AR(1) process, i.e.

$$y_t^h = \bar{y}_h \exp(\tilde{y}_t) \quad (4.26)$$

$$\tilde{y}_t = \zeta_h \tilde{y}_{t-1} + \varepsilon_t \quad \text{with} \quad \varepsilon_t \sim N(0, \sigma_h^2) \quad (4.27)$$

Using the data from the TVSEP we can estimate the mean income for the different education groups \bar{y}_h . Using two different waves furthermore allows us to derive estimates for the persistence of income ζ_h and the variance of the idiosyncratic income shock σ_h^2 . Note that we normalised all income levels with respect to the mean income of the lowest education group. Hence we will measure all costs and expenditures in this calibration exercise in these Basic Income Units [BIU].

For the numerical calculation and simulation we discretized the space of potential incomes using the Tauchen algorithm (see Tauchen (1986)) into five different income realisations. The numerical education specific values can be found in Table 4.1.

Living costs

The living costs are one of the rather abstract measurements within our framework. We decided to use the value of land owned by the households as a measure for the living costs and calculate an imputed rent. Using the average disaster experience within the same village as an approximation for the exposure level of the households within their village allows us to derive an appropriate formulation of the living costs in relation to the exposure level. Among several functional forms,

$$p_E(E) = \underline{p}_E \exp(\phi(1 - E)^2) - \tilde{p}_E \quad (4.28)$$

has shown to provide an adequate fit to the data and exhibits intuitive qualitative characteristics. The parameter \underline{p}_E provides the lower bound for the living costs in the most exposed locations and the marginal costs for a reduction in the exposure level in these points are equal to zero. The estimated parameters are rather similar for Thailand and Vietnam and we chose the average values for each parameter between the countries for our calibration, i.e. $\underline{p}_E = 0.482$ and $\phi = 1.075$.

For the numerical solution of the model we introduce the parameter \widetilde{p}_E , which adjusts the lower bound of costs according to the minimum working income a household generates. This adjustment makes sure that households are able to afford their basic needs of consumption \underline{c} , physical assets \underline{W} and housing in the most exposed area, i.e. $p_E(1)$.

Prevention costs

For the prevention costs we start with the actuarially fair insurance premium of $p_P(E, W, P) = (1 - \delta)EWP$.¹⁵ However we assume, that for households with better access to prevention measures, the first measures taken are cheaper and only for full protection the same total costs would arise. We decided to reflect this convex cost structure by adding an exponent $(1 + \varphi_P)$ to the preventive effort.

$$p_P(E, W, P) = (1 - \delta) \cdot E \cdot W \cdot P^{1+\varphi_P} \quad (4.29)$$

¹⁵Note that the households only pay for the depreciated value of the physical assets $(1 - \delta)W$. We assume that the disaster occurs between two time periods and the physical assets from the last period get depreciated right at the beginning of the next time period. This means, that each physical asset, that the household has insured against damage from a natural disaster and gets replaced in case of occurrence also depreciates right afterwards. Hence the household actually only gets the present value of the assets from the insurance and hence should also only pay a premium according to the same value.

Share of phys. assets lost during relocation	Δ^W	0.2
Value of phys. assets after liquidation	κ_Z	0.8
Interest rate	r	$(1 + 0.04)^2 - 1 = 0.0816$
Depreciation of phys. assets	δ	$1 - (1 - 0.02)^2 = 0.0396$
Share of income lost after disaster	Δ^y	0.1499
$u^1(c) = \frac{(c-\underline{c})^{1-\gamma}-1}{1-\gamma} + \theta \cdot \frac{(W-W)^{1-\beta}-1}{1-\beta}$	γ	1.5
	β	2.0
	θ	1.0
	\underline{c}	0.0001
	\underline{W}	0.0001
	$p_E(E) = \underline{p}_E \exp(\phi(1-E)^2) - \widetilde{p}_E$	\underline{p}_E
ϕ		1.075
\widetilde{p}_E		0.111
Education category	H_0	(1,2,3,4,5)
$y_t^h = \bar{y}_h \exp(\tilde{y}_t)$ $\tilde{y}_t = \zeta_h \tilde{y}_{t-1} + \varepsilon_t$ with $\varepsilon_t \sim N(0, \sigma_h^2)$	\bar{y}_h	(1.0, 1.148, 1.641, 2.409, 5.375)
	ζ_h	(0.322, 0.315, 0.430, 0.489, 0.486)
	σ_h	(0.775, 0.756, 0.742, 0.730, 0.638)
Awareness	a^h	(0.7, 0.825, 0.95)
Time preference	ρ^h	(0.1306, 0.1534, 0.1763)
Access to prevention	φ_P^h	(0.2, 0.6, 1.0)

Table 4.1: Summary of functional specifications and parameters in the model.

4.5.2 Empirical distributions

For the construction and simulation of a synthetic population resembling the empirical data, we need to pin down the distribution across the key household characteristics (i) education, (ii) awareness, (iii) prevention access and (iv) time preference within the TVSEP data set.

- (i) For the education grouping we classify households into 5 different groups according to the median educational attainment within the household.
- (ii) For the awareness categorisation we use the information on the number of shocks the household is expecting to face in the next years and compare them with the number of shocks the household actually suffered as stated in the consequent wave of the questionnaire. We assigned households into three different categories according to the accuracy of their prediction of number of shocks suffered.
- (iii) For the access to prevention measures we use the access to financial support as a proxy. Taking the number of days each household would need to acquire a given amount of money enables us to assign each household an access score between zero and one, which we can consequently also classify into three categories.
- (iv) As information on time preference of households is not part of the TVSEP, we propose a negative correlation of -0.3 between education level and time preference, which reflects the fact, that higher educated households are likely to be more forward looking. Therefore we assign each household a

random time preference category in a way that replicates the desired correlation over the whole population. Consequently, we again grouped households into three categories with high, mid and low time preference rates.

The respective parameters a^h , ρ^h and φ_P^h presented in chosen to cover a reasonable range of values (awareness between 0.7 and 0.95 resp. access to prevention between 0.2 and 1.0) and lead to realistic outcomes of the model (time preference is deliberately chosen higher than the interest rate, since otherwise households would have an unrealistically high incentive for financial savings.)

Figure 4.2 provides an overview of the distribution of households across education, awareness and prevention access described above. The largest share of the population is classified as high awareness, with low and mid awareness representing population groups of more similar size. According to these empirical distributions households predominantly have high access to prevention measures (across all education and awareness categories) and those households with mid or even low access are more likely in a lower education level.

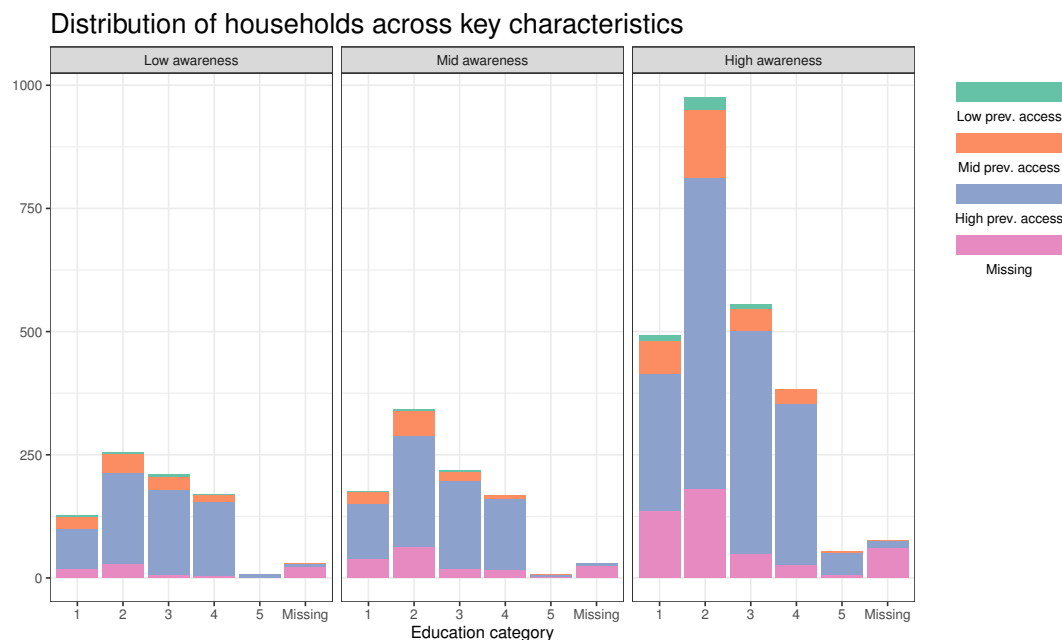


Figure 4.2: Distribution of households across the key characteristics education, awareness and prevention access.

4.6 Numerical results

For the discussion of the numerical results, we constructed and simulated a synthetic population of households reflecting the compositional distribution of characteristics presented in Figure 4.2. In Section 4.6.1, we first discuss the long-run equilibrium distribution of the population across the different state variables and other factors of disaster risk. Furthermore we assess the impact of the household characteristics with a specific focus on the effects of education on the equilibrium outcomes in Section 4.6.2. We investigate the decision making of households in Section 4.6.3 with regard to the different situations and scenarios households are facing. Again we will identify the impact of the key household characteristics in selected settings. Lastly, in Section 4.6.4 we present the impact of education, awareness, time preference, prevention access and income on the households disaster risk level and its components exposure and vulnerability.

4.6.1 Equilibrium distributions

Before discussing the distributions in equilibrium along the different state and control variables, we break down the correlations between some of the key variables of the model. Figure 4.3 shows the pairwise correlations for consumption, income, disaster experience, physical and financial assets and exposure both in the simulated data (left panel) and empirical data (right panel). Although Figure 4.3 shows that we manage to replicate the empirical correlations qualitatively well with our model, differences in quantitative values for some of the variables are apparent.

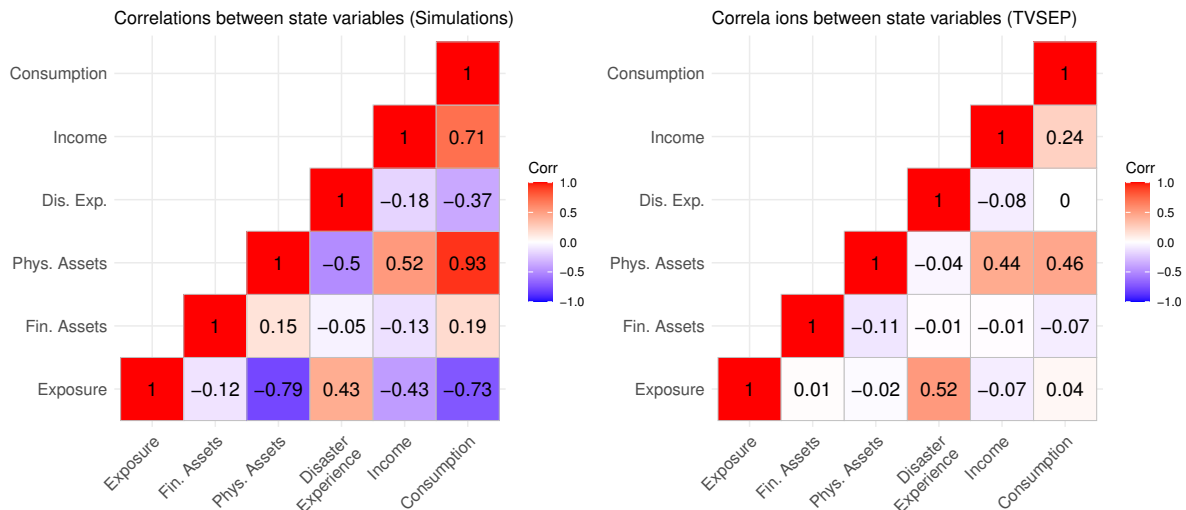


Figure 4.3: Correlations between the state variables in equilibrium for the simulation results (left) and the empirical data (right)

Comparing the two panels we see that the equilibrium outcomes of the model simulations have a tendency to exhibit stronger correlations compared to the empirical data.¹⁶ In this regard, first note that we match the positivity of the correlation between income and consumption, but overestimate it to some degree (0.71 and 0.24 respectively). On the other hand the empirical correlation between income and disaster experience of -0.08 is replicated closely in the model with -0.18 . However, while the model would predict a negative correlation of -0.36 between disaster experience and consumption, the TVSEP data does not find any significant correlation between these two variables.

Concerning physical assets we are able to replicate the positive correlations with income and consumption (with the correlation between physical assets and consumption being close to 1 and therefore considerably stronger compared to the empirical data), whereas the rather small negative correlation with disaster experience (-0.04) is predicted to be substantially more pronounced in the model (-0.5).

For the correlations with respect to financial assets in the fifth row in Figure 4.3 we do not find the same sign in the model as in the data for two of them, however, the correlations are all comparatively close to zero (between ± 0.2), which we are able to fully replicate with our model (i.e. phys. assets, disaster experience, income and consumption).

Finally we find the arguably largest differences in the correlations with respect to the exposure level. Whereas the correlations with financial assets (0.01 resp. -0.12) and disaster experience (0.52 resp. 0.43) are closely matched¹⁷, the model predicts strong negative correlations with physical assets, income and

¹⁶This indicates, that our framework is still a considerable simplification of reality. We abstract from e.g. other defining attributes of the costs of the settlement location and hence do not account for a variety of potentially mitigating factors. On the other hand also the empirical data for the corresponding variables of the model had to be constructed from other available data. Therefore the data is not able to perfectly represent the specific model variables.

¹⁷The positive correlation of exposure and disaster experience is to some degree explained by construction in the data

consumption, which we do not find in the TVSEP dataset.

These differences between model outcomes and empirical data can be explained through the absence of physical modelling of disaster risk and exact geo-specific location data of each household not being readily available in the survey. For the assignment of an exposure level to each household in the TVSEP we used the information on the number of natural disasters each household was affected by during the last two years. We use this information to construct the average probability within each village of a household suffering from at least one disaster in the past and use that value as the estimate for the exposure level of each household in the village.¹⁸ Hence the exposure indicator in the data should be taken with a grain of salt. In future work, we aim to obtain proper disaster risk data for household settlement locations, whereas the current definition should still be sufficient for the illustration of the working of our model framework.

In a next step Figure 4.4 presents the equilibrium distributions of exposure, financial and physical assets, consumption and prevention decision for the simulated data in comparison to the distribution within the TVSEP.¹⁹ Note that all variables related to financial units (e.g. income, financial and physical assets, consumption) are measured in terms of basic income units [*BIU*], i.e. the mean period income of the lowest education group (see also the specification for the income process in Table 4.1).

Exposure: In the simulated data as well as in the empirical data we find households allocated along all different exposure levels. The most exposed locations are the least populated and a majority of the population is situated in locations with intermediate exposure between 0.25 and 0.75. The most significant difference between the empirical and the simulated data is in the share of households living at locations with zero (or close to zero) exposure, which the model would predict to be substantially higher. Otherwise the model is able to replicate the empirical distribution fairly well.

Financial assets: The financial assets show the most significant discrepancy between simulation and data. The data from the TVSEP exhibits a drastically lower range of financial savings. While the empirical data shows mostly values of financial savings between $\pm 2.5[BIU]$, the simulated data exhibits a significantly larger range from about $-6[BIU]$ to $+10[BIU]$ for the majority of households. The small range in the data is rather surprising, as some households in the dataset have a period income of more than $5[BIU]$. For such high income levels we would intuitively expect them to be able to accumulate higher levels of financial savings than can be found in the data. Future work will investigate the phenomenon further and we will adjust the model accordingly.

Physical assets: The distributions for physical assets overall coincide nicely for data and simulations. Nevertheless we see slightly more households with physical assets in the range of $3.5 + [BIU]$ in the simulations.

Consumption: As for the physical assets the empirical consumption distribution is replicated qualitatively well through the model. Still the empirical distribution exhibits a heavier right tail and therefore higher consumption levels are slightly underrepresented in the simulated data.

Prevention: For the prevention decision, we were not able to construct a corresponding variable in the empirical dataset, so we only discuss the simulation results. In equilibrium the majority of households decide to protect between 25% and 50% of their physical assets against disaster risk. Overall the distribution is left-skewed with the mode of the distribution being close to 0.5 and a steep decrease of the distribution for higher prevention levels.

resp. the definitions in the model. The fact that the two are relatively close quantitatively, however, can be seen as a confirmation for the validity of the model.

¹⁸We take a cross-population average of disaster suffering instead of a longitudinal average across time for each household, so we do not require information across multiple waves of the TVSEP or have to account for households changing their settlement location.

¹⁹There is no appropriate comparison for the prevention decision in the TVSEP dataset, so we omitted plotting distribution of prevention for the empirical data.

Overall, we were able to reproduce all empirical distributions qualitatively (and for parts also quantitatively) quite well. The wider range of financial assets and underestimated number of higher consumption levels would at first thought indicate, that the discount rate for the simulations were chosen to low (higher time preference would in general give present consumption higher priority than savings). However, as Table 4.2 will show, higher time preference actually implies lower financial savings and lower consumption in the long-run equilibrium.

4.6.2 Impact of education and other household characteristics

In this section we want to assess the impact of education and other household characteristics on the equilibrium distributions. Figure 4.5 illustrates the distributional effects of different levels of educational attainments by households in the simulated and empirical data.

It is most apparent that within each educational group the distributions in the empirical data are more widely and equally spread compared to the simulated data. Furthermore the different education groups also cover more similar ranges of values in the empirical data. As a result, the impact of the education level on the equilibrium distributions can be seen as less pronounced in reality compared to the prediction of the model. However the qualitative impacts of education are replicated well by the model and we will discuss potential sources and origins for discrepancies in quantitative terms.

Exposure: For most education groups (accept the highest educated) the model predicts an equilibrium distribution covering a wide range of different exposure levels. For the lower educated groups the distribution is progressively shifted towards the right, i.e. they exhibit higher exposure levels on average. The empirical data is less consistent in this regards and except for the highest education group, there is no distinct impact of different educational levels on the distributions.²⁰

Financial assets: For increasing educational levels the distribution of financial assets becomes flatter and covers a wider range of positive and negative values. This qualitative aspect is present in both the simulation and empirical data, but the impact of education is more pronounced in the simulations. Furthermore we see that the distributions are consistently centered around zero in the empirical data. In contrast, the simulated distributions indicate that lower educated households tend to hold slightly positive financial assets, whereas the mean value of assets decreases and becomes negative for higher educated households.

Physical assets: Concerning physical assets the impact of education is qualitatively similar in the empirical and simulated data. The variance of the distribution increases for higher educated groups implying flatter distributions and the mean values are shifted to the right. However, as discussed before regarding other variables the model also suggest a considerably stronger impact of education than can be found in the data. Note that the model is still able to reproduce, that a significant share of households in the second highest education group faces similarly low levels of physical assets as average households on the lowest two educational levels.

Consumption: The effect of education on consumption is qualitatively analogous to the effect on physical assets both in the simulation and empirical data. Hence we refrain from repeating the conclusions of the previous bullet point here.

Prevention: For the simulated equilibrium distributions of the prevention decision we find that interestingly the education level has no considerable effect for the first four education groups. Only the households with the highest educational attainment show significantly higher preventive efforts on average (in equilibrium).

²⁰We refer to a previous discussion on the construction and applicability of the exposure indicator in the data.

Equilibrium distributions

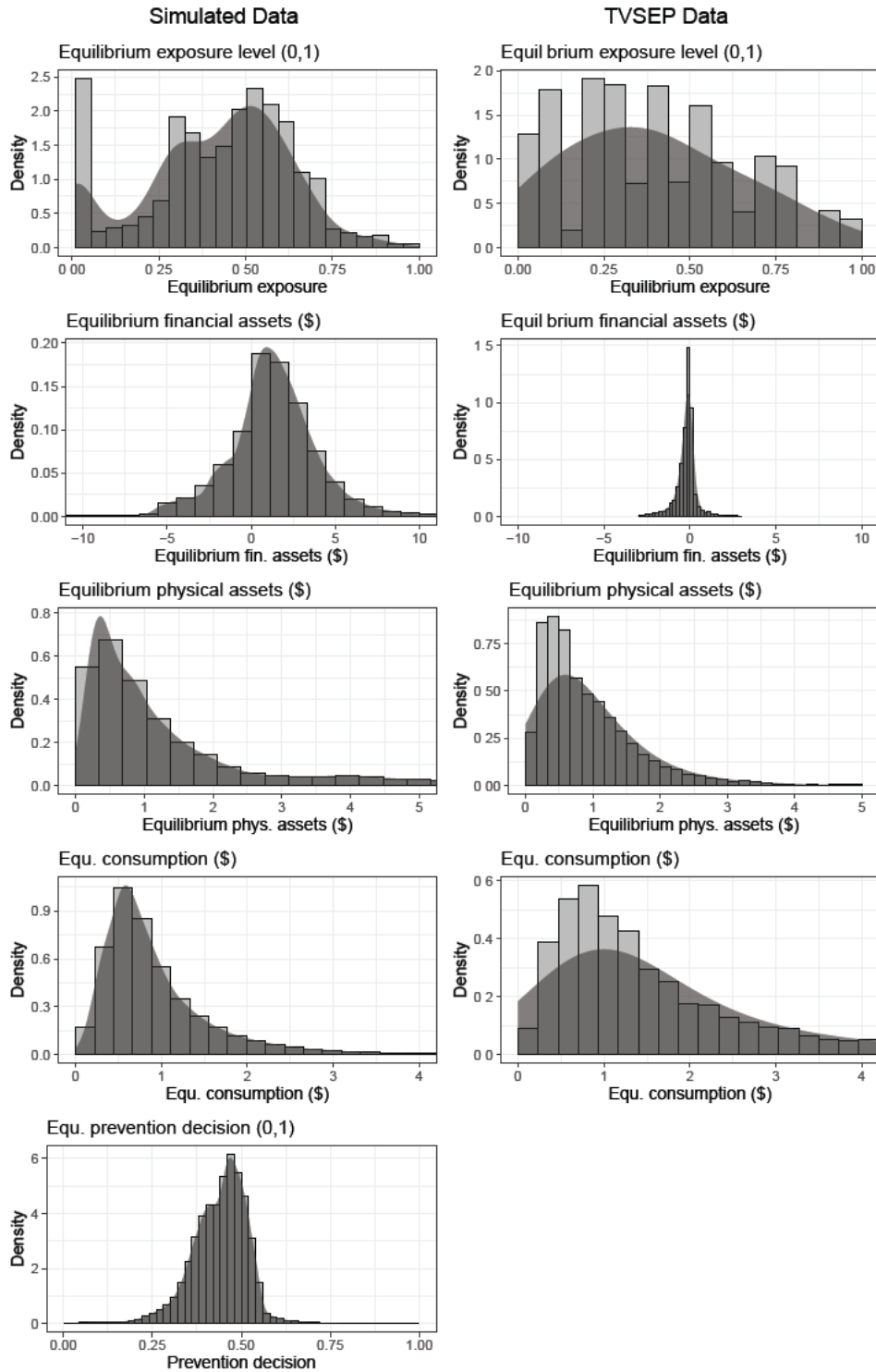


Figure 4.4: Equilibrium distribution of the main state and decision variables in the simulated and empirical data.

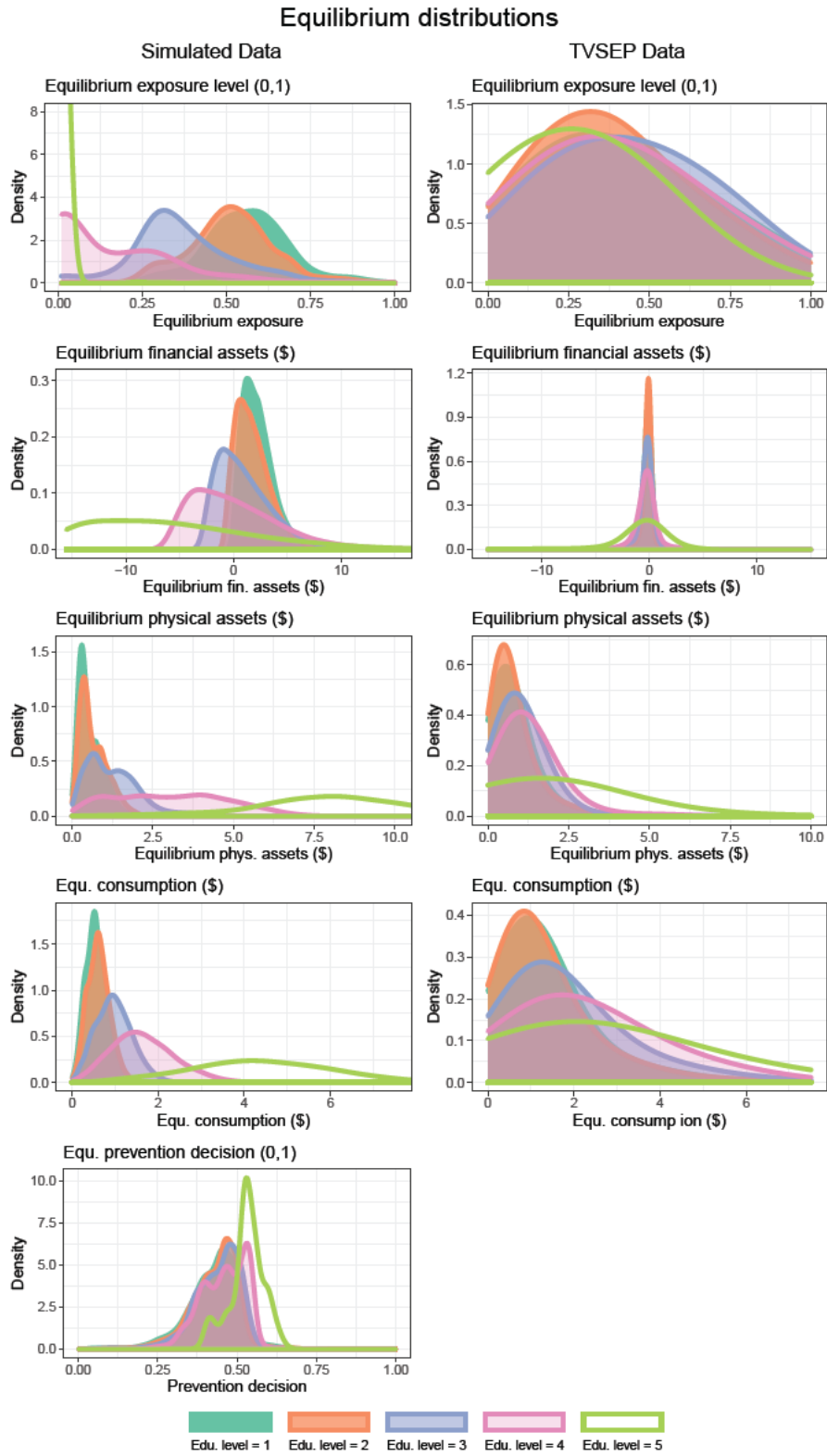


Figure 4.5: Equilibrium distributions of control and state variables by education level in simulated and empirical data.

In Table 4.2 we assess the impact of the education level on the equilibrium distributions of the simulated synthetic population using regression analyses. The first column for each variable of interest shows the coefficients for an (appropriately chosen) OLS- or logit-regression with the education classes as the independent variables. However, we also investigate the effect of the other household characteristics as shown in the second column for each variable. For the regression analyses the household characteristics are represented by categorical variables (except for income, which is treated as a continuous variable). As presented in Table 4.1 we distinguish between five education groups and three different groups for each awareness, time preference, and prevention access. Consequently, each estimated parameter covers the effect of a certain category in comparison to a household (i) with the lowest education level, (ii) low disaster risk awareness, (iii) low access to prevention and (iv) a low time preference rate. Comparing the estimated coefficients of the same education level between the two scenario (i.e. the first and the second column for each education group) allows us to identify how much of the total effect of education can be attributed to other household characteristics and which share is a direct effect of education.

Since we simulate 500,000 households, all estimated coefficients are highly statistically significance. Therefore, the statistical significance of each coefficient should be considered carefully and not be used to make specific conclusions.

Exposure

The first column in Table 4.2 shows the estimated parameters of a logit-regression for the exposure level of the simulated households. The estimated effect coefficients for the different categorical variables are quite intuitive:

- With each increasingly higher education level, the impact on the equilibrium exposure level is progressively more pronounced in absolute value implying lower exposure levels on average (compared to the base households). However, note that the parameter for the highest education class is more than twofold the one of the second highest class. This reiterates the dramatically different distribution of the highest education group shown in Figure 4.5.
- Higher awareness also leads to households being less exposed to natural hazards in equilibrium.
- Interestingly we also find, that better access to prevention actually results in households settling in more exposed areas on average. This indicates, that prevention measures being more affordable leads to households choosing less expensive, but more disaster prone settlement locations as it becomes more cost-effective to protect the physical wealth against disasters, compared to paying the premium for a settlement location in less exposed areas.
- The time discount rate also acts rather intuitive as households that discount the future to a higher extent choose more exposed settlement location, since they are rather myopic. Furthermore, the parameters for mid and high time discount rate are on very similar levels, suggesting diminishing effects for increasingly high time discount rates.
- The negative parameter for income shows that within each education group, households being in a better income situation reside in slightly less exposed areas.
- Considering the coefficients of the different education groups, accounting for the other household characteristics only marginally affects their values. This is consistent with the correlation between the other household characteristics and education and the sign of their respective effect coefficients. High awareness, low time preference and high prevention access are all positively correlated with higher education groups. Whereas the former two imply lower exposure levels for the households, the contrary holds for the latter. Hence, when including the combined effect of the three characteristics in the education effect (as shown in the first column) they partially cancel each other out. As a results the effect coefficients for the educational level hardly change when we control for other household characteristics.

	Equilibrium Values									
	Exposure Level		Financial Assets		Physical Assets		Consumption		Prevention	
	<i>glm: quasibinomial</i> <i>link = logit</i>		<i>OLS</i>		<i>OLS</i>		<i>OLS</i>		<i>glm: quasibinomial</i> <i>link = logit</i>	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Edu. Class = 2	-0.201*** (0.003)	-0.162*** (0.003)	-0.433*** (0.011)	-0.690*** (0.011)	0.123*** (0.004)	0.080*** (0.004)	0.082*** (0.002)	0.035*** (0.002)	0.029*** (0.001)	0.022*** (0.001)
Edu. Class = 3	-0.806*** (0.003)	-0.775*** (0.003)	-1.281*** (0.012)	-1.693*** (0.012)	0.718*** (0.004)	0.640*** (0.004)	0.438*** (0.002)	0.270*** (0.002)	0.064*** (0.001)	0.058*** (0.001)
Edu. Class = 4	-1.965*** (0.004)	-1.889*** (0.004)	-2.129*** (0.013)	-3.017*** (0.015)	2.504*** (0.005)	2.325*** (0.005)	1.118*** (0.002)	0.746*** (0.002)	0.142*** (0.001)	0.137*** (0.001)
Edu. Class = 5	-4.557*** (0.030)	-4.148*** (0.029)	-7.881*** (0.030)	-10.763*** (0.035)	7.431*** (0.011)	6.767*** (0.012)	3.861*** (0.005)	2.739*** (0.005)	0.428*** (0.003)	0.337*** (0.003)
Mid Awareness		-0.258*** (0.003)		0.414*** (0.012)		0.179*** (0.004)		0.047*** (0.002)		0.247*** (0.001)
High Awareness		-0.455*** (0.002)		0.703*** (0.010)		0.290*** (0.003)		0.074*** (0.002)		0.491*** (0.001)
Mid Prevention Access		0.088*** (0.007)		-0.092*** (0.027)		0.030*** (0.010)		0.033*** (0.004)		0.491*** (0.003)
High Prevention Access		0.182*** (0.006)		-0.214*** (0.026)		0.031*** (0.009)		0.044*** (0.004)		0.744*** (0.002)
Mid Time Discount Rate		0.176*** (0.003)		-1.388*** (0.012)		-0.270*** (0.004)		-0.098*** (0.002)		0.011*** (0.001)
High Time Discount Rate		0.255*** (0.004)		-1.842*** (0.014)		-0.307*** (0.005)		-0.122*** (0.002)		0.032*** (0.001)
Income		-0.047*** (0.001)		0.326*** (0.003)		0.085*** (0.001)		0.209*** (0.001)		0.002*** (0.0003)
Constant	0.263*** (0.002)	0.285*** (0.008)	2.265*** (0.009)	3.101*** (0.030)	0.549*** (0.003)	0.481*** (0.011)	0.537*** (0.002)	0.300*** (0.005)	-0.309*** (0.001)	-1.356*** (0.003)
Observations	500,000	500,000	500,000	500,000	500,000	500,000	500,000	500,000	500,000	500,000
R ²			0.158	0.213	0.637	0.649	0.622	0.718		
Adjusted R ²			0.158	0.213	0.637	0.649	0.622	0.718		

Note: *p<0.1; **p<0.05; ***p<0.01

Table 4.2: Impact of education levels, disaster risk awareness, prevention access, time discount rate and income on the main state and decision variables in the long-run equilibrium.

Financial assets

For the financial assets we use a standard OLS regression for the estimation of the impact of the different categorical variables.

- As already indicated by Figure 4.5 higher educated households on average accrue higher financial debts. Financial assets decreasing with the level of education of a household is a rather counter-intuitive result at first. However, consider that financial assets do not contribute to period utility directly (in contrast to physical assets) and serve as a way to transfer wealth into the future without potential losses in case of a disaster. As discussed before, higher educated households generally reside in less exposed areas and consequently have less incentive to keep risk-free but period-utility neutral financial assets instead of physical assets. Furthermore the generally higher income level and expectation of similar income levels in the future allows the households to invest more in physical assets and consumption without keeping financial savings. They can afford to hold fewer financial assets and go into debt (to finance physical assets), as they can always expect to have sufficient labour income to account for the basic expenditures and pay the interest on the debt in the future.
- Table 4.2 shows that this phenomenon becomes even more pronounced when accounting for the other household characteristics, which turn out to have a mitigating effect on the effect coefficient of education in the uncontrolled analysis.
- The impact of the higher discount rates intuitively implies less financial savings as households are less future oriented.
- Higher income in general allows households to accumulate more financial assets. Hence within each education group households save in times of higher income to look ahead for potential negative income shocks in the future. Meanwhile (as discussed in the first bullet point) it still holds, that the financial assets in general are lower for the higher educated households with higher average income. This represents a rather interesting duality aspect of income: While higher income in isolation implies higher levels of financial assets, the expectation of higher income in the future (among other effects) disincentivises holding financial assets.
- The positive impact of higher awareness levels on financial assets highlights that risk awareness can have a capacity building aspect. Households with higher awareness for disaster risk keep more financial assets to increase their capabilities to react and adapt to the impacts of a disaster after it has occurred. As a result they are more resilient against natural disasters being able to maintain their living standards without substantial adjustments.
- In similar fashion better access to prevention measures reduces the necessity for financial assets to some degree. Better access makes it more cost efficient to protect physical assets ex-ante against loss or damage compared to replacing them ex-post (with the usage of financial resources).

Physical assets

As for the financial assets we conduct a standard OLS regression for the estimation of the impact of the different categorical variables.

- For the physical assets we obtain the converse results compared to the financial assets. Higher education levels in general indicate higher levels of physical assets, with this effect being more moderate when controlled for the other household characteristics.
- The effect coefficients of higher awareness are positive and qualitatively similar to the case of financial assets, but on a substantially lower level (i.e. coefficients are half to one third the value in absolute terms). On the other hand the long-run perspective plays a crucial role here. As can be found in Table 4.4 in Appendix 4.8.2 higher awareness actually implies slightly lower investments

in physical assets in the short-run. However, the positive impact of awareness on the other decision and state variable (exposure reduction, higher prevention investment, ...) leads to households with higher awareness on average being less exposed and losing fewer assets in case they are hit by a disaster. Consequently, this allows them to accumulate comparatively higher levels of physical assets in the long-run.

- Better access to prevention leads to households accumulating higher levels of physical assets. Here the impact is twofold: (i) Better access to prevention reduces the costs for protection of assets and households can use parts of the savings in expenditure for further investment into physical assets. (ii) Better prevention access implies higher levels of prevention (see last column in Table 4.4) and hence fewer assets get destroyed in case of a disaster favoring asset accumulation in general. However compared to the other effect coefficients, the magnitude of this effect is rather small.
- Although it can be seen in Table 4.4 that higher time discount rates in the short-run counter-intuitively correlate positively with higher investments in physical assets, the correlation with the long-run distribution is clearly negative. Again this indicates, that the influence of higher time preference on other variables in the long-run can lead to an overcompensation of the direct negative short-term effects.
- As for the financial assets, higher income in general correlates with higher levels of physical assets, although the effect is substantially smaller for physical assets.

Consumption

The OLS-regressions for the consumption levels in equilibrium result in the following conclusions:

- As intuitively expected higher education also leads to higher consumption in equilibrium. However, in comparison to the other directly utility generating factor, i.e. the physical assets, the coefficients with respect to consumption are significantly smaller. This furthermore indicates that education has a more pronounced effect on stock variables that accumulate over time compared to flow variables like consumption.
- For the effect coefficient of the discount rate, we observe a similar result as already discussed for the financial assets. While a higher discount rate intuitively would suggest a higher preference for utility in the present and therefore could indicate higher consumption (as is confirmed in Table 4.4), the long term effects lead to a reversal of the effect. Being more future oriented (i.e. lower time discount rate) leads to households showing higher consumption levels in the long-run equilibrium.
- The effects of awareness, prevention access and income are all positive. While the impact of income on consumption is quantitatively comparable to the role of income on financial and physical assets, the coefficients of the former two characteristics are rather small in size.

Prevention

To assess the impact of the household characteristics on the prevention decision we performed a logit-regression analysis.

- Again we observe that higher educational levels lead to higher levels of prevention on average. Compared to the impact of education on exposure, however, the coefficients are rather small. This reflects the observation in the discussion of Figure 4.5. Furthermore, the coefficients of education do not substantially change, if we control for other household characteristics.
- The effect coefficient of the time discount rate and income are positive, but the coefficients are comparatively small in magnitude.²¹

²¹Again note that the statistical significance of the parameters results from the sample size of 500,000 simulated households and the significance level should not be used as an argument for the importance of an effect coefficient .

- Awareness on the other hand substantially affects the prevention level decision in equilibrium. Changing from low to high awareness on average has a more pronounced positive effect on the prevention decision than switching from the lowest to the highest education group.
- The most distinct factor for the equilibrium prevention decision is given by the level of access to prevention. Better access allows households to protect considerably higher shares of physical assets against disaster risk for the same or even lower levels of expenditures.

4.6.3 Optimal policy functions

Due to the stochastic elements in our framework, households do not stay in a fixed equilibrium of the state variables in the long-run, but follow different paths according to the realisations of the stochastic processes of labour income and disaster experience. In this section we analyse the optimal situational household strategies depending on the combination of state variables (exposure, financial and physical assets, disaster experience and income) a household faces in its present situation. Furthermore we also investigate the impact of the education level and the other household characteristics on the optimal decision making process. However, the illustration of the optimal household strategies is far from trivial itself, as finding a compact way to illustrate seven different decision variables, which depend on combinations of five different state variables and four household characteristics is obviously quite intricate. We decided to focus on the presentation of each decision variable in isolation and with respect to at most three different state variables or characteristics, while averaging over the other factors. A more in-depth regression analysis of all household decision variables with respect to their current state variables (with and without controlling for the household characteristics) is relegated to Appendix 4.8.2.

Relocation and Exposure decision

In the left panel of Figure 4.6 we illustrate the relocation decision of households depending on the exposure level they face at their current settlement location and their education level. The right panel shows the exposure decision in case the household decides to relocate depending on the same variables.

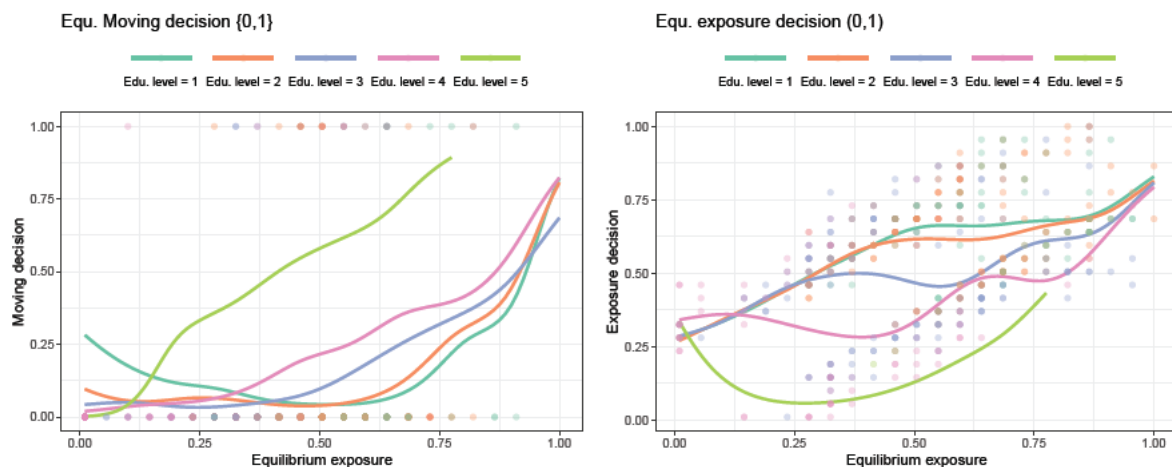


Figure 4.6: Optimal relocation and exposure decision by households depending on their educational level.

Starting with the left panel we observe, that the share of households deciding to relocate increases monotonously with the current settlement exposure level for the higher educational levels. For the lower educated groups we identify an U-shaped pattern. This reflects the fact that lower educated households can afford to keep their settlement in the least exposed areas only under the most beneficial conditions (no disaster occurrence, high income realisations, ...). Facing any kind of adversity such as losses from

a (in this area rare) disaster or a negative income shock can induce the (still rational) choice to relocate to a more exposed location. The range of exposure levels between 0.4 and 0.8 shows the most distinct differences between the different educational groups. In this range the higher educated households are substantially more likely to relocate compared to the lower educated households, which remain at the same settlement location in the majority of cases.

The right panel of Figure 4.6 shows, whether households have the tendency to choose a more or less exposed new settlement in case they decide to relocate. Starting at the areas of lower exposure, our simulations indicate that households actually choose similar (higher) exposure levels independent of their education level. However, for each education group, a threshold exists, where the optimal new location switches from more to less exposed (in comparison to the current one). The exposure level at which this switch happens is lower for higher educated households and vice versa.

Financial assets

In Figure 4.7 we plot the financial net-savings, i.e. the net-change in financial assets depending on the level of financial assets and the educational level. The figure shows, that the optimal strategies are rather



Figure 4.7: Optimal decisions for the net-investments into financial assets depending on the current status and the households educational level.

similar for the first four education groups for most positive values of current financial assets. However as financial assets turn negative, the optimal net-savings become more distinct with the lower education groups aiming for higher net-savings. Higher educated households can afford to keep net-savings low or even negative for a wider range of current asset levels, as they have higher expected labour income enabling them to repay accrued debt more easily. The income advantage of the highest educated group furthermore is so strong such that the net-savings are significantly below those of the other education groups for all current levels of financial assets.

Physical assets

For the net-investments in physical assets with respect to the current level of physical assets and education in Figure 4.8 we can identify an interesting hump-shaped pattern for the lower wealth levels followed by a shallow but monotonous decrease in the higher wealth levels. This seems rather counter-intuitive at first, but can be explained when additionally grouping the decision making by disaster experience as shown in Figure 4.9. This figure highlights two distinct investment patterns depending on whether or not a household suffered from a natural disaster in the last period. As the left panel shows, house-

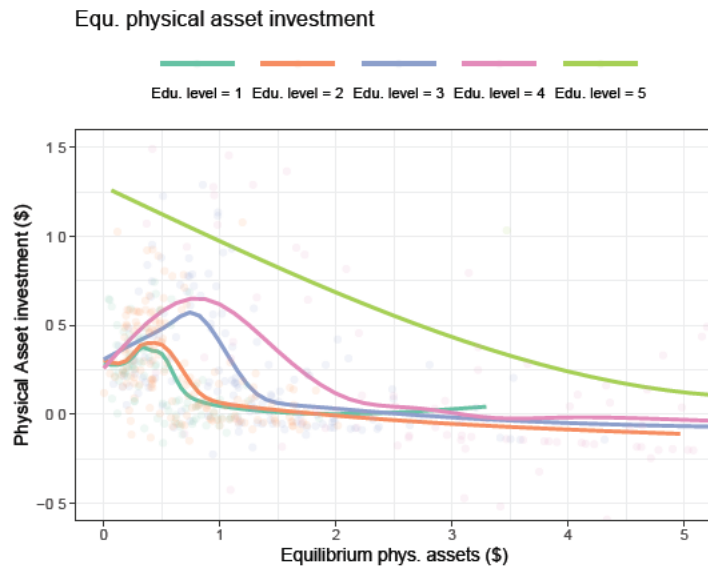


Figure 4.8: Optimal decisions for the net-investments into physical assets depending on the current status and the households educational level.

holds that were not hit by a natural disaster exhibit a similar optimal investment pattern for physical assets as for financial assets. The lower the current physical assets, the higher the investments are likely to be. However, in contrast to the financial assets, the higher educated households generally tend to invest more than the lower educated groups. On the other hand the right panel highlights that households increasingly invest in physical assets after suffering from a natural disaster. This indicates, that households sustaining higher levels of physical assets after a disaster shock also have other resources to reinvest in physical assets more extensively as a strategy to cope with the losses.²²

Equ. physical asset investment

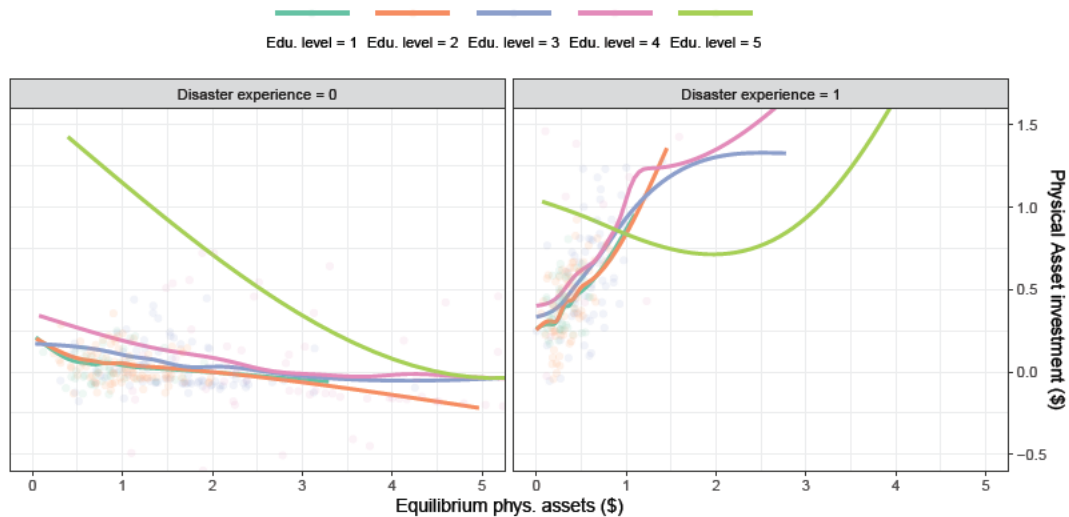


Figure 4.9: Optimal decisions for the net-investments into physical assets depending on the current status and the households educational level grouped by disaster experience.

²²The profile of the highest education group should not be used for any type of conclusion. As we have seen before, almost all households in the highest education group are located at the lowest exposed areas. Hence the number of households with disaster experience within the highest education group is negligible. We omitted to plot the confidence intervals for the curve to increase clarity of the plot, but the confidence interval for the investment curve of the highest educated group covers a range of some $\pm 1[BIU]$ around the estimated decision strategy.

4.6.4 Exposure, Vulnerability and disaster risk

In this last section we want to return to the concept of disaster risk at the household level and how it is related to the educational level of the household as presented in Figure 4.1. While the model exhibits a variable directly representing the exposure of a household to natural disasters and the hazard is an exogenous factor for the households, we need to define an indicator capturing the third argument of disaster risk, i.e. vulnerability. This value should represent the extent to which a household is affected by a natural disaster in case that a disaster happens. We decided to define vulnerability as the loss in long-term expected utility in case of disaster experience relative to the utility in the no-disaster-scenario. As expected long-term utility is captured by the value function in our framework, this definition can be represented by equation (4.30). Thereby we adjust the value function by its minimum value over the feasible region of state variables \underline{V} to ensure a positive denominator in equation (4.30).²³

$$Vul := \frac{\mathbb{E}_{\mathcal{Y}}[V(E, S, W, \mathcal{Y}, D = 0)] - \mathbb{E}_{\mathcal{Y}}[V(E, S, W, \mathcal{Y}, D = 1)]}{\mathbb{E}_{\mathcal{Y}}[V(E, S, W, \mathcal{Y}, D = 0)] - \underline{V}} \quad (4.30)$$

We find this definition to be appropriate as it incorporates several important aspects of vulnerability:

- It considers a long-term perspective compared to e.g. using the period utility function. A household temporarily reducing consumption and/or adjusting in other ways, while being able to properly return to its previous living standard after 1-2 periods of time, intuitively should be marked as less vulnerable compared to a household, which is put on a completely different trajectory by the shock for longer periods of time.
- It provides a holistic point of view on the impact of disasters. The value function is the result of all future decisions by the household and therefore includes multiple different impact channels compared to alternatively focusing on consumption or physical assets in isolation.
- The same absolute losses in consumption, physical assets, etc. should in general have less impact on wealthier households when discussing their vulnerability. Our definition accounts for this aspects in two ways. (i) We scale the difference in expected utility with the expected value function for the no disaster case. Hence, the same absolute difference in utility in the nominator results in lower vulnerability for households being well off in the first place compared to households in more precarious situations. (ii) The decreasing marginal utility of consumption and physical assets in the period utility function implies that losses in those two factors have less impact on utility if they are normally on a higher level already. Since the value function originates from the period utility, the difference in the nominator directly accounts for this wealth aspect to some degree.
- We obtain an indicator which is normalised between 0 and 1 and is independent of the units or scales used in the framework.

Applying this definition of vulnerability we define the disaster risk level R of a household as the product of exposure and vulnerability.²⁴

$$R = E \times Vul \quad (4.31)$$

Table 4.3 summarises the impact of education before and after controlling for the other household characteristics on exposure, vulnerability and disaster risk. We apply a logistic regression for each of the three dependent variables.

The columns for exposure are identical with the corresponding ones in Table 4.2 and whereas we include them here for completeness, for a related discussion we refer to Section 4.6.2. Regarding the impact of the different educational groups on vulnerability we again observe a strong negative gradient

²³Note that adjusting the value function by this term in general would not change equation (4.30), as this term would cancel out in the nominator.

²⁴Again, we omit hazard as it is an exogenous variable and would only scale risk by a certain factor for all households.

	Equilibrium Values					
	Exposure		Vulnerability		Disaster Risk	
	(1)	(2)	(3)	(4)	(5)	(6)
Edu. Class = 2	-0.201*** (0.003)	-0.162*** (0.003)	-0.140*** (0.003)	-0.070*** (0.003)	-0.221*** (0.005)	-0.109*** (0.004)
Edu. Class = 3	-0.806*** (0.003)	-0.775*** (0.003)	-0.478*** (0.003)	-0.257*** (0.003)	-0.820*** (0.006)	-0.493*** (0.005)
Edu. Class = 4	-1.965*** (0.004)	-1.889*** (0.004)	-0.778*** (0.004)	-0.320*** (0.004)	-1.783*** (0.009)	-1.111*** (0.008)
Edu. Class = 5	-4.557*** (0.030)	-4.148*** (0.029)	-1.624*** (0.017)	-0.320*** (0.014)	-5.261*** (0.159)	-3.287*** (0.125)
Mid Awareness		-0.258*** (0.003)		-0.142*** (0.003)		-0.258*** (0.004)
High Awareness		-0.455*** (0.002)		-0.262*** (0.002)		-0.471*** (0.004)
Mid Prevention Access		0.088*** (0.007)		-0.125*** (0.006)		-0.084*** (0.010)
High Prevention Access		0.182*** (0.006)		-0.198*** (0.006)		-0.117*** (0.010)
Mid Time Discount Rate		0.176*** (0.003)		0.072*** (0.003)		0.157*** (0.006)
High Time Discount Rate		0.255*** (0.004)		0.132*** (0.004)		0.251*** (0.006)
Income		-0.047*** (0.001)		-0.308*** (0.001)		-0.520*** (0.002)
Constant	0.263*** (0.002)	0.285*** (0.008)	-3.453*** (0.002)	-2.849*** (0.007)	-3.936*** (0.004)	-3.146*** (0.012)
Observations	500,000	500,000	500,000	500,000	500,000	500,000

Note:

*p<0.1; **p<0.05; ***p<0.01

Table 4.3: Impact of education levels, disaster risk awareness, prevention access, time discount rate and income on exposure, vulnerability and disaster risk.

with higher educated households showing substantially lower vulnerability. However, once we control for the other household characteristics the gradient becomes substantially smaller and especially for the education groups 3-5 we observe hardly any difference in their impact. In this regard, we find that higher awareness and income both imply considerable reductions in vulnerability (as also shown for exposure). The time discount rate also has a similar effect on vulnerability as it has for exposure. Households with less future orientation exhibit higher vulnerability on average. Finally, whereas better access to prevention leads to households residing in more exposed areas on average, it substantially reduces the vulnerability of households at the same.

The column for disaster risk now illustrates the combined effect on disaster risk. As the majority of impact factors had the same leading sign for coefficients in the exposure and vulnerability analysis, the signs of these coefficients for the disaster risk estimate are consistent with the other analyses. However, since better access to prevention measures had mixed impacts on vulnerability and exposure, this analysis shows that better access ultimately reduces the disaster risk of households. Hence, the positive effect reducing vulnerability actually overcompensates the negative effect on exposure.

4.7 Conclusions and discussion

In this paper we develop a framework for the modelling and analysis of household behaviour when households are subject to the hazard of a natural disaster. The modelling efforts are motivated through numerous empirical studies on the importance of education on different aspects of disaster risk. However, a holistic conceptual framework, which systemically incorporates and interconnects these aspects in a formal way, has been missing in the literature. The work presented in this paper aims to fill this gap.

We introduce a dynamic household model in discrete time, with households facing the hazard of natural disasters. Households can utilize several different ex-ante and ex-post strategies to mitigate the

negative impact of natural disasters on their expected utility. Potential strategies include relocating, precautionary savings, and loss reduction efforts. We find that all households use a combination of those strategies in their optimal decisions and using the first order optimality conditions and the Bellman-equations we are able to derive equations describing the optimal trade-offs between different decisions variables. However, since they either contain an intricate expectation operator or the unknown value function of the household problem, we are not able to derive further analytical insights and focus on the numerical solution.

We calibrate the parameters and functional forms of the model to empirical data from Thailand and Vietnam. Using a synthetic population of simulated households following the optimal decision rules proposed by our model, we are able to replicate the empirical distributions of various variables of our model and the correlations between them qualitatively well, while just missing quantitative aspects for some variables. While some part of the variations in outcomes and behaviour can be explained through the stochasticity of the labour income, certain household characteristics have a significant impact in the short and long-run. We find that they impact the decision making in the present due to the implied differences in the preferences and expectations (i.e. time preference and awareness) and more or less restrictive constraints (i.e. working income, financial and physical assets, prevention access). As the characteristics are consistent over time, these short-term variations in decision making under similar circumstances can lead to considerable differences in the long-run outcomes for households. Some of the key results are:

1. Higher education significantly reduces the long-term disaster risk of households by reducing their exposure as well as their vulnerability. However, while education remains the main driving factor for exposure when controlled for other characteristics, the positive impact of education on vulnerability can be in large part explained through effects resulting from awareness, prevention access, time preference and income.
2. Some of the observed short-run effects of household characteristics are overcompensated in the long-run through developments across other household variables within our multi-faceted framework. E.g. higher awareness correlates with lower investments in physical assets short-term, whereas in the long-run households with higher awareness are able to accumulate higher levels of physical assets.
3. For the long-term distributions of most variables the aggregated positive impact of education is attenuated to different degrees when controlling for the other household properties (as they are often positively correlated with education). For financial assets, on the other hand, the impact of education becomes even more pronounced.

Overall we can attest higher education consistently and inherently positively affects both, the short-term behaviour and long-run outcomes with respect to natural hazards. Although our model contains a multitude of decision and state variables it is still limited in its ability to replicate some aspects of reality due to some short-comings in the framework. First of all, the model still only represents a partial equilibrium, as our framework features no production sector, which would allow us to endogenously define wages and interest rates. In the same regard, we have no direct interaction between households in the model. Furthermore we assume, that the infinitely living households exhibit an exogenously given education level. An alternative approach for a future adaptation of the model would contain the introduction of a life-cycle model for households making endogenous education decisions over their finite life-time. We also abstract from reality by using a rather simplified functional form describing the relationship between exposure level and living costs. Similarly, the different levels of access to prevention measures enter the prevention cost function in a rather stylistic way, which could warrant further consideration.

On the other hand the model has substantial potential for extension and other future work. Since the model considers endogenous decision making by households resulting from a desire for utility maximization, it is significantly more capable of providing unbiased predictions and estimations in the evaluation

process of policy interventions. Whereas empirical correlations can lose their validity under systemic changes (e.g. introduction of new insurance schemes, construction of new community wide disaster protection measures), our proposed model is based on households subjectively maximizing their expected utility and therefore endogenously adjusting to systemic changes. This not only makes predictions from our model more convincing, but would also allow deeper analysis of the effects of different policy measures across different socio-economic groups and for different time horizons. Future work of the authors intend investigate the effectiveness of different policy interventions for calibrated case studies.

4.8 Appendix

4.8.1 Analytical Results

In this section we present the proof for Proposition 4 and an additional Proposition 5.

Proof for Proposition 4

For the proof of Proposition 4 we first set up the Lagrange-function

$$\begin{aligned} \mathcal{L} = \mathbb{E}_{t=0} \left\{ \sum_{t=1}^{\infty} \left(\frac{1}{1+\rho} \right)^t u(c_t, W_{t+1}) + \lambda_t^S \left[-S_{t+1} + y_t(1 - \Delta^y D_t) + (1+r_t)S_t - c_t - p^w(w_t) - \right. \right. \\ \left. - p^P(E_{t+1}, W_{t+1}, P_{t+1}) - p^E(E_{t+1}) \right] + \lambda_t^W \left[-W_{t+1} + (1-\delta)(1 - \Delta^W I_t)(1 - (1-P_t)D_t)W_t + w_t \right] + \\ \left. \lambda_t^I (E_{t+1} - E_t)(1 - I_t) \right\} \end{aligned} \quad (4.32)$$

For the optimal decision made at the beginning of time period t , i.e. $(c_t, w_t, S_{t+1}, W_{t+1}, P_{t+1}, E_{t+1}, I_t)$, the household has access to the full information on the state of all current state variables and previous decisions made, i.e. $(E_t, S_t, W_t, P_t, D_t, y_t)$.

Hence taking the derivative of the Lagrange-function with respect to the decisions variables leads to the first-order optimality conditions.

$$\frac{\partial \mathcal{L}}{\partial c_t} = \mathbb{E}_t \left\{ \left(\frac{1}{1+\rho} \right)^t u_c(c_t, W_{t+1}) - \lambda_t^S \right\} = 0 \quad \implies \quad u_c(c_t, W_{t+1}) \left(\frac{1}{1+\rho} \right)^t = \lambda_t^S \quad (4.33)$$

$$\frac{\partial \mathcal{L}}{\partial w_t} = \mathbb{E}_t \left\{ -\lambda_t^S \frac{dp^w}{dw}(w_t) + \lambda_t^W \right\} = 0 \quad \implies \quad \lambda_t^S \frac{dp^w}{dw}(w_t) = \lambda_t^W \quad (4.34)$$

$$\frac{\partial \mathcal{L}}{\partial S_{t+1}} = \mathbb{E}_t \left\{ (1+r_t)\lambda_{t+1}^S - \lambda_t^S \right\} = 0 \quad \implies \quad \lambda_t^S = (1+r_t)\mathbb{E}_t \left\{ \lambda_{t+1}^S \right\} \quad (4.35)$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial W_{t+1}} = \mathbb{E}_t \left\{ \left(\frac{1}{1+\rho} \right)^t u_W(c_t, W_{t+1}) - \lambda_t^S \cdot p_W^P(E_{t+1}, W_{t+1}, P_{t+1}) - \right. \\ \left. - \lambda_t^W + \lambda_{t+1}^W (1-\delta)(1 - \Delta^W I_{t+1})(1 - (1-P_{t+1})D_{t+1}) \right\} = 0 \end{aligned} \quad (4.36)$$

$$\frac{\partial \mathcal{L}}{\partial P_{t+1}} = \mathbb{E}_t \left\{ -\lambda_t^S \cdot p_P^P(E_{t+1}, W_{t+1}, P_{t+1}) + \lambda_{t+1}^W (1-\delta)(1 - \Delta^W I_{t+1})D_{t+1}W_{t+1} \right\} = 0 \quad (4.37)$$

Combining equations (4.33) and (4.35) leads to

$$u_c(c_t, W_{t+1}) \left(\frac{1}{1+\rho} \right)^t = (1+r_t)\mathbb{E} \left\{ u_c(c_{t+1}, W_{t+2}) \left(\frac{1}{1+\rho} \right)^{t+1} \right\}$$

$$u_c(c_t, W_{t+1}) = \frac{1+r_t}{1+\rho} \mathbb{E} \{u_c(c_{t+1}, W_{t+2})\}$$

and consequently proof the consumption Euler equation (4.16). Equation (4.34) provides a direct relationship between the shadow-prices of physical and financial assets. Using this equality to substitute λ_{t+1}^W in equation (4.37) implies

$$\lambda_t^S \cdot p_P^P(E_{t+1}, W_{t+1}, P_{t+1}) = \mathbb{E}_t \left\{ \lambda_{t+1}^S \frac{dp^w}{dw}(w_{t+1})(1-\delta)(1-\Delta^W I_{t+1}) D_{t+1} W_{t+1} \right\}.$$

Using again equation (4.33) we can reformulate the previous equation to

$$u_c(c_t, W_{t+1}) p_P^P(E_{t+1}, W_{t+1}, P_{t+1}) = \frac{1-\delta}{1+\rho} \mathbb{E}_t \left\{ u_c(c_{t+1}, W_{t+2}) \frac{dp^w}{dw}(w_{t+1})(1-\Delta^W I_{t+1}) W_{t+1} D_{t+1} \right\}.$$

This equation directly corresponds to equation (4.17) in Proposition 4. The derivations for the FOC (4.18) follow analogously from combinations of equation (4.33), (4.34) and (4.36).

For the proof of FOC (4.19) we use (4.18) in the following simplified form.

$$\begin{aligned} u_c(c_t, W_{t+1}) &= \frac{u_W(c_t, W_{t+1})}{g(t)} + \left(\frac{1-\delta}{1+\rho} \right) \frac{\mathbb{E}_t \{u_c(c_{t+1}, W_{t+2}) f(t+1)\}}{g(t)} \\ g(t) &:= \left[p_W^P(E_{t+1}, W_{t+1}, P_{t+1}) + \frac{dp^w}{dw}(w_t) \right] \\ f(t) &:= \frac{dp^w}{dw}(w_t)(1-\Delta^W I_t)(1-(1-P_t)D_t) \end{aligned} \quad (4.38)$$

We can now substitute the term $u_c(c_{t+1}, W_{t+2})$ on the right hand side of (4.38) by using (4.38) for the time period $t+1$. This results in

$$u_c(c_t, W_{t+1}) = \frac{u_W(t)}{g(t)} + \frac{1-\delta}{1+\rho} \mathbb{E}_t \left\{ \frac{u_W(t+1) + \frac{1-\delta}{1+\rho} \mathbb{E}_{t+1} \{u_c(t+2) f(t+2)\}}{g(t+1)} \frac{f(t+1)}{g(t)} \right\} \quad (4.39)$$

As the information set at time $t+1$ is a subset of the information set at t , this equation can be simplified to

$$u_c(c_t, W_{t+1}) = \frac{u_W(t)}{g(t)} + \frac{1-\delta}{1+\rho} \mathbb{E}_t \left\{ \frac{u_W(t+1) + \frac{1-\delta}{1+\rho} \mathbb{E}_{t+1} \{u_c(t+2) f(t+2)\}}{g(t+1)} \frac{f(t+1)}{g(t)} \right\} \quad (4.40)$$

$$u_c(c_t, W_{t+1}) = \frac{u_W(t)}{g(t)} + \mathbb{E}_t \left\{ \left(\frac{1-\delta}{1+\rho} \right) \frac{u_W(t+1)}{g(t)} \frac{f(t+1)}{g(t+1)} + \left(\frac{1-\delta}{1+\rho} \right)^2 u_c(t+2) \frac{f(t+1)}{g(t+1)} \frac{f(t+2)}{g(t)} \right\} \quad (4.41)$$

Taking a next step and replacing $u_c(t+2)$ in the last term makes the systematic structure finally even more apparent.

$$u_c(c_t, W_{t+1}) = \frac{u_W(t)}{g(t)} + \mathbb{E}_t \left\{ \left(\frac{1-\delta}{1+\rho} \right) \frac{u_W(t+1)}{g(t)} \frac{f(t+1)}{g(t+1)} + \right. \quad (4.42)$$

$$\left. + \left(\frac{1-\delta}{1+\rho} \right)^2 \left(\frac{u_W(t+2)}{g(t+2)} + \left(\frac{1-\delta}{1+\rho} \right) \frac{\mathbb{E}_{t+2} \{u_c(t+3) f(t+3)\}}{g(t+2)} \right) \frac{f(t+1)}{g(t+1)} \frac{f(t+2)}{g(t)} \right\} \quad (4.43)$$

$$u_c(c_t, W_{t+1}) = \frac{u_W(t)}{g(t)} + \mathbb{E}_t \left\{ \left(\frac{1-\delta}{1+\rho} \right) \frac{u_W(t+1)}{g(t)} \frac{f(t+1)}{g(t+1)} + \right. \quad (4.44)$$

$$+ \left. \left(\frac{1-\delta}{1+\rho} \right)^2 \frac{u_W(t+2) f(t+1) f(t+2)}{g(t) g(t+1) g(t+2)} + \left(\frac{1-\delta}{1+\rho} \right)^3 \frac{u_c(t+3) f(t+3) f(t+1) f(t+2)}{g(t) g(t+1) g(t+2)} \right\} \quad (4.45)$$

Using induction the validity of the FOC (4.19) in Proposition 4 can be easily shown. \blacksquare

First order conditions for Bellman formulation

Proposition 5 (First order conditions (II))

Assume that the value function $V(E, S, W, Y, D)$ is continuous differentiable in S and W . Further assume that the optimal solution $(E^*, S^*, W^*, P^*, I^*, c^*, w^*)$ is an interior solution in S^* , W^* , c^* and P^* . Then the optimal solution has to fulfil the first order optimality conditions (4.46), (4.47), and (4.48).

$$\begin{aligned} u_c(c, W) &= \frac{1}{1+\rho} \mathbb{E}_{\mathcal{D}, \mathcal{Y}} \left\{ \frac{\partial V}{\partial S} \right\} \\ &= \frac{1}{1+\rho} \left[aEH \cdot \mathbb{E}_{\mathcal{Y}} \frac{\partial V}{\partial S}(E, S, W \cdot P, \mathcal{Y}, D=1) + (1-aEH) \cdot \mathbb{E}_{\mathcal{Y}} \frac{\partial V}{\partial S}(E, S, W, \mathcal{Y}, D=0) \right] \end{aligned} \quad (4.46)$$

$$\begin{aligned} u_c(c, W) \cdot \left[\frac{\partial p^w}{\partial w}(w) + \frac{\partial p^P}{\partial W}(E, W, P) \right] &= u_W(c, W) + \\ &+ \frac{1}{1+\rho} \left[aEH \cdot P \cdot \mathbb{E}_{\mathcal{Y}} \frac{\partial V}{\partial W}(E, S, W \cdot P, \mathcal{Y}, 1) + (1-aEH) \cdot \mathbb{E}_{\mathcal{Y}} \frac{\partial V}{\partial W}(E, S, W, \mathcal{Y}, 0) \right] \\ &= u_W(c, W) + \frac{1}{1+\rho} \left[\mathbb{E}_{\mathcal{D}, \mathcal{Y}} \left\{ \frac{\partial V}{\partial W} \right\} - aEH(1-P) \cdot \mathbb{E}_{\mathcal{Y}} \left\{ \frac{\partial V}{\partial W}(E, S, W \cdot P, \mathcal{Y}, 1) \right\} \right] \end{aligned} \quad (4.47)$$

$$u_c(c, W) \cdot \frac{\partial p^P}{\partial P}(E, W, P) = \frac{1}{1+\rho} aEH \cdot W \cdot \mathbb{E}_{\mathcal{Y}} \left\{ \frac{\partial V}{\partial W}(E, S, W \cdot P, \mathcal{Y}, 1) \right\} \quad (4.48)$$

For the proof we set up the Lagrange function \mathcal{L} .

$$\mathcal{L} = u(c, W) + \frac{1}{1+\rho} \left[a_t E_t H_t \cdot \mathbb{E}_{\mathcal{Y}} V(E_t, S_t, W_t \cdot P_t, \mathcal{Y}, D_t=1) + (1-a_t E_t H_t) \cdot \mathbb{E}_{\mathcal{Y}} V(E_t, S_t, W_t, \mathcal{Y}, D_t=0) \right] + \quad (4.49)$$

$$+ \lambda_t^S \left(-S_t + y_t \cdot (1 - \Delta^y D_{t-1}) + (1+r_t) S_{t-1} - c_t - p^w(w_t) - p^P(E_t, W_t, P_t) - p^E(E_t) \right) + \quad (4.50)$$

$$+ \lambda_t^W \left(-W_t + (1-\delta)(1-\Delta^W I_t) W_{t-1} + w_t \right) + \lambda_t^I \left(E_t - E_{t-1} \right) (1 - I_t) \quad (4.51)$$

Taking the derivative of the Lagrange function with respect to the decision variables c, W

$$\frac{d\mathcal{L}}{dc} = u_c - \lambda_S \quad (4.52)$$

$$\frac{d\mathcal{L}}{dW} = u_W + \frac{1}{1+\rho} \left[a_t E_t H_t \cdot \mathbb{E}_{\mathcal{Y}} \frac{dV}{dW}(E, S, W \cdot P_t, \mathcal{Y}, 1) \cdot P_t + (1-a_t E_t H_t) \cdot \mathbb{E}_{\mathcal{Y}} \frac{dV}{dW}(E, S, W, \mathcal{Y}, 0) \right] + \quad (4.53)$$

$$+ \lambda_S \left(-p_W^P(E, W, P) \right) + \lambda_W \cdot (-1) \quad (4.54)$$

$$\frac{d\mathcal{L}}{dS} = \frac{1}{1+\rho} \left[a_t E_t H_t \cdot \mathbb{E}_{\mathcal{Y}} \frac{dV}{dS}(E_t, S_t, W_t \cdot P_t, \mathcal{Y}, 1) + (1-a_t E_t H_t) \cdot \mathbb{E}_{\mathcal{Y}} \frac{dV}{dS}(E_t, S_t, W_t, \mathcal{Y}, 0) \right] + \lambda_S \cdot (-1) \quad (4.55)$$

$$\frac{d\mathcal{L}}{dw} = -\lambda_S \cdot \frac{d}{dw} [p^w(w)] + \lambda_W \quad (4.56)$$

Combining these equations to eliminate the shadow prices directly results in the equations presented in the Proposition. \blacksquare

Equations (4.46), (4.47), and (4.48) nicely illustrate some of the trade-offs the households have to consider in their decision making.

- (4.46) shows that for the optimal behaviour the marginal utility of consumption is equal to the expected marginal benefits of a marginal unit of additional financial assets. These benefits consist of the expected (with respect to stochastic income) marginal change in the value-function for the disaster and no-disaster scenarios, which are in turn weighted by the subjective probabilities of occurrence. Finally they are discounted at rate ρ as they only materialise in the next time period (compared to consumption).
- (4.47) illustrates the trade-offs between consumption and physical assets. The left hand side contains the marginal utility gains if one marginal unit of physical assets is used for consumption instead (the marginal utility of consumption multiplied with the marginal cost savings of less physical assets. On the right hand side, we see that the benefits of physical assets are twofold in contrast to the benefits of consumption and financial assets. While consumption only implies utility gains through the period utility and financial assets only affect future utility through the value function, physical assets contribute through both channels. However, since they can get destroyed in case of a natural disaster (if not protected), the expected value of the derivative of the value function is smaller (in the disaster case the marginal gains are scaled with the prevention level P).
- (4.48) similarly relates the marginal utility gains through consumption (resulting from decreased preventive efforts) to the expected gains from higher prevention through the discounted value function in the next period. Since prevention efforts only become relevant in case of disaster occurrence, the benefits are limited to this case. Also as the marginal prevention costs, also the benefits are more pronounced for higher levels of physical assets due to the multiplicative nature within the value function.

4.8.2 Numerical solution

Model transformation

While the introduction of the indicator decision variable I_t allowed us to formulate the problem in a very compact form for the numerical solution a different approach is more efficient. We formulate the decision whether to relocate or not similar to an option value approach splitting the optimisation problem (4.20)-(4.23) into two separate problems.

Solving the problem (4.20)-(4.23) at each point in time is equivalent to the household solving two optimisation problems (P1) and (P2) at each point in time.

- (P1) The first problem is to maximize the expected utility (4.57) conditional on the household deciding to relocate. This only directly affects the wealth accumulation equation as can be seen in (4.59).

$$V^1(S_{t-1}, W_{t-1}, y_t, D_{t-1}) = \max_{\substack{E_t, S_t, W_t, \\ P_t, c_t, w_t}} \left\{ u(c_t, W_t) + \frac{1}{1 + \rho} \left[a_t E_t H_t \times \right. \right. \\ \left. \left. \times \mathbb{E}_Y V(E_t, S_t, W_t \cdot P_t, \mathcal{Y}, D_t = 1) + (1 - a_t E_t H_t) \cdot \mathbb{E}_Y V(E_t, S_t, W_t, \mathcal{Y}, D_t = 0) \right] \right\} \quad (4.57)$$

$$S_t = y_t \cdot (1 - \Delta^y D_{t-1}) + (1 + r_t) S_{t-1} - c_t - p^w(w_t) - p^P(E_t, W_t, P_t) - p^E(E_t) \quad (4.58)$$

$$W_t = (1 - \delta)(1 - \Delta^W) W_{t-1} + w_t \quad (4.59)$$

It quickly becomes apparent, that the optimal decisions do not depend on the initial exposure level E_{t-1} any more. However this optimisation still can be done for every combination of (S, W, y, D) and the optimal objective values are captured in the function V_1 .

- (P2) The second problem is to maximize the expected utility (4.60) conditional on the household staying at the same settlement location. Again, this mainly affects the wealth accumulation equation

(see (4.62)). There is no decision variable E_t anymore in this case and E_t gets replaced by E_{t-1} in all equation (4.60)-(4.62).

$$V_2(E_{t-1}, S_{t-1}, W_{t-1}, y_t, D_{t-1}) = \max_{\substack{S_t, W_t, \\ P_t, c_t, w_t}} \left\{ u(c_t, W_t) + \frac{1}{1+\rho} \left[a_t E_{t-1} H_t \times \right. \right. \\ \left. \left. \times \mathbb{E}_{\mathcal{Y}} V(E_{t-1}, S_t, W_t \cdot P_t, \mathcal{Y}, D_t = 1) + (1 - a_t E_{t-1} H_t) \cdot \mathbb{E}_{\mathcal{Y}} V(E_{t-1}, S_t, W_t, \mathcal{Y}, D_t = 0) \right] \right\} \quad (4.60)$$

$$S_t = y_t \cdot (1 - \Delta^y D_{t-1}) + (1 + r_t) S_{t-1} - c_t - p^w(w_t) - p^P(E_{t-1}, W_t, P_t) - p^E(E_{t-1}) \quad (4.61)$$

$$W_t = (1 - \delta) W_{t-1} + w_t \quad (4.62)$$

Similar to the first case we store the maximized objective value for each possible combination of (E, S, W, y, D) in function V_2 .

After calculating the optimal objective value for both scenarios the households bases its relocation decision on which scenario yields a higher expected utility. For the value function it consequently holds that

$$V(E, S, W, y, D) = \max \left\{ V_1(S, W, y, D), V_2(E, S, W, y, D) \right\}.$$

This problem transformation allows us to use a value function iteration approach to find the unknown value function $V(\cdot)$ and the optimal policy function for all decision variables.

Numerical methods

To derive an approximation for the unknown value function and optimal policy function, in a first step we discretize the state space. However, as our model consists of five state and seven decision variables, we face the curse of dimensionality when applying a value function iteration strategy over the discrete grid of state variables.²⁵ To increase computational efficiency and reduce the overall computation time, we use a combination of several strategies:

- When solving the maximization problem on the right hand side of (4.20) we use a coordinate ascend method to reduce the number of functional evaluations compared to a full grid search over all possible combination of decision variables. Since the restriction in search directions for the coordinate ascend method, we check all surrounding neighbors of a fix point of the algorithm to ensure local optimality.

Furthermore we conduct a complete search over a more extended neighborhood once the value function has converged. This allows us to escape local optima and we continue the value-function-iteration process until all policies cannot be improved within the extended neighborhood.

- Between iterations conducting an optimisation as described above, we apply Howards-Policy-Iteration-Strategy to increase the speed of convergence and reduce computation times.
- After convergence of the value function over a discrete approximation of the control/state space, we continue with iterations over the continuous control space until convergence²⁶ using a linear interpolation of the value function.

²⁵Doubling the number of grid points for exposure, financial and physical assets would imply, that we not only have to conduct 8-times the number of optimisation problems, but also that for a complete grid search the number of combination of state combinations to check for optimality would increase eight fold. In total that would roughly increase the computational efforts 64-fold .

²⁶Due to numerical issues we kept the exposure level decision over a discrete control space.

- Finally we make use of GPU-accelerated computation methods to conduct the steps described above for thousands of combinations of state variables in parallel.

After having derived the optimal policy functions of the households, we use Monte-Carlo-Simulations to obtain the equilibrium distributions of households across all state variables. Thereby we again use GPU-accelerated methods to decrease computations times.

Impact of state variables and household characteristics on decision making

	Dependent variable:											
	Relocation		Exposure Decision		Fin. Net-Savings		Phys. Ass. invest.		Consumption		Prevention level	
	<i>logistic</i>	<i>glm: quasinomial link = logit</i>	<i>glm: quasinomial link = logit</i>	<i>OLS</i>	<i>OLS</i>	<i>OLS</i>	<i>OLS</i>	<i>OLS</i>	<i>OLS</i>	<i>glm: quasinomial link = logit</i>	<i>glm: quasinomial link = logit</i>	
Equ. Exposure	-0.690*** (0.057)	0.183*** (0.064)	3.151*** (0.004)	3.153*** (0.005)	1.149*** (0.004)	0.914*** (0.003)	-0.348*** (0.002)	-0.302*** (0.002)	0.081*** (0.002)	0.111*** (0.001)	-0.175*** (0.003)	0.152*** (0.003)
Equ. Fin. Assets	-0.198*** (0.003)	-0.042*** (0.007)	0.003*** (0.0002)	0.024*** (0.0004)	0.024*** (0.0002)	-0.121*** (0.0002)	0.026*** (0.0001)	0.085*** (0.0002)	0.034*** (0.0001)	0.125*** (0.0001)	-0.006*** (0.0002)	-0.001*** (0.0002)
Equ. Phys. Assets	-1.634*** (0.031)	-2.024*** (0.041)	-0.442*** (0.001)	-0.507*** (0.001)	-0.283*** (0.001)	0.161*** (0.001)	-0.142*** (0.0003)	-0.327*** (0.0005)	0.385*** (0.0003)	0.096*** (0.0003)	0.036*** (0.001)	0.031*** (0.001)
Equ. Income	0.305*** (0.006)	0.267*** (0.006)	-0.072*** (0.0004)	-0.080*** (0.0004)	0.590*** (0.0004)	0.638*** (0.0003)	0.177*** (0.0002)	0.158*** (0.0002)	0.190*** (0.0002)	0.190*** (0.0001)	0.160*** (0.0001)	0.001*** (0.0003)
Equ. Dis. Exp.	1.215*** (0.023)	0.987*** (0.027)	-0.213*** (0.001)	-0.254*** (0.001)	-0.769*** (0.001)	-0.483*** (0.001)	0.381*** (0.001)	0.261*** (0.001)	0.135*** (0.001)	0.135*** (0.001)	-0.051*** (0.0004)	0.054*** (0.001)
Edu. Class = 2		0.117*** (0.018)		0.023*** (0.001)		-0.177*** (0.001)		0.065*** (0.001)		0.116*** (0.0004)		0.027*** (0.001)
Edu. Class = 3		0.687*** (0.030)		0.128*** (0.002)		-0.683*** (0.001)		0.246*** (0.001)		0.431*** (0.001)		0.076*** (0.001)
Edu. Class = 4		1.553*** (0.056)		0.168*** (0.004)		-1.596*** (0.003)		0.633*** (0.002)		0.923*** (0.001)		0.143*** (0.003)
Edu. Class = 5		2.474*** (0.241)		0.573*** (0.017)		-5.258*** (0.008)		2.123*** (0.005)		3.463*** (0.003)		0.213*** (0.008)
Mid Awareness		0.039** (0.018)		-0.059*** (0.001)		0.045*** (0.001)		-0.039*** (0.001)		-0.019*** (0.0004)		0.254*** (0.001)
High Awareness		0.094*** (0.015)		-0.102*** (0.001)		0.082*** (0.001)		-0.075*** (0.001)		-0.036*** (0.0003)		0.503*** (0.001)
Mid Prev. Acc.		-0.505*** (0.035)		0.064*** (0.003)		-0.029*** (0.002)		0.007*** (0.001)		0.040*** (0.001)		0.485*** (0.001)
High Prev. Acc.		-0.673*** (0.033)		0.119*** (0.003)		-0.047*** (0.002)		0.017*** (0.001)		0.065*** (0.001)		0.735*** (0.002)
Mid Disc. Rate		0.068*** (0.023)		0.039*** (0.001)		-0.133*** (0.001)		0.044*** (0.001)		0.100*** (0.0004)		0.011*** (0.001)
High Disc. Rate		0.138*** (0.025)		0.052*** (0.002)		-0.188*** (0.001)		0.064*** (0.001)		0.134*** (0.0005)		0.029*** (0.001)
Constant	-2.354*** (0.055)	-2.345*** (0.069)	-1.039*** (0.003)	-1.111*** (0.005)	-0.897*** (0.002)	-0.690*** (0.003)	0.047*** (0.001)	0.045*** (0.002)	-0.055*** (0.001)	-0.167*** (0.001)	-0.240*** (0.002)	-1.484*** (0.003)
Observations	500,000	500,000	500,000	500,000	500,000	500,000	500,000	500,000	500,000	500,000	500,000	500,000
R ²					0.843	0.921	0.760	0.832	0.951	0.988		
Adjusted R ²					0.843	0.921	0.760	0.832	0.951	0.988		
Akaike Inf. Crit.	198,256.700	196,807.300										

Note: *p<0.1; **p<0.05; ***p<0.01

Table 4.4: Impact of the state variables in equilibrium, education levels, disaster risk awareness, prevention access, time discount rate and income on the decision making with respect to relocation, exposure, financial savings, physical asset accumulation, consumption and prevention.

Chapter 5

Conclusions

In this final chapter overall findings and conclusions of this thesis are presented. We recap on the key findings of each model within the overarching context and with a discussion on future related work and potential extensions of each framework.

5.1 Overall conclusion

This thesis explores the optimal behaviour at the individual or macro level with respect to risk in various settings. New perspectives on already established model frameworks allow for additional and deeper insights into the optimal strategies. In all investigated settings we find combinations of ex-ante preventive and ex-post adaptation strategies aiming to lessen the impacts of risk.

In Chapter 2 we extend a standard life-cycle model of modelling health decisions by a stochastic health shock, which can put the whole life-course on a different trajectory. Through this extension we were able to substantially expand the theory on the value of life to more specific and differentiated health aspects of an individual's life (general health and its preventive and survival effects, acute and long-run effects of a specific disease). Our framework allows for distinction between the value of health, value of prevention, value of acute survival and value of morbidity, which each describe the willingness to pay for improvements in a part of general health. Furthermore we found that these valuations are the defining factors for the timing and the trade-offs between different health investments.

By finding an analytic decomposition for each term, we identify the forces that determine preventive behaviour before the health shock and adaptive strategies after a shock has occurred. Although the different terms are hard to sort with respect to their magnitude in general, our numerical example allowed us to pin down how strong general health investments are driven by the shock-preventing aspects of good health. We find that until the age of 70 health expenditure is shifted towards younger ages through this preventive effect, whereas the indirect incentive to have better health after the shock works in the opposite direction, but is insignificant in magnitude. For consumption we find that a discrepancy between the mortality rate and the returns to annuities incentives a postponement of consumption to older ages for cancer-free individuals. The higher marginal utility of consumption in the cancer-free state implies advancements of consumption and partially off-sets the first effect. Overall we obtain a postponement of consumption over the life-cycle compared to a complete risk-free setting.

On the other hand adaptive behaviour after a cancer diagnosis is swift and pronounced. Extensive funds are invested in chronic care in the first years to reduce the disease stock at a fast pace. This implies that young individuals go significantly stronger into debt compared to their cancer-free counterparts. Older individuals similarly dissolve their accumulated assets considerably earlier, if they are diagnosed with cancer at later stages in life. Also the consumption path significantly adjusts after the diagnosis while also being sensitive to the age at diagnosis. Whereas the increased mortality risk leads to an advancement of consumption close to the time of diagnosis, the negative impact of the cancer stock on utility incentivises deferring consumption to later stages after the impact of cancer has been reduced. For diagnoses early in life we consequently find a U-shaped consumption-profile (always below

the comparable cancer free level), meanwhile in older ages the mortality risk is the dominating effect throughout and consumption at the time of diagnosis jumps upward.

In Chapter 3 the distinction between preventive and adaptive strategies is less clear-cut, as we do not consider a regime changing shift explicitly. However, the network-structure still lets us identify preventive and adaptive aspects of the optimal target transmission rates and testing strategy for the containment of an infectious disease. On the one hand lockdown measures can be seen as a preventive strategy to reduce new infections, on the other hand they can also be a reaction to a spike in infection numbers and therefore interpreted as an adaptive intervention. In a standard setting, these two considerations are hard to disentangle. However, in our numerical example with an initial hotspot of infections in one region, the immediate reduction in the target transmission within this region can intuitively be seen as a reaction to this initial shock of high infections. Then again, the complete cut-down of transmission of group 1 on the other two to the lowest possible level intuitively resembles more the preventive aspect of lockdowns. This measure is primarily geared to avoid infections in groups 2 and 3 as far as possible.

Testing also incorporates both aspects of risk behaviour. On the adaptive side testing can be a reactive measure to reduce the number of infected individuals from the active population. Through the preventive perspective testing at the same time reduces the probability of new infections within the susceptible population. Furthermore in case of early treatment possibilities, testing can identify infected people and allows them to be treated before they can potentially escalate to a heavy case (another preventive aspect). In the numerical simulations we find that testing becoming available reduces the lockdown measures, while still reducing the aggregate burden on society. This indicates that testing and lockdowns not only both contain aspects of preventive and adaptive strategies, but can also be substituted by each other.

In the third model presented in Chapter 4 we can again observe explicit preventive and adaptive behaviour with respect to the risk of natural disasters. Even more explicitly we can make the distinction between loss prevention and loss reduction strategies. The household changing the location of its settlement effects the probability of being affected by a natural disaster (loss prevention), whereas explicit investments in insurance or other disaster protection methods reduce the damages to physical assets in case a disaster strikes (loss reduction). Considering adaptation, we also derive the optimal reaction in case a household suffered losses from a natural disaster in form of physical assets or future income. This reaction beside others contains adjustments to consumption and financial savings, but also potentially different preventive decision in preparation for future shocks.

As the optimal decision rules do not allow for a thorough analytical investigation, we rely on numerical simulations to determine the impact of education in comparison to other household characteristics. Through our numerical calibration exercise we find that the effects of education (through its impact on the stochastic income process) on outcomes and decisions of the household are generally most pronounced among all household characteristics considered. Depending on the variable of interest the effects of education on exposure, vulnerability or other decision and state variables can be either enhanced or dampened by improvements along the other household characteristics. E.g. while higher awareness increases the positive effect of education on exposure reduction, better access to prevention measures implies the contrary. Among other results, this example is another indicator for the substitutability between different prevention strategies.

In conclusion, this thesis illustrates that behaviour with respect to risk consists of preventive and adaptive strategies. However, as the examples show, classifying an action strictly into one of the two categories is often hard to justify. In each of the three models we introduce multiple control variables to disentangle these effects as far as possible. The interconnectedness of different strategies and co-existence of multiple risk factors (risk of large health shock in presence of general mortality risk, or idiosyncratic income risk in addition to the risk of natural disasters) still adds substantial complexity to this task. By using technical frameworks specifically chosen for each research question, we are still able to identify some of the driving forces behind the optimal decision strategies either analytically or numerically. We also find trade-offs between preventive and adaptive efforts in all models, however

generally applicable conclusions on optimal strategies cannot be drawn. Risk behaviour and the forces, which are driving it remain highly context specific and require explicit consideration.

5.2 Potential extensions and outlook

With our three different frameworks we not only managed to extend the existing literature, but also paved the way for numerous further developments based on the work presented. For the first model in Chapter 2 we can find several other types of health shocks (e.g. cardio-vascular disease, stroke, Alzheimer disease) for numerical calibration and analysis. Consequently we can compare their respective outcomes and find differences in their impacts. Furthermore a different annuity market and the introduction of a health insurance market could be of interest for future research. This would allow to pin down how varying market structures affect the optimal behaviour and health outcomes at the individual level. Finally from a technical point of view, we would certainly expect interesting results from extending the uncertainty elements even further. First we could allow health shock to not only be stochastic in its timing, but also its immediate impact (e.g. stage of cancer at time of diagnosis can vary). Alternatively, the introduction of multiple potential shocks over the life-course could substantially improve the realism and relevance of the model (consider the reemergence of cancer or multiple heart attacks for the same individual). However, the last extension described requires further progress in the theory of distributed optimal control models before it can be applied.

Although the model of Chapter 3 might appear quite topical to the Covid-19 pandemic, research on the optimal containment of infectious diseases should not stop due to the declining impact of the current pandemic. The sub-optimal and inconsistent management strategies around the world especially in the early phases of the pandemic showed how valuable scientific expertise from all relevant fields could be. To be prepared for future outbreaks of highly contagious diseases, models analysing and incorporating many different aspects of policy intervention should be continuously developed and can help mitigate the financial and social burden for society. Furthermore the (as of April 2022 still looming) threat of potential virus mutations with unknown changes in infectiousness and lethality should be incentive enough for continued research. In this regard, extending the second paper with the disruptive shock of an emerging mutation (using the framework of the first paper) could give considerable insight on how a mutation should be handled after its emergence and which preventive actions should be taken.

For the dynamic household model in Chapter 4 we could first extend it from a partial to a general equilibrium model. Adding a production sector to the economy (and potentially a housing market) can close the system. This would not only justify the differences in wages for different educational groups, but would also allow to examine the impacts of natural hazards on the production side of the economy in the short- and long-run. This extension also makes the assessment of different policy experiments with respect to their effectiveness in curbing the individual and societal risk with respect to natural disasters possible.

List of Tables

2.1	Summary of all state and control variables in the basic framework	30
2.2	Summary of functional specifications and parameters in the model. Parameters indicated by ** are estimated before the solution process using only the empirical data. Parameters indicated with * result from the calibration process within the solution process. Parameters marked with + are chosen from the literature. Parameters without indications are either educated guesses (ρ, r, κ_i) or are manually chosen to improve the calibration $(\delta_1, \delta_2, \delta_3)$	41
3.1	Compartments (state variables) of the extended SIR model. The first three columns show the mathematical symbol, the description and the abbreviation used in the paper. The last column relates the compartments to those of the classical SIR model.	78
3.2	Parameters for all numerical scenarios.	90
3.3	Qualitative comparison of the four different cases analysed	90
3.4	Summary and comparison of the endstates of the pandemic and the costs of the pandemic across the three different regions for the four scenarios analysed. Endstates are given in percentage of the initial population size of each region, resp. the total population. Costs are given in units of GDP per capita per day (GDPPCPD).	92
3.5	Comparison of the 4 (main) scenarios.	100
3.6	Terminal states at $T = 400$ of the population (susceptibles S , recovered light R_L cases, diagnosed/heavy cases R_D , and deceased M) and cumulative costs in units of GDP per capita per day at days 100, 200, 300 and 400 in the "Uncontrolled" case.	106
3.7	Terminal states of the population (susceptibles S , recovered light R_L cases, diagnosed/heavy cases R_D , and deceased M) and cumulative costs in units of GDP per capita per day at days 100, 200, 300 and 400 in the "No testing" case.	107
3.8	Terminal states of the population (susceptibles S , recovered light R_L cases, diagnosed/heavy cases R_D , and deceased M) and cumulative costs in units of GDP per capita per day at days 100, 200, 300 and 400 in the "Ineffective testing" case.	107
3.9	Terminal states of the population (susceptibles S , recovered light R_L cases, diagnosed/heavy cases R_D , and deceased M) and cumulative costs in units of GDP per capita per day at days 100, 200, 300 and 400 in the "Perfect testing" case.	108
4.1	Summary of functional specifications and parameters in the model.	123
4.2	Impact of education levels, disaster risk awareness, prevention access, time discount rate and income on the main state and decision variables in the long-run equilibrium.	131
4.3	Impact of education levels, disaster risk awareness, prevention access, time discount rate and income on exposure, vulnerability and disaster risk.	138
4.4	Impact of the state variables in equilibrium, education levels, disaster risk awareness, prevention access, time discount rate and income on the decision making with respect to relocation, exposure, financial savings, physical asset accumulation, consumption and prevention.	146

List of Figures

2.1	Calibration results: Comparison between expenditures predicted by the model and expenditure data	42
2.2	Calibration results: Comparison of survival and mortality profiles between the model and the data	43
2.3	Consumption profiles in the first and in the second stage	44
2.4	Illustrations of the impacts of the different parts of the Euler equation for consumption before the diagnosis	46
2.5	Illustrations of the impacts of the different parts of the Euler equations for consumption after the diagnosis.	47
2.6	First row: General and cancer specific health expenditures before and after the diagnosis. Second row: Financial assets before and after the diagnosis.	48
2.7	Health/survival state and cancer deficits before and after the diagnosis	49
2.8	Valuations of Health	50
2.9	Valuations of Health	51
2.10	Illustrations of the impacts of the different parts of the Euler equations for general health expenditure before a cancer diagnosis	54
2.11	Illustrations of the impacts of the different parts of the Euler equations	54
2.12	Illustrations of the impacts of the different parts of the Euler equations for cancer specific care after a diagnosis.	56
2.13	Graphical illustration of the change in the order of integration	61
3.1	Flow chart of the SIR-type model. Grey boxes: compartments, grey ellipses: controls, black arrows: flows between the compartments, dashed arrows: flows that can be changed by controls.	78
3.2	Original and final smoothed version of the mortality function describing congestion in the health sector.	89
3.3	Pandemic development for the "Uncontrolled" case.	91
3.4	Comparison of the pandemic development for the "No testing", "Ineffective Testing" and "Perfect testing" case.	94
3.5	Optimal target transmission rates for the "No testing", "Ineffective Testing" and "Perfect testing" case.	96
3.6	Optimal testing strategy in the "ineffective" and "perfect" testing case.	97
3.7	Origins of costs for the "No testing" (left), "Ineffective testing" (middle) and "Perfect testing" case (right).	99
4.1	Conceptual framework	113
4.2	Distribution of households across the key characteristics education, awareness and prevention access.	124
4.3	Correlations between the state variables in equilibrium for the simulation results (left) and the empirical data (right)	125
4.4	Equilibrium distribution of the main state and decision variables in the simulated and empirical data.	128

4.5	Equilibrium distributions of control and state variables by education level in simulated and empirical data.	129
4.6	Optimal relocation and exposure decision by households depending on their educational level.	134
4.7	Optimal decisions for the net-investments into financial assets depending on the current status and the households educational level.	135
4.8	Optimal decisions for the net-investments into physical assets depending on the current status and the households educational level.	136
4.9	Optimal decisions for the net-investments into physical assets depending on the current status and the households educational level grouped by disaster experience.	136

Bibliography

- Abel, Guy J., Michael Brottrager, Jesus Crespo Cuaresma, and Raya Muttarak (Jan. 2019). “Climate, conflict and forced migration”. en. In: *Global Environmental Change* 54, pp. 239–249. ISSN: 0959-3780. DOI: 10.1016/j.gloenvcha.2018.12.003. URL: <https://www.sciencedirect.com/science/article/pii/S0959378018301596> (visited on 03/07/2022).
- Acemoglu, Daron, Victor Chernozhukov, Iván Werning, Michael D Whinston, et al. (2020). *A multi-risk SIR model with optimally targeted lockdown*. Vol. 2020. National Bureau of Economic Research Cambridge, MA.
- Adger, W Neil, Terry P Hughes, Carl Folke, Stephen R Carpenter, and Johan Rockstrom (2005). “Social-ecological resilience to coastal disasters”. In: *Science* 309(5737), pp. 1036–1039.
- Aerts, Jeroen CJH, Wouter J Botzen, Keith C Clarke, Susan L Cutter, Jim W Hall, Bruno Merz, Erwann Michel-Kerjan, Jaroslav Mysiak, Swenja Surminski, and Howard Kunreuther (2018). “Integrating human behaviour dynamics into flood disaster risk assessment”. In: *Nature Climate Change* 8(3), pp. 193–199.
- Alam, Mansoor, J Lynn, and V Sarma (Sept. 1976). “Optimal maintenance policy for equipment subject to random deterioration and random failure A modern control theory approach”. In: *International Journal of Systems Science* 7(9), pp. 1071–1080. DOI: 10.1080/00207727608941989.
- Aldy, Joseph E and W Kip Viscusi (2008). “Adjusting the value of a statistical life for age and cohort effects”. In: *The Review of Economics and Statistics* 90(3), pp. 573–581.
- Alvarez, Fernando, David Argente, and Francesco Lippi (2021). “A simple planning problem for COVID-19 lock-down, testing, and tracing”. In: *American Economic Review: Insights* 3(3), pp. 367–82.
- Angelov, G, R Kovacevic, NI Stilianakis, and VM Veliov (2021). “Optimal vaccination strategies using a distributed epidemiological model applied to COVID-19”. In.
- Aspri, Andrea, Elena Beretta, Alberto Gandolfi, and Etienne Wasmer (2021). “Mortality containment vs. economics opening: optimal policies in a SEIARD model”. In: *Journal of mathematical economics* 93, p. 102490.
- Baker, David P, Juan Leon, Emily G Smith Greenaway, John Collins, and Marcela Movit (2011). “The education effect on population health: a reassessment”. In: *Population and development review* 37(2), pp. 307–332.
- Bellman, Richard (1954). “The theory of dynamic programming”. In: *Bulletin of the American Mathematical Society* 60(6), pp. 503–515.
- Bíró, Anikó (2013). “Subjective mortality hazard shocks and the adjustment of consumption expenditures”. In: *Journal of Population Economics* 26(4), pp. 1379–1408.
- Black, Richard, W Neil Adger, Nigel W Arnell, Stefan Dercon, Andrew Geddes, and David Thomas (2011). “The effect of environmental change on human migration”. In: *Global environmental change* 21, S3–S11.
- Blair, Clancy, David Gamson, Steven Thorne, and David Baker (2005). “Rising mean IQ: Cognitive demand of mathematics education for young children, population exposure to formal schooling, and the neurobiology of the prefrontal cortex”. In: *Intelligence* 33(1), pp. 93–106.
- Bliman, Pierre-Alexandre, Michel Duprez, Yannick Privat, and Nicolas Vauchelet (2021). “Optimal Immunity Control and Final Size Minimization by Social Distancing for the SIR Epidemic Model”. In: *Journal of Optimization Theory and Applications* 189(2), pp. 408–436.

- Bloom, David E, Michael Kuhn, and Klaus Prettnner (2020). *Modern infectious diseases: Macroeconomic impacts and policy responses*. Tech. rep. National Bureau of Economic Research.
- Blumen, Helen, Kathryn Fitch, and Vincent Polkus (2016). “Comparison of treatment costs for breast cancer, by tumor stage and type of service”. In: *American health & drug benefits* 9(1), p. 23.
- Bonnans, J Frédéric and Justina Gianatti (2020). “Optimal control techniques based on infection age for the study of the COVID-19 epidemic”. In: *Mathematical Modelling of Natural Phenomena* 15, p. 48.
- Boucekkine, Raouf, Andrés Carvajal, Shankha Chakraborty, and Aditya Goenka (2021). “The economics of epidemics and contagious diseases: An introduction”. In: *Journal of Mathematical Economics*.
- Boukas, El-Kébir, Alain Haurie, and Philippe Michel (1990). “An optimal control problem with a random stopping time”. In: *Journal of optimization theory and applications* 64(3), pp. 471–480.
- Bruine de Bruin, Wändi, Andrew M Parker, and Baruch Fischhoff (2007). “Individual differences in adult decision-making competence.” In: *Journal of personality and social psychology* 92(5), p. 938.
- Buratto, A., M. Muttoni, S. Wrzaczek, and M. Freiberger (2022). “Should the COVID-19 lockdown be relaxed or intensified in case a vaccine becomes available?” In: *Submitted*.
- Buratto, Alessandra, Maddalena Muttoni, Stefan Wrzaczek, and Michael Freiberger (2021). “Should the COVID-19 lockdown be relaxed or intensified in case a vaccine becomes available?” In: *Plos one (submitted)*.
- Capatina, Elena (2015). “Life-cycle effects of health risk”. In: *Journal of Monetary Economics* 74, pp. 67–88.
- Card, David (1999). “The causal effect of education on earnings”. In: *Handbook of labor economics* 3, pp. 1801–1863.
- Carli, Raffaele, Graziana Cavone, Nicola Epicoco, Paolo Scarabaggio, and Mariagrazia Dotoli (2020). “Model predictive control to mitigate the COVID-19 outbreak in a multi-region scenario”. In: *Annual Reviews in Control* 50, pp. 373–393.
- Carlson, Dean A, Alain B Haurie, and Arie Leizarowitz (2012). *Infinite horizon optimal control: deterministic and stochastic systems*. Springer Science & Business Media.
- Caulkins, Jonathan P, Dieter Grass, Gustav Feichtinger, Richard F Hartl, Peter M Kort, Alexia Prskawetz, Andrea Seidl, and Stefan Wrzaczek (2021). “The optimal lockdown intensity for COVID-19”. In: *Journal of Mathematical Economics* 93, p. 102489.
- Caulkins, Jonathan P, Dieter Grass, Gustav Feichtinger, Richard F Hartl, Peter M Kort, Alexia Prskawetz, Andrea Seidl, and Stefan Wrzaczek (2022). “COVID-19 and optimal lockdown strategies: The effect of new and more virulent strains”. In: *Pandemics: Insurance and Social Protection*. Springer, Cham, pp. 163–190.
- Caulkins, Jonathan, Dieter Grass, Gustav Feichtinger, Richard Hartl, Peter M Kort, Alexia Prskawetz, Andrea Seidl, and Stefan Wrzaczek (2020). “How long should the COVID-19 lockdown continue?” In: *Plos one* 15(12), e0243413.
- Ceci, Stephen J (1991). “How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence.” In: *Developmental psychology* 27(5), p. 703.
- Centre for Research on the Epidemiology of Disasters (CRED) (2022). *EM-DAT: The Emergency Events Database*. URL: www.emdat.be.
- Chew, Soo Hong, James Heckman, Junjian Yi, Junsen Zhang, and Songfa Zhong (2010). “Education and preferences: Experimental evidence from Chinese adult twins”. In: *Education* 1, p. 13.
- Cole, Harold L, Soojin Kim, and Dirk Krueger (2019). “Analysing the Effects of Insuring Health Risks: On the Trade-off between Short-Run Insurance Benefits versus Long-Run Incentive Costs”. In: *The Review of Economic Studies* 86(3), pp. 1123–1169.
- Courbage, Christophe, Béatrice Rey, and Nicolas Treich (2013). “Prevention and precaution”. In: *Handbook of insurance*, pp. 185–204.
- Dalgaard, Carl-Johan and Holger Strulik (2014). “Optimal aging and death: understanding the Preston curve”. In: *Journal of the European Economic Association* 12(3), pp. 672–701.
- Dasgupta, Partha (1998). “The economics of poverty in poor countries”. In: *Scandinavian Journal of Economics* 100(1), pp. 41–68.

- De Nardi, Mariacristina, Eric French, and John B Jones (2010). “Why do the elderly save? The role of medical expenses”. In: *Journal of Political Economy* 118(1), pp. 39–75.
- De Zeeuw, Aart and Amos Zemel (2012). “Regime shifts and uncertainty in pollution control”. In: *Journal of Economic Dynamics and Control* 36(7), pp. 939–950.
- Drabo, Alassane and Linguère Mously Mbaye (2015). “Natural disasters, migration and education: an empirical analysis in developing countries”. In: *Environment and Development Economics* 20(6), pp. 767–796.
- Eeckhoudt, Louis, Christian Gollier, and Nicolas Treich (2005). “Optimal consumption and the timing of the resolution of uncertainty”. In: *European Economic Review* 49(3), pp. 761–773.
- Eeckhoudt, Louis and Harris Schlesinger (2008). “Changes in risk and the demand for saving”. In: *Journal of Monetary Economics* 55(7), pp. 1329–1336.
- Ehrlich, Isaac (2000). “Uncertain lifetime, life protection, and the value of life saving”. In: *Journal of Health Economics* 19(3), pp. 341–367.
- Eichenbaum, Martin S, Sergio Rebelo, and Mathias Trabandt (2021). “The macroeconomics of epidemics”. In: *The Review of Financial Studies* 34(11), pp. 5149–5187.
- Eslinger, Paul J, Clancy Blair, JianLi Wang, Bryn Lipovsky, Jennifer Realmuto, David Baker, Steven Thorne, David Gamson, Erin Zimmerman, Lisa Rohrer, et al. (2009). “Developmental shifts in fMRI activations during visuospatial relational reasoning”. In: *Brain and cognition* 69(1), pp. 1–10.
- Fabbri, Giorgio, Fausto Gozzi, and Giovanni Zanco (2021). “Verification results for age-structured models of economic–epidemics dynamics”. In: *Journal of Mathematical Economics* 93, p. 102455.
- Federico, Salvatore and Giorgio Ferrari (2021). “Taming the spread of an epidemic by lockdown policies”. In: *Journal of mathematical economics* 93, p. 102453.
- Feichtinger, Gustav, Gernot Tragler, and Vladimir M. Veliov (2003). “Optimality conditions for age-structured control systems”. In: *Journal of Mathematical Analysis and Applications* 288(1), pp. 47–68.
- Finkelstein, Amy, Erzo FP Luttmer, and Matthew J Notowidigdo (2013). “What good is wealth without health? The effect of health on the marginal utility of consumption”. In: *Journal of the European Economic Association* 11(suppl_1), pp. 221–258.
- Fonseca, Raquel, Pierre-Carl Michaud, Arie Kapteyn, and Titus J Galama (2013). “Accounting for the rise of health spending and longevity”. In.
- Fothergill, Alice and Lori A Peek (2004). “Poverty and disasters in the United States: A review of recent sociological findings”. In: *Natural hazards* 32(1), pp. 89–110.
- Frankenberg, Elizabeth, Bondan Sikoki, Cecep Sumantri, Wayan Suriastini, and Duncan Thomas (2013). “Education, vulnerability, and resilience after a natural disaster”. In: *Ecology and society: a journal of integrative science for resilience and sustainability* 18(2), p. 16.
- Frankovic, Ivan, Michael Kuhn, and Stefan Wrzaczek (2020). “On the anatomy of medical progress within an overlapping generations economy”. In: *De Economist*, pp. 1–43.
- Freiberger, Michael and Michael Kuhn (2020). *Modelling health processes through survival*. Mimeo.
- Freiberger, Michael, Dieter Grass, Michael Kuhn, Andrea Seidl, and Stefan Wrzaczek (2022a). *Allocating tests and lockdowns to contain pandemic spread and medical overload within a network*. In final revision for publication in the Journal of Public Economic Theory.
- Freiberger, Michael, Michael Kuhn, and Stefan Wrzaczek (2022b). *Integrating large health shocks into life-cycle models - Modelling impacts of a cancer diagnosis*. In preparation for submission.
- Freiberger, Michael, Roman Hoffmann, and Alexia Prskawetz (2022c). *Modelling Disaster Risks: A Dynamic Household Model*. In preparation for submission.
- French, Eric (2005). “The effects of health, wealth, and wages on labour supply and retirement behaviour”. In: *The Review of Economic Studies* 72(2), pp. 395–427.
- French, Eric and John Bailey Jones (2011). “The effects of health insurance and self-insurance on retirement behavior”. In: *Econometrica* 79(3), pp. 693–732.
- Gaimon, Cheryl and Gerald L Thompson (1984). “Optimal preventive and repair maintenance of a machine subject to failure”. In: *Optimal Control Applications and Methods* 5(1), pp. 57–67.

- Gamchi, Nafiseh Shamsi, S Ali Torabi, and Fariborz Jolai (2021). “A novel vehicle routing problem for vaccine distribution using SIR epidemic model”. In: *OR Spectrum* 43(1), pp. 155–188.
- Garbero, Alessandra and Raya Muttarak (2013). “Impacts of the 2010 droughts and floods on community welfare in rural Thailand: differential effects of village educational attainment”. In: *Ecology and Society* 18(4).
- Gersovitz, Mark and Jeffrey S Hammer (2004). “The economical control of infectious diseases”. In: *The Economic Journal* 114(492), pp. 1–27.
- Goenka, Aditya, Lin Liu, and Manh-Hung Nguyen (2021). “SIR economic epidemiological models with disease induced mortality”. In: *Journal of Mathematical Economics* 93, p. 102476.
- Gollier, Christian (2020). “Cost–benefit analysis of age-specific deconfinement strategies”. In: *Journal of Public Economic Theory* 22(6), pp. 1746–1771.
- Goodwin, James S, Jonathan M Samet, Charles R Key, Charles Humble, Daniel Kutvirt, and Curtis Hunt (1986). “Stage at diagnosis of cancer varies with the age of the patient”. In: *Journal of the American Geriatrics Society* 34(1), pp. 20–26.
- Gori, Luca, Piero Manfredi, Simone Marsiglio, and Mauro Sodini (2021). “COVID-19 epidemic and mitigation policies: Positive and normative analyses in a neoclassical growth model”. In: *Journal of Public Economic Theory*.
- Grass, Dieter, Jonathan P Caulkins, Gustav Feichtinger, Gernot Tragler, Doris A Behrens, et al. (2008). “Optimal control of nonlinear processes”. In: *Berlino: Springer*.
- Grossman, Michael (1972). “On the concept of health capital and the demand for health”. In: *Journal of Political economy* 80(2), pp. 223–255.
- Grossman, Michael (2006). “Education and nonmarket outcomes”. In: *Handbook of the Economics of Education* 1, pp. 577–633.
- Grundel, Sara M, Stefan Heyder, Thomas Hotz, Tobias KS Ritschel, Philipp Sauerteig, and Karl Worthmann (2021). “How to coordinate vaccination and social distancing to mitigate SARS-CoV-2 outbreaks”. In: *SIAM Journal on Applied Dynamical Systems* 20(2), pp. 1135–1157.
- Hall, Robert E and Charles I Jones (2007). “The value of life and the rise in health spending”. In: *The Quarterly Journal of Economics* 122(1), pp. 39–72.
- Hansen, Elsa and Troy Day (2011). “Optimal control of epidemics with limited resources”. In: *Journal of mathematical biology* 62(3), pp. 423–451.
- Hansen, Gary D and Selahattin İmrohoroğlu (2008). “Consumption over the life cycle: The role of annuities”. In: *Review of Economic Dynamics* 11(3), pp. 566–583.
- Hassager, Christian, Ken Nagao, and David Hildick-Smith (2018). “Out-of-hospital cardiac arrest: in-hospital intervention strategies”. In: *The Lancet* 391(10124), pp. 989–998.
- Heckman, James J, John Eric Humphries, and Gregory Veramendi (2018). “Returns to education: The causal effects of education on earnings, health, and smoking”. In: *Journal of Political Economy* 126(S1), S197–S246.
- Heuser, Louis, John S Spratt, and Hiram C Polk Jr (1979). “Growth rates of primary breast cancers”. In: *Cancer* 43(5), pp. 1888–1894.
- Hoffmann, Roman and Raya Muttarak (2017). “Learn from the past, prepare for the future: Impacts of education and experience on disaster preparedness in the Philippines and Thailand”. In: *World Development* 96, pp. 32–51.
- Hoffmann, Roman and Daniela Blecha (2020). “Education and disaster vulnerability in Southeast Asia: Evidence and policy implications”. In: *Sustainability* 12(4), p. 1401.
- Holling, Crawford S (1973). “Resilience and stability of ecological systems”. In: *Annual review of ecology and systematics* 4(1), pp. 1–23.
- Hugonnier, Julien, Florian Pelgrin, and Pascal St-Amour (2013). “Health and (other) asset holdings”. In: *Review of Economic Studies* 80(2), pp. 663–710.
- Hunter, Lori M, Jessie K Luna, and Rachel M Norton (2015). “Environmental dimensions of migration”. In: *Annual Review of Sociology* 41, pp. 377–397.

- Ikefuji, Masako and Ryo Horii (2007). “Wealth heterogeneity and escape from the poverty–environment trap”. In: *Journal of Public Economic Theory* 9(6), pp. 1041–1068.
- Institute for Health Metrics and Evaluation (IHME) (2016). *Tracking personal health care spending in the US*. URL: <http://vizhub.healthdata.org/dex>.
- IPCC (2022). *Climate Change 2022: Impacts, Adaptation, and Vulnerability*. Cambridge University Press. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [H.-O. Pörtner, D. C. Roberts, M. Tignor, E. S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama (eds.)]
- Irmansyah, I, Suryo Dharmono, Albert Maramis, and Harry Minas (2010). “Determinants of psychological morbidity in survivors of the earthquake and tsunami in Aceh and Nias”. In: *International Journal of Mental Health Systems* 4(1), pp. 1–10.
- Jung, Dawoon, Tushar Bharati, and Seungwoo Chin (2021). “Does education affect time preference? Evidence from Indonesia”. In: *Economic Development and Cultural Change* 69(4), pp. 1451–1499.
- Jung, Juergen and Chung Tran (2016). “Market inefficiency, insurance mandate and welfare: US health care reform 2010”. In: *Review of Economic Dynamics* 20, pp. 132–159.
- Kamien, Morton I and Nancy L Schwartz (1971). “Optimal maintenance and sale age for a machine subject to failure”. In: *Management Science* 17(8), B–495.
- Kermack, M and A Mckendrick (1927). “Contributions to the mathematical theory of epidemics. Part I”. In: *Proc. r. soc. a* 115(5), pp. 700–721.
- Khwaja, Ahmed, Frank Sloan, and Sukyung Chung (2006). “Learning about individual risk and the decision to smoke”. In: *International Journal of Industrial Organization* 24(4), pp. 683–699.
- Kirschenbaum, Alan (2006). “Families and disaster behavior: a reassessment of family preparedness”. In: *International Journal of Mass Emergencies and Disasters* 24(1), p. 111.
- Kohn, Sivan, Jennifer Lipkowitz Eaton, Saad Feroz, Andrea A Bainbridge, Jordan Hoolachan, and Daniel J Barnett (2012). “Personal disaster preparedness: an integrative review of the literature”. In: *Disaster medicine and public health preparedness* 6(3), pp. 217–231.
- Kopecky, Karen A and Tatyana Koreshkova (2014). “The impact of medical and nursing home expenses on savings”. In: *American Economic Journal: Macroeconomics* 6(3), pp. 29–72.
- Krueger, Dirk and Alexander Ludwig (2016). “On the optimal provision of social insurance: Progressive taxation versus education subsidies in general equilibrium”. In: *Journal of Monetary Economics* 77, pp. 72–98.
- Kuhlicke, Christian, Sebastian Seebauer, Paul Hudson, Chloe Begg, Philip Bubeck, Cordula Dittmer, Torsten Grothmann, Anna Heidenreich, Heidi Kreibich, Daniel F Lorenz, et al. (2020). “The behavioral turn in flood risk management, its assumptions and potential implications”. In: *Wiley Interdisciplinary Reviews: Water* 7(3), e1418.
- Kuhn, Michael, Stefan Wrzaczek, and Jim Oeppen (2010). “Recognizing progeny in the value of life”. In: *Economics Letters* 107(1), pp. 17–21.
- Kuhn, Michael, Stefan Wrzaczek, Alexia Prskawetz, and Gustav Feichtinger (2011). “Externalities in a life cycle model with endogenous survival”. In: *Journal of Mathematical Economics* 47(4–5), pp. 627–641.
- Kuhn, Michael, Stefan Wrzaczek, Alexia Prskawetz, and Gustav Feichtinger (2015). “Optimal choice of health and retirement in a life-cycle model”. In: *Journal of Economic Theory* 158, pp. 186–212.
- Kuhn, Michael and Stefan Wrzaczek (2021). “Rationally risking addiction: a two-stage approach”. In: *Dynamic Economic Problems with Regime Switches*. Springer, pp. 85–110.
- Kydland, Finn E and Edward C Prescott (1982). “Time to build and aggregate fluctuations”. In: *Econometrica: Journal of the Econometric Society*, pp. 1345–1370.
- Laporte, Audrey and Brian S Ferguson (2007). “Investment in health when health is stochastic”. In: *Journal of Population Economics* 20(2), pp. 423–444.
- Lave, Tamara R and Lester B Lave (1991). “Public perception of the risks of floods: Implications for communication”. In: *Risk analysis* 11(2), pp. 255–267.

- Lechowska, Ewa (2018). “What determines flood risk perception? A review of factors of flood risk perception and relations between its basic elements”. In: *Natural Hazards* 94(3), pp. 1341–1366.
- Lee, James J (2010). “Review of Intelligence and how to get it: Why schools and cultures count.” In.
- Lee, Ronald Demos and Andrew Mason (2011). *Population aging and the generational economy: A global perspective*. Edward Elgar Publishing.
- Leitmann, George and Harold Stalford (1971). “A sufficiency theorem for optimal control”. In: *Journal of Optimization Theory and Applications* 8(3), pp. 169–174.
- Lengwiler, Yvan (2009). “The origins of expected utility theory”. In: *Vinzenz Bronzin’s Option Pricing Models*. Springer, pp. 535–545.
- Li, Chuan-Zhong, Anne-Sophie Crépin, Carl Folke, et al. (2018). “The economics of resilience”. In: *International Review of Environmental and Resource Economics* 11(4), pp. 309–353.
- Lutz, Wolfgang, Raya Muttarak, and Erich Striessnig (2014). “Universal education is key to enhanced climate adaptation”. In: *Science* 346(6213), pp. 1061–1062.
- Mandelblatt, Jeanne, Howard Andrews, Jon Kerner, Ann Zauber, and William Burnett (1991). “Determinants of late stage diagnosis of breast and cervical cancer: the impact of age, race, social class, and hospital type”. In: *American journal of public health* 81(5), pp. 646–649.
- Marti, Joachim and Michael R Richards (2017). “Smoking response to health and medical spending changes and the role of insurance”. In: *Health economics* 26(3), pp. 305–320.
- Masson-Delmotte, V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield (n.d.). “IPCC, 2018: Annex I: Glossary [Matthews, J.B.R. (ed.)]” In: *Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty* ().
- Meyer, Andrew (2015). “Does education increase pro-environmental behavior? Evidence from Europe”. In: *Ecological economics* 116, pp. 108–121.
- Murphy, Kevin M and Robert H Topel (2006). “The value of health and longevity”. In: *Journal of political Economy* 114(5), pp. 871–904.
- Muttarak, Raya and Wiraporn Pothisiri (2013). “The role of education on disaster preparedness: case study of 2012 Indian Ocean earthquakes on Thailand’s Andaman Coast”. In: *Ecology and Society* 18(4).
- Muttarak, Raya and Wolfgang Lutz (2014). “Is education a key to reducing vulnerability to natural disasters and hence unavoidable climate change?” In: *Ecology and Society* 19(1).
- Nawrotzki, Raphael J, Fernando Riosmena, Lori M Hunter, and Daniel M Runfola (2015). “Amplification or suppression: Social networks and the climate change—migration association in rural Mexico”. In: *Global Environmental Change* 35, pp. 463–474.
- Norris, Fran H, Tenbroeck Smith, and Krzysztof Kaniasty (1999). “Revisiting the experience–behavior hypothesis: The effects of Hurricane Hugo on hazard preparedness and other self-protective acts”. In: *Basic and Applied Social Psychology* 21(1), pp. 37–47.
- Obokata, Reiko, Luisa Veronis, and Robert McLeman (2014). “Empirical research on international environmental migration: a systematic review”. In: *Population and environment* 36(1), pp. 111–135.
- Oreopoulos, Philip and Kjell G Salvanes (2011). “Priceless: The nonpecuniary benefits of schooling”. In: *Journal of Economic perspectives* 25(1), pp. 159–84.
- Ouardighi, Fouad El, EUGENE Khmelnsky, and Suresh Sethi (2020). “Control of an Epidemic With Endogenous Treatment Capability Under Popular Discontent and Social Fatigue”. In: *Available at SSRN 3731673*.
- Palumbo, Michael G (1999). “Uncertain medical expenses and precautionary saving near the end of the life cycle”. In: *The Review of Economic Studies* 66(2), pp. 395–421.
- Paton, Douglas and David Johnston (2001). “Disasters and communities: vulnerability, resilience and preparedness”. In: *Disaster Prevention and Management: An International Journal*.

- Peters, Ellen, Daniel Västfjäll, Paul Slovic, CK Mertz, Ketti Mazzocco, and Stephan Dickert (2006). “Numeracy and decision making”. In: *Psychological science* 17(5), pp. 407–413.
- Pichler, Adelheid and Erich Striessnig (2013). “Differential vulnerability to hurricanes in Cuba, Haiti, and the Dominican Republic: the contribution of education”. In: *Ecology and society* 18(3).
- Picone, Gabriel, Martín Uribe, and R Mark Wilson (1998). “The effect of uncertainty on the demand for medical care, health capital and wealth”. In: *Journal of health economics* 17(2), pp. 171–185.
- Polasky, Stephen, Aart De Zeeuw, and Florian Wagener (2011). “Optimal management with potential regime shifts”. In: *Journal of Environmental Economics and management* 62(2), pp. 229–240.
- Pontryagin, Lev Semenovich (1987). *Mathematical theory of optimal processes*. CRC press.
- Quartz, Steven R and Terrence J Sejnowski (1997). “The neural basis of cognitive development: A constructivist manifesto”. In: *Behavioral and brain sciences* 20(4), pp. 537–556.
- Reed, William J (1984). “The effects of the risk of fire on the optimal rotation of a forest”. In: *Journal of environmental economics and management* 11(2), pp. 180–190.
- Reed, William J (1987). “Protecting a forest against fire: optimal protection patterns and harvest policies”. In: *Natural Resource Modeling* 2(1), pp. 23–53.
- Reed, William J (1989). “Optimal investment in the protection of a vulnerable biological resource”. In: *Natural Resource Modeling* 3(4), pp. 463–480.
- Reichling, Felix and Kent Smetters (2015). “Optimal annuitization with stochastic mortality and correlated medical costs”. In: *American Economic Review* 105(11), pp. 3273–3320.
- Reininger, Belinda M, Mohammad H Rahbar, MinJae Lee, Zhongxue Chen, Sartaj R Alam, Jennifer Pope, and Barbara Adams (2013). “Social capital and disaster preparedness among low income Mexican Americans in a disaster prone area”. In: *Social Science & Medicine* 83, pp. 50–60.
- Richard, Quentin, Samuel Alizon, Marc Choisy, Mircea T Sofonea, and Ramsès Djidjou-Demasse (2021). “Age-structured non-pharmaceutical interventions for optimal control of COVID-19 epidemic”. In: *PLoS computational biology* 17(3), e1008776.
- Rosen, Sherwin (1988). “The value of changes in life expectancy”. In: *Journal of Risk and uncertainty* 1(3), pp. 285–304.
- Al-Rousan, Tala M, Linda M Rubenstein, and Robert B Wallace (2014). “Preparedness for natural disasters among older US adults: a nationwide survey”. In: *American journal of public health* 104(3), pp. 506–511.
- Rowen, Donna and John Brazier (2011). “Health utility measurement”. In: *The oxford handbook of health economics*.
- Russell, Lisa A, James D Goltz, and Linda B Bourque (1995). “Preparedness and hazard mitigation actions before and after two earthquakes”. In: *Environment and behavior* 27(6), pp. 744–770.
- Sachs, Jeffrey, John W McArthur, Guido Schmidt-Traub, Margaret Kruk, Chandrika Bahadur, Michael Faye, and Gordon McCord (2004). “Ending Africa’s poverty trap”. In: *Brookings papers on economic activity* 2004(1), pp. 117–240.
- Satariano, William A, Steven H Belle, and G Marie Swanson (1986). “The severity of breast cancer at diagnosis: a comparison of age and extent of disease in black and white women”. In: *American Journal of Public Health* 76(7), pp. 779–782.
- Schechtman, Jack (1976). “An income fluctuation problem”. In: *Journal of Economic Theory* 12(2), pp. 218–241.
- Schuenemann, Johannes, Holger Strulik, and Timo Trimborn (2017). “Going from bad to worse: Adaptation to poor health health spending, longevity, and the value of life”. In: *Journal of Economic Behavior & Organization* 140, pp. 130–146.
- Selin, Steve and Deborah Chevez (1995). “Developing a collaborative model for environmental planning and management”. In: *Environmental management* 19(2), pp. 189–195.
- Sharma, Upasna, Anand Patwardhan, and Anthony G Patt (2013). “Education as a determinant of response to cyclone warnings: evidence from coastal zones in India”. In: *Ecology and Society* 18(2).
- Shepard, Donald S and Richard J Zeckhauser (1984). “Survival versus consumption”. In: *Management Science* 30(4), pp. 423–439.

- Siders, AR, Miyuki Hino, and Katharine J Mach (2019). “The case for strategic and managed climate retreat”. In: *Science* 365(6455), pp. 761–763.
- Smith, Diane L and Stephen J Notaro (2009). “Personal emergency preparedness for people with disabilities from the 2006-2007 Behavioral Risk Factor Surveillance System”. In: *Disability and health journal* 2(2), pp. 86–94.
- Smith, James E and Ralph L Keeney (2005). “Your money or your life: A prescriptive model for health, safety, and consumption decisions”. In: *Management Science* 51(9), pp. 1309–1325.
- Smith, V Kerry, Donald H Taylor Jr, Frank A Sloan, F Reed Johnson, and William H Desvousges (2001). “Do smokers respond to health shocks?” In: *Review of Economics and Statistics* 83(4), pp. 675–687.
- Solberg, Christian, Tiziana Rossetto, and Helene Joffe (2010). “The social psychology of seismic hazard adjustment: re-evaluating the international literature”. In: *Natural Hazards and Earth System Sciences* 10(8), pp. 1663–1677.
- Sorger, Gerhard (1991). “Maximum principle for control problems with uncertain horizon and variable discount rate”. In: *Journal of optimization theory and applications* 70(3), pp. 607–618.
- Surveillance Research Program, National Cancer Institute (Sept. 2020). *SEER*Explorer: An interactive website for SEER cancer statistics [Internet]*. URL: <https://seer.cancer.gov/explorer/>.
- Talkington, Anne and Rick Durrett (2015). “Estimating tumor growth rates in vivo”. In: *Bulletin of mathematical biology* 77(10), pp. 1934–1954.
- Tauchen, George (1986). “Finite state markov-chain approximations to univariate and vector autoregressions”. In: *Economics letters* 20(2), pp. 177–181.
- Thieken, Annegret H, Heidi Kreibich, Meike Müller, and Bruno Merz (2007). “Coping with floods: preparedness, response and recovery of flood-affected residents in Germany in 2002”. In: *Hydrological Sciences Journal* 52(5), pp. 1016–1037.
- Tomiyaama, Ken (1985). “Two-stage optimal control problems and optimality conditions”. In: *Journal of Economic Dynamics and Control* 9(3), pp. 317–337.
- Tomiyaama, Ken and Robert J Rossana (1989). “Two-stage optimal control problems with an explicit switch point dependence: Optimality criteria and an example of delivery lags and investment”. In: *Journal of Economic Dynamics and Control* 13(3), pp. 319–337.
- Tsur, Yacov and Amos Zemel (1995). “Uncertainty and irreversibility in groundwater resource management”. In: *Journal of environmental Economics and Management* 29(2), pp. 149–161.
- Tsur, Yacov and Amos Zemel (1996). “Accounting for global warming risks: resource management under event uncertainty”. In: *Journal of Economic Dynamics and Control* 20(6), pp. 1289–1305.
- Tsur, Yacov and Amos Zemel (1998). “Pollution control in an uncertain environment”. In: *Journal of Economic Dynamics and Control* 22(6), pp. 967–975.
- United Nations (2013). *National transfer accounts manual: Measuring and analysing the generational economy*. UN. URL: <https://www.ntaccounts.org/web/nta/show/Browse%20database>.
- University of California, Berkeley (USA) and Max Planck Institute for Demographic Research (Germany) (2020). *Human Mortality Database*.
- Van den Driessche, Pauline and James Watmough (2002). “Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission”. In: *Mathematical biosciences* 180(1-2), pp. 29–48.
- Veliov, Vladimir M. (2003). “Newton’s method for problems of optimal control of heterogeneous systems”. In: *Optimization Methods and Software* 18(6), pp. 689–703.
- Viscusi, W Kip and Michael J Moore (1989). “Rates of time preference and valuations of the duration of life”. In: *Journal of public economics* 38(3), pp. 297–317.
- Wamsler, Christine, Ebba Brink, and Oskari Rantala (2012). “Climate Change, Adaptation and Formal Education: The Role of Schooling for Increasing Societies’ Adaptive Capacities in El Salvador and Brazil”. In: *Ecology and Society* 17(2).
- Wisner, Ben, Piers Blaikie, Terry Cannon, and Ian Davis (2014). *At risk: natural hazards, people’s vulnerability and disasters*. Routledge.

- Witvorapong, Nopphol, Raya Muttarak, and Wiraporn Pothisiri (2015). “Social participation and disaster risk reduction behaviors in tsunami prone areas”. In: *PLoS one* 10(7), e0130862.
- Wrzaczek, Stefan, Michael Kuhn, and Ivan Frankovic (2020). “Using Age Structure for a Multi-stage Optimal Control Model with Random Switching Time”. In: *Journal of Optimization Theory and Applications* 184(3), pp. 1065–1082.
- Xiao, Chenyang and Aaron M McCright (2007). “Environmental concern and sociodemographic variables: A study of statistical models”. In: *The Journal of Environmental Education* 38(2), pp. 3–14.
- Yaari, Menahem E (1965). “Uncertain lifetime, life insurance, and the theory of the consumer”. In: *The Review of Economic Studies* 32(2), pp. 137–150.
- Yancik, Rosemary, Margaret N Wesley, Lynn AG Ries, Richard J Havlik, Brenda K Edwards, and Jerome W Yates (2001). “Effect of age and comorbidity in postmenopausal breast cancer patients aged 55 years and older”. In: *Jama* 285(7), pp. 885–892.
- Yogo, Motohiro (2016). “Portfolio choice in retirement: Health risk and the demand for annuities, housing, and risky assets”. In: *Journal of Monetary Economics* 80, pp. 17–34.

Curriculum Vitae

Personal information

Michael Freiberger

Date of birth

Place of birth

Nationality

Austria

Working Experience

International Institute for Applied Systems Analysis (IIASA), Laxenburg *5/2021 - present*

Research Scholar

Economic Frontiers Programme

Wittgenstein Centre (Vienna Institute of Demography), Vienna *4/2018 - 4/2021*

Projectassistant

FWF-Project “Life-cycle behaviour in the face of large shocks to health”

Supervisor: Dr. Stefan Wrzaczek

TU Wien, Vienna *4/2017 - 3/2018*

Projectassistant (FWF)

Doctoral Programme on Water Resource Systems

TU Wien, Vienna *3/2014 - 6/2014*

Projectassistant

EU-Project “WWWforEurope”

TU Wien, Vienna *8/2011 - 9/2011*

Projectassistant

EU-Project “WWWforEurope”

LKH, Steyr *8/2009 - 4/2010*

Civilianservice in patientstransport

Education

TU Wien, Vienna *4/2017 - present*

PhD studies in technical mathematics

Supervisor: Alexia Fürnkranz-Prskawetz

Stockholm University, Stockholm	8/2015 - 1/2016
Erasmus scholarship	
TU Wien, Vienna	11/2014 - 3/2017
Master studies Statistics and applied Mathematics in Economics	
Passed with distinction	
TU Wien, Vienna	10/2010 - 11/2014
Bachelor studies Statistics and applied Mathematics in Economics	
Passed with distinction	
BG/BRG Werndlpark, Steyr	2001 - 2009
Matura passed with distinction	

Teaching Experience

TU Wien, Vienna	10/2018 - 1/2019
Lecturer	
“Dynamic Macroeconomics” (Exercises)	
TU Wien, Vienna	10/2018 - 1/2019
Lecturer	
“Principles of Microeconomics” (Exercises)	
TU Wien, Vienna	3/2013 - 6/2013
Tutor	
“Econometrics for Business Informatics” (Lectures with exercises)	

Publications

Quantifying Economic Dependency.

Loichinger, E., Hammer, B., Prskawetz, A., Freiberger, M., & Sambt, J. (2017).
European Journal of Population, 1-30.

Language abilities

German	Native language
English	Fluent written and spoken
French	Matura level
Latin	Matura level

Technical skills

Programming	Matlab, Julia, GAMS, R, Stata, LaTeX, C, C++
Office Programs	Word, Excel, PowerPoint

Other Information

Club chairman*1/2016 - current*

Chairman of Honeybees Flagfootball
(playing in the first division of the Austrian flagfootball league)

FLA2-League manager*9/2017 - 11/2018*

League-manager of the second division in the Austrian flagfootball league