

Essays in Applied Microeconomics

Inaugural-Dissertation

zur Erlangung des Grades eines Doktors
der Wirtschafts- und Gesellschaftswissenschaften

durch

die Rechts- und Staatswissenschaftliche Fakultät der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Janoś Gabler

aus Crailsheim

Bonn

2022

Dekan: Prof. Dr. Jürgen von Hagen
Erstreferent: Prof. Dr. Hans-Martin von Gaudecker
Zweitreferent: Prof. Dr. Flavio Cunha
Tag der mündlichen Prüfung: 11.08.2022

Acknowledgements

First of all, I want to thank Hans-Martin von Gaudecker for being the best supervisor anyone could hope for during such a challenging experience as one's PhD. It is hard to put into words the many many ways Hans-Martin has influenced me in my research and how he has helped me to grow since I first met him in an econometrics class eight years ago. No matter how busy he was, he always found time for me. His support went far beyond what anyone could expect from a supervisor. The German word "Doktorvater" is a more accurate description. He also makes delicious espresso, which I enjoyed on many bike rides to his place.

I want to thank Flavio Cunha for being my second supervisor. Several of my projects, starting with my master thesis, are inspired by and build on his work. His input and feedback on those projects were invaluable. More importantly, he is an inspiration and role model as a researcher. I have never seen anyone more curious and passionate about discovering knowledge and understanding the drivers of skill formation than Flavio.

I also want to thank Joachim Freyberger for being on my committee and for the many interesting discussions we had about the identification of latent factor models. Using completely different ways of thinking, we always arrived at the same conclusions. Given the ongoing debates about this topic in the literature, this was a pleasant and refreshing experience.

I am thankful to my coauthors Philipp Eisenhauer, Hans-Martin von Gaudecker, Lena Janys, Jürgen Maurer, Mariam Petrosyan, Klara Röhl, and Tobias Raabe for all the challenges we have overcome together. Klara and Tobias are not only coauthors on the most daunting chapter of my dissertation but also close friends.

I want to thank all people involved in Open Source Economics and everyone who contributed to the amazing packages we have created over the years.

I am grateful as well for support from the Bonn Graduate School of Economics, IZA - Institute of Labor Economics, briq Institute on Behavior and Inequality, CRC 224 Transregio, TRA Modelling, and Google Cloud CoViD-19 research credits program.

Finally, I want to thank the people who are closest to me: my mother for her unconditional support and love and for raising me the way she did; my brother and sister for always being there when I needed it; and Mariam for sharing her life with me and being amazing.

Contents

Acknowledgements	iii
List of Figures	xii
List of Tables	xiv
Introduction	1
References	4
1 The Effectiveness of Testing, Vaccinations and Contact Restrictions for Containing the CoViD-19 Pandemic	7
1.1 Introduction	7
1.2 Model Description	8
1.3 Second and Third Waves of the Covid-19 Pandemic in Germany	11
1.4 Results	13
1.5 Discussion and Conclusions	16
Appendix 1.A Materials and Methods	17
1.A.1 Course of Disease	17
1.A.2 The Synthetic Population	20
1.A.3 Number of Contacts	22
1.A.4 Assortativity	24
1.A.5 Policies	25
1.A.6 Rapid Test Demand	28
1.A.7 Share of Detected Cases	31
1.A.8 PCR Testing and Behavioral Response	32
1.A.9 Estimated Parameters	34
1.A.10 Shapley Values	39
1.A.11 Overview Model Parameters	41
1.A.12 Reproducibility	46
Appendix 1.B Supplementary Text	46
1.B.1 Literature Review	46

1.B.2	Summary	48
1.B.3	Modeling Numbers of Contacts	49
1.B.4	Reducing Numbers of Contacts via NPIs	50
1.B.5	Matching Individuals	51
1.B.6	Course of Disease	54
1.B.7	Testing	54
1.B.8	Seasonality	56
1.B.9	Initial Conditions	57
1.B.10	Model Fit	58
1.B.11	Model Validation	63
1.B.12	Robustness to assumptions about rapid test sensitivity	66
1.B.13	Share of Detected Cases	69
1.B.14	Simulated Rapid Tests	69
1.B.15	Scenarios	73
	References	77
2	Structural Models for Policy-Making: Coping with Parametric Uncertainty	89
2.1	Introduction	89
2.2	Structural Models for Policy-Making	92
2.2.1	Uncertainty Sets	94
2.2.2	Statistical Decision Theory	95
2.3	Eckstein-Keane-Wolpin Models	98
2.3.1	General Structure	98
2.3.2	The Career Decisions of Young Men	99
2.3.3	Confidence Set Bootstrap	105
2.4	Results	106
2.4.1	General Subsidy	107
2.4.2	Targeted Subsidy	108
2.5	Conclusion	111
	Appendix 2.A Supplementary Material	112
2.A.1	Computation	112
2.A.2	Data	118
2.A.3	Results	121
	References	123
3	Mens Sana in Corpore Sano?	129
3.1	Introduction	129
3.2	Data and Measurements	134
3.2.1	Physical Capacity	135
3.2.2	Cognitive Capacity	138

3.2.3	Exercise and Cognitive Stimulation	139
3.2.4	Raw Correlations in the Data	140
3.2.5	Example Transitions	143
3.3	Model	144
3.3.1	The Technology of Aging	144
3.3.2	Identification and Interpretation of Parameters	146
3.3.3	Estimation	148
3.4	Results	149
3.4.1	Measurement System	149
3.4.2	Transition Equations	152
3.4.3	Dynamic Effects Over Several Periods	155
3.5	Conclusions and Outlook	157
Appendix 3.A Additional Background on the Data and Measurements		158
Appendix 3.B The Maximum Likelihood Estimator		160
3.B.1	State Estimation	160
3.B.2	The Likelihood Interpretation of the Kalman Filter	164
3.B.3	Numerical Stability	164
Appendix 3.C Detailed Model Setup		168
3.C.1	Background on Identification	168
Appendix 3.D Additional Tables and Figures for the Main Specification		169
3.D.1	Complete Set of Parameters of the Measurement System	169
3.D.2	Correlations Between Measurements and Factors	176
3.D.3	Factor Distributions	182
3.D.4	Transition Equations	188
3.D.5	Distributions of Initial Factors and of Shocks to Factors	196
Appendix 3.E Results for a Linearized Model		198
3.E.1	Measurement System	198
3.E.2	Transition Equations	205
References		211

List of Figures

1.2.1 A description of the model can be found in Supplementary Material B. Figure 1a shows the influence of an agent's contacts to other agents on infections. Demographic characteristics set the baseline number of contacts in different networks (η). The agents may reduce the number of contacts due to NPIs, showing symptoms, or testing positively for SARS-CoV-2 (τ). Infections may occur when a susceptible agent meets an infectious agent; the probability depends on the type of contact (β_c), on seasonality (α_c), and on NPIs ($\rho_{c,t}$). If infected, the infection progresses as depicted in Figure 1b. If rapid tests are available, agents' demand is modeled as in Figure 1c. All reasons trigger a test only for a fraction of individuals depending on an individual compliance parameter; the thresholds for triggering test demand differ across reasons and they may depend on calendar time ($\pi_{c,t}$ and $\tau_{c,t}$). Figure 1d shows the model of translating all infections in the simulated data to age-specific recorded infections. The model uses data on the aggregate share of recorded cases (ψ), the share of positive PCR tests triggered by symptoms ($\chi_{symptom}$), and the false positive rate of rapid tests ($p_{positive|infected, i, t}$). The lower part of the graph is relevant only for periods where rapid tests are available. All parameters are explained in Supplementary Material A.11. 9

1.3.1 Evolution of the pandemic, its drivers, and model fit, September 2020 to May 2021: Data sources are described in Supplementary Material A. Age- and region-specific analogues to Figure 2a can be found in Supplementary Material B.10. For legibility reasons, all lines in Figure 2b are rolling 7-day averages. The Oxford Response Stringency Index is scaled as $2 \cdot (1 - x/100)$, so that a value of one refers to the situation at the start of our sample period and zero means that all NPIs included in the index are turned on. The other lines in Figure 2b show the product of the effect of contact reductions, increased hygiene regulations, and seasonality. See Appendix A.5 for separate plots of the three factors by contact type. 12

1.4.1	The effect of different interventions on recorded and actual infections. The blue line in Figure 3a is the same as in Figure 2a and refers to our baseline scenario, so does the blue line in Figure 3b. The red lines refer to a situation where NPIs evolve as in the baseline scenario and the B.1.1.7 variant is introduced in the same way; vaccinations, rapid tests, and seasonality remain at their January levels. The other scenarios turn each of these three factors on individually. The decompositions in Figures 3c and 3d are based on Shapley values, which are explained more thoroughly in Appendix A.10. For legibility reasons, all lines are rolling 7-day averages.	14
1.5.1	Effects of different scenarios for policies regarding schools and workplaces. Blue lines in both figures refer to our baseline scenario; they are the same as in Figure 3b. Interventions start at Easter because there were no capacity constraints for rapid tests afterwards. For legibility reasons, all lines are rolling 7-day averages.	16
1.A.1	Vaccination Rates by Age Group	22
1.A.2	Number of Contacts of the Different Contact Types	23
1.A.3	Number of Household Contacts	24
1.A.4	Assortativity by Age Group for Non Recurrent Other and Work Contacts	25
1.A.5	The Contact Reduction Effects of School and Work Attendance Policies	27
1.A.6	Share of Individuals Doing a Rapid Test.	30
1.A.7	Share of Detected Cases in the Absence of Rapid Tests	32
1.A.8	Share of Positive PCR Tests Administered to Symptomatic Individuals	33
1.A.9	Daily share of infections by contact type	37
1.A.10	Share of Pre-Pandemic Other Contacts Taking Place with Infection Potential	38
1.B.1	Seasonality by Type of Contact	57
1.B.2	Model Fit of the Reported Cases and the Effective Replication Number	59
1.B.3	Simulated and Empirical Infections by Age Group	60
1.B.4	Simulated and Empirical Infections by Federal State	61
1.B.5	Model Fit of the Share of B.1.1.7 and Vaccinations	62
1.B.6	Share of Individuals With Rapid Tests	63
1.B.7	Rapid Test Introduction in the Three Scenarios	64
1.B.8	Out of Sample Prediction for Reported and Total Cases from March to June 2021.	65
1.B.9	Rapid test sensitivity profiles	68
1.B.10	Shapley decompositions for different values of rapid test sensitivity	68
1.B.11	Share of Detected Cases by Age Group	69
1.B.12	Rapid Test Shares in the Population by Channel	70
1.B.13	Simulated Rapid Test Statistics	72
1.B.14	False Positive and False Negative Rates by Channel	73
1.B.15	The Effect of Different Work Scenarios on Reported and Total Cases	74

1.B.16	The Effect of Different School Scenarios on Reported and Total Cases	75
1.B.17	The Role of Targeted and Compliance Driven Rapid Test Demand	76
2.2.1	Model comparison	94
2.2.2	Comparing policy predictions	96
2.2.3	Comparing policy regret	97
2.3.1	Timing of events	99
2.3.2	Decision tree	100
2.3.3	Data overview	103
2.3.4	Model fit	104
2.4.1	General subsidy	107
2.4.2	Time preference	108
2.4.3	Type heterogeneity	109
2.4.4	Targeted subsidy	109
2.4.5	Policy ranking	110
2.A.1	Sample size	119
2.A.2	Initial schooling	119
2.A.3	Average choices by initial schooling	120
2.A.4	Transition matrix	120
2.A.5	Model fit	121
2.A.6	Targeted subsidy for all types	122
2.A.7	Policy impact and time preference	122
3.2.1	Average measurements by age	137
3.2.2	Cross factor measurement correlations (female).	141
3.2.3	Cross factor measurement correlations (male).	142
3.2.4	Trajectories for decline of health and cognitive capacity	144
3.4.1	Next period states as a function of current states, other factors evaluated at the median	153
3.A.1	Standard deviation of measurements by age	159
3.D.1	Correlations across implied factors and measurement correlations – females aged 70	176
3.D.2	Correlations across implied factors and measurement correlations – females aged 80	177
3.D.3	Correlations across implied factors and measurement correlations – females aged 90	178
3.D.4	Correlations across implied factors and measurement correlations – males aged 70	179
3.D.5	Correlations across implied factors and measurement correlations – males aged 80	180
3.D.6	Correlations across implied factors and measurement correlations – males aged 90	181

xii | List of Figures

3.D.7	Factor distributions – females aged 70	182
3.D.8	Factor distributions – females aged 80	183
3.D.9	Factor distributions – females aged 90	184
3.D.10	Factor distributions – males aged 70	185
3.D.11	Factor distributions – males aged 80	186
3.D.12	Factor distributions – males aged 90	187
3.D.13	Transition equations for all factors (other factors evaluated at the median), females	188
3.D.14	Transition equations for all factors (other factors evaluated at the median), males	189
3.E.1	Transition equations (other factors evaluated at the median)	205

List of Tables

1.A.1	Estimated Parameters	35
1.A.2	Contacts, Matching and Policies	41
1.A.3	Infection Probabilities and Virus Variants	42
1.A.4	Disease and Vaccination Model	43
1.A.5	Rapid Testing	44
1.A.6	PCR Testing, Case Detection and Behavioral Response	45
2.A.1	Overview of Parameters in the Keane and Wolpin extended model	116
3.4.1	Loadings and Measurement Standard Deviations	150
3.4.2	6-year-ahead effects of changing exercise or cognitive stimulation, females	156
3.4.3	6-year-ahead effects of changing exercise or cognitive stimulation, males	156
3.D.1	Intercepts, Loadings, and Measurement Standard Deviations for Physical Capacity, Females	170
3.D.2	Intercepts, Loadings, and Measurement Standard Deviations for Physical Capacity, Males	171
3.D.3	Intercepts, Loadings, and Measurement Standard Deviations for Cognitive Capacity, Females	172
3.D.4	Intercepts, Loadings, and Measurement Standard Deviations for Cognitive Capacity, Males	173
3.D.5	Intercepts, Loadings, and Measurement Standard Deviations for Exercise, Females	174
3.D.6	Intercepts, Loadings, and Measurement Standard Deviations for Exercise, Males	174
3.D.7	Intercepts, Loadings, and Measurement Standard Deviations for Cognitive Stimulation, Females	175
3.D.8	Intercepts, Loadings, and Measurement Standard Deviations for Cognitive Stimulation, Males	175
3.D.9	Transition Parameters for Physical Capacity, Females	190
3.D.10	Transition Parameters for Physical Capacity, Males	191

3.D.11	Transition Parameters for Cognitive Capacity, Females	192
3.D.12	Transition Parameters for Cognitive Capacity, Males	193
3.D.13	Transition Parameters for Exercise, Females	194
3.D.14	Transition Parameters for Exercise, Males	194
3.D.15	Transition Parameters for Cognitive Stimulation, Females	195
3.D.16	Transition Parameters for Cognitive Stimulation, Males	195
3.D.17	Distribution of the initial states, females	196
3.D.18	Distribution of the intitial states, males	196
3.D.19	Standard deviations of shocks	197
3.E.1	Loadings and Measurement Standard Deviations for Physical Capacity, Females	199
3.E.2	Loadings and Measurement Standard Deviations for Physical Capacity, Males	200
3.E.3	Loadings and Measurement Standard Deviations for Cognitive Capacity, Females	201
3.E.4	Loadings and Measurement Standard Deviations for Cognitive Capacity, Males	202
3.E.5	Loadings and Measurement Standard Deviations for Exercise, Females	203
3.E.6	Loadings and Measurement Standard Deviations for Exercise, Males	203
3.E.7	Loadings and Measurement Standard Deviations for Cognitive Stimulation, Females	204
3.E.8	Loadings and Measurement Standard Deviations for Cognitive Stimulation, Males	204
3.E.9	Transition Parameters for Physical Capacity, Females	206
3.E.10	Transition Parameters for Physical Capacity, Males	206
3.E.11	Transition Parameters for Cognitive Capacity, Females	207
3.E.12	Transition Parameters for Cognitive Capacity, Males	207
3.E.13	Transition Parameters for Exercise, Females	208
3.E.14	Transition Parameters for Exercise, Males	208
3.E.15	Transition Parameters for Cognitive Stimulation, Females	209
3.E.16	Transition Parameters for Cognitive Stimulation, Males	209
3.E.17	Standard deviations of shocks	210

Introduction

The world is complex and full of challenges. Structural economists and other scientists create insights that help policy-makers to decide how to react to those challenges. Their toolkit for creating insights contains three building blocks: *Empirical data*, which are scarce and imperfect but provide the fundamental link between theoretical models and the real world. *Modeling*, which is the creative process of building less complex – wrong but useful – versions of the world. And *Computation*, which is necessary to fit models to data and simulate quantities of interest. Under the names *Observation*, *Theory*, and *Computation*, these steps are often called the three pillars of science.

The three go hand in hand and one component can often substitute for another: Doing more computations can relax the need for simplifying assumptions when creating models; large sample sizes allow data driven model selection; explicitly modeling decision-making of individuals makes it possible to use data that was not collected in randomized controlled trials.

This substitutability is important for two reasons: Firstly, the costs of the three components have developed very differently over the last decades. Datasets have increased in size but only at a slow rate; the cost of creative modeling is unlikely to ever change; computation is getting cheaper at an exponential rate. Secondly, there are sometimes hard restrictions. Collecting panel data, for example, is not only expensive but inherently takes a lot of time.

This dissertation consists of three independent research studies. All of them use empirical datasets, modeling, and computation to shed light on important topics. The first chapter develops an extremely detailed agent based model of disease transmission and uses it to quantify the effects of vaccinations, seasonality, and rapid testing on the spread of CoViD-19. The second chapter acknowledges that any outcome of a complex model comes with a lot of uncertainty. It embeds standard techniques for uncertainty quantification in a decision-theoretic framework. The third chapter uses methods from the skill formation literature to analyze the decline of health and cognitive capacity in the last years of life. Below, I provide a brief summary of chapters two and three. The summary of the first chapter is embedded in a short history of the project and the evolution of the pandemic in Germany.

Chapter 1: The Effectiveness of Testing, Vaccinations and Contact Restrictions for Containing the CoViD-19 Pandemic: We started working on the project in the summer of 2020. Due to a strict initial lockdown that was relaxed only slowly, the case numbers were low. However, it was not clear whether they would stay low when measures were to be relaxed further. It was also unknown whether seasonality would have an effect. There were no vaccinations and no rapid tests. Official contact tracing was seen as a primary reaction to potentially rising case numbers. The highest case numbers Germany had seen were less than 60 cases per million and there was hope that it would never get higher.

Our main motivation was to find optimal combinations of non-pharmaceutical interventions (such as contact tracing, PCR testing and contact reductions in specific sectors) to keep case numbers constant at a low level and to determine thresholds that would warrant a full lockdown. We had collected data on beliefs about the pandemic and compliance with policies that we hoped to feed into such a model. Moreover, we thought that it would be important to use as many data sources as possible to calibrate realistic contact networks and reactions to policies. It became apparent that common epidemiological models were not designed to leverage such different data sources. In addition, many of those models do not have policy invariant parameters, rendering them unsuitable for ex-ante evaluation of fine-grained policies.

At the core of our model are physical contacts between heterogeneous agents. Each contact between infectious and susceptible individuals bears the risk of transmitting the virus. Contacts occur in four networks: Within the household, at work, at school, or in other settings (leisure activities, grocery shopping, medical appointments, etc.). Some contacts recur regularly; others occur at random. Pre-pandemic contact diaries, census data and real-time mobility data are used to calibrate realistic contact networks. While the number of contacts is influenced by policies, the infection probabilities of each contact are policy invariant parameters that are estimated with the method of simulated moments from observed case numbers.

The first application of the model took place at the beginning of November 2020 (Dorn, Gabler, Gaudecker, Peichl, Raabe, et al., 2020). Case numbers had skyrocketed to about 180 per million and the so-called “lockdown light” took effect. It was a much more nuanced policy than previous lockdowns. Our model correctly predicted that the lockdown light would be enough to stop the exponential increase but not to bring case numbers down. Moreover, we predicted that schools do play a role for the transmission of the disease but that split class approaches combined with strict hygiene measures are almost as effective as complete school closures. Doing such fine-grained simulations of school policies was possible because we imposed almost no simplifying assumptions on schools. We simulated full classrooms that have classes with multiple teachers per day.

In December, the case numbers grew even higher and every new week brought new policies. It became essential to model all these policies in detail, so we could

continue to fit case numbers and use longer time series to better estimate the transmission probabilities. To the best of our knowledge, we are the only research group that fit a model to actual case numbers of an entire country over more than nine months.

Moreover, it became clear that the share of undetected cases had increased dramatically as contact tracing became infeasible due to the sheer number of cases. We thus added PCR testing to our model and calibrated the tests to fit available estimates of the share of unknown cases. In December and January, we published two policy reports to show the importance of private contact tracing during the Christmas holidays (Gabler, Raabe, Röhr, and Gaudecker, 2020) and the effect of working from home (Gabler, Raabe, Röhr, and Gaudecker, 2021). Our predicted effect of increasing the share of people working from home lined up well with reduced form estimates from other groups (Fadinger and Schymik, 2020).

At the beginning of March 2021, cases rose sharply due to the new Alpha variant. In April, they dropped suddenly, taking many forecasters by surprise. Vaccinations, freely available rapid tests and seasonality were the most likely explanations for the decline. We thus incorporated extremely detailed models of rapid testing and behavioral reactions to positive test results into the model. Moreover, we added the more contagious Alpha variant, vaccinations and an exogenously calibrated seasonality effect to our model. This specification is the one described in chapter 1. The results show that all three factors are important but that rapid tests play a much larger role than expected at the time. Extensive robustness checks – which we performed with data that only became available much later (see section 1.B.12) – show that this result is very robust even though our original assumptions on the sensitivity of rapid tests had been optimistic.

Chapter 2: Structural models for policy-making: Coping with parametric uncertainty Structural Microeconomists use complex models of economic agents making optimal decisions, subject to constraints. Usually, these models have parametric utility functions and the parameters of those functions are estimated by fitting the model to data. The estimated model is then used to investigate economic mechanisms, predict the impact of proposed policies, and inform optimal policy-making. Typically, the uncertainty around the estimated parameters is neither taken into account in the model predictions nor in the calculation of optimal policies.

We develop an approach that copes with parametric uncertainty in computational models and embeds model informed policy-making in a decision-theoretic framework. This approach treats the model as a black box and is thus not specific to microeconomic models. As an illustration, we apply the method to the seminal career decisions of young men model (Keane and Wolpin, 1997). The model is well known and akin to many contemporary dynamic discrete choice models. We document considerable uncertainty in the models' policy predictions and highlight the resulting policy recommendations obtained from using different formal rules of decision-making under uncertainty.

Chapter 3: Mens Sana in Corpore Sano? Development and maintenance of human capital throughout the life-cycle enables individuals to lead longer, more productive and more fulfilling lives. While there is a vast literature on the development of human capital during childhood, much fewer studies in economics have analyzed the decline of physical and cognitive capacities during later stages of life within a dynamic framework.

In this paper, we take a broad and systematic approach to model the interdependency of physical and cognitive capacity over individuals' later part of the life-cycle. To this end, we adapt the *Technology of Skill Formation* model by Cunha, Heckman, and Schennach (2010) to the context of aging. Our model has four latent factors: physical capacity, cognitive capacity, physical exercise, and cognitive stimulation. While the latent factors are not observable, the model allows us to identify their joint distribution from their effects on observable variables. Moreover, the model enables us to estimate dynamic relationships between the latent factors that are uncontaminated by measurement error.

We incorporate a simple model of mortality into the model to remove the selection bias that would arise from ignoring that mortality carries information about latent health. We estimate the model parameters with a maximum likelihood estimator on data from the Health and Retirement Study. The estimator is similar to the one used by Cunha, Heckman, and Schennach (2010), but we were forced to derive a numerically more robust version due to challenges that stem from the large number of periods and high correlation between factors. The estimation is done separately for men and women to account for gender differences in life expectancy and aging.

Our key findings are as follows: 1) There is substantial noise in all observed variables. While most measurements have a high correlation with the latent factor they measure, no single measurement dominates to an extent where it would be justified to just use a single variable and ignore the measurement error in the econometric analysis. 2) Despite a strong decline in means for physical and cognitive capacity, the rank order of these latent factors is remarkably stable. 3) Physical and cognitive capacity can be influenced by investments until very high ages. Cognitive stimulation is a specific investment into cognitive capacity. Physical exercise has a larger effect on physical capacity and a small effect on cognitive capacity.

References

- Cunha, Flavio, James J. Heckman, and Susanne Schennach.** 2010. "Estimating the technology of cognitive and noncognitive skill formation." *Econometrica* 78(3): 883–931. [4]
- Dorn, Florian, Janoś Gabler, Hans-Martin von Gaudecker, Andreas Peichl, Tobias Raabe, and Klara Röhr.** 2020. "Wenn Menschen (keine) Menschen treffen: Simulation der Auswirkungen von Politikmaßnahmen zur Eindämmung der zweiten Covid-19-Welle." IFO Schnelldienst. [2]

- Fadinger, Harald, and Jan Schymik.** 2020. "The Effects of Working From Home on COVID-19 Infections and Production - A Macroeconomic Analysis for Germany." CRC TR 224 Discussion Paper Series. University of Bonn, and University of Mannheim, Germany. [3]
- Gabler, Janoś, Tobias Raabe, Klara Röhl, and Hans-Martin von Gaudecker.** 2020. "Die Bedeutung individuellen Verhaltens über den Jahreswechsel für die Weiterentwicklung der Covid-19-Pandemie in Deutschland." IZA Standpunkte. [3]
- Gabler, Janoś, Tobias Raabe, Klara Röhl, and Hans-Martin von Gaudecker.** 2021. "Der Effekt von Homeoffice auf die Entwicklung der Covid-19-Pandemie in Deutschland." IZA Standpunkte. [3]
- Keane, Michael P., and Kenneth I. Wolpin.** 1997. "The career decisions of young men." *Journal of Political Economy* 105 (3): 473–522. [3]

Chapter 1

The Effectiveness of Testing, Vaccinations and Contact Restrictions for Containing the CoViD-19 Pandemic

Joint with Tobias Raabe, Klara Röhl and Hans-Martin von Gaudecker

1.1 Introduction

Since early 2020, the CoViD-19 pandemic has presented an enormous challenge to humanity on many dimensions. The development of highly effective vaccines holds the promise of containment in the medium term. However, most countries find themselves many months—and often years—away from reaching vaccination levels that would end the pandemic or even protect the most vulnerable (Mathieu, Ritchie, Ortiz-Ospina, Roser, Hasell, et al., 2021). In the meantime, it is of utmost importance to employ an effective mix of strategies for containing the virus. The most frequent initial response was a set of non-pharmaceutical interventions (NPIs) to reduce contacts between individuals. While this has allowed some countries to sustain equilibria with very low infection number—see Contreras, Dehning, Mohr, Bauer, Spitzner, et al. (2021) for a theoretical equilibrium at low case numbers which is sustained with test-trace-and-isolate policies—, most have seen large fluctuations of infection rates over time. Containment measures have become increasingly diverse and now include rapid testing, more nuanced NPIs, and contact tracing. Neither

* The authors are grateful for support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1– 390838866 – and through CRC-TR 224 (Projects A02 and C01), by the IZA Institute of Labor Economics, and by the Google Cloud CoViD-19 research credits program.

these policies' effects nor the influence of seasonal patterns or of more infectious virus strains are well understood in quantitative terms.

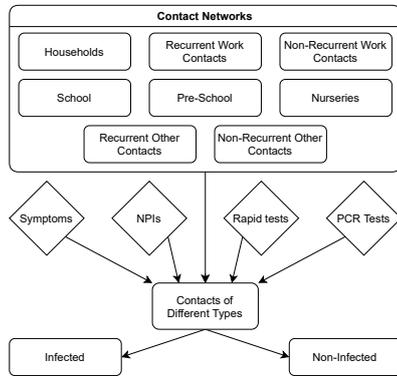
This paper develops a quantitative model incorporating these factors simultaneously. The framework allows to combine a wide variety of data and mechanisms in a timely fashion, making it useful to predict the effects of various interventions. Behavioral reactions to symptoms or positive tests are explicitly taken into account. We apply the model to Germany, where new infections fell by almost 80% during May 2021. Our analysis shows that, aside from seasonality, frequent and large-scale rapid testing caused the bulk of this decrease, which is in line with prior predictions (Mina and Andersen, 2021). We conclude that it should have a large role for at least as long as vaccinations have not been offered to an entire population.

1.2 Model Description

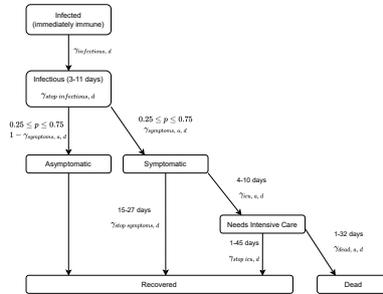
At the core of our agent-based model (Aleta, Martín-Corral, Piontti, Ajelli, Litvinova, et al., 2020; Hinch, Probert, Nurtay, Kendall, Wymant, et al., 2021)—we review more literature in Supplementary Material B.1—are physical contacts between heterogeneous agents (Figure 1a). Each contact between an infectious individual and somebody susceptible to the disease bears the risk of transmitting the virus. Contacts occur in up to four networks: Within the household, at work, at school, or in other settings (leisure activities, grocery shopping, medical appointments, etc.). Some contacts recur regularly, others occur at random. Empirical applications can take the population and household structure from census data and the network-specific frequencies of contacts from diary data measuring contacts before the pandemic (Mossong, Hens, Jit, Beutels, Auranen, et al., 2008; Hoang, Coletti, Melegaro, Wallinga, Grijalva, et al., 2019). Within each network, meeting frequencies depend on age and geographical location (see Supplementary Material A.4).

The four contact networks are chosen so that the most common NPIs can be modeled in great detail. NPIs affect the number of contacts or the risk of transmitting the disease upon having physical contact. The effect of different NPIs will generally vary across contact types. For example, a mandate to work from home will reduce the number of work contacts to zero for a fraction of the working population. Schools and daycare can be closed entirely, operate at reduced capacity—including an alternating schedule—, or implement mitigation measures like masking requirements or air filters (Lessler, Grabowski, Grantz, Badillo-Goicoechea, Metcalf, et al., 2021). Curfews may reduce the number of contacts in settings outside of work and school. In any setting, measures like masking requirements would reduce the probability of infection associated with a contact (Cheng, Ma, Witt, Rapp, Wild, et al., 2021).

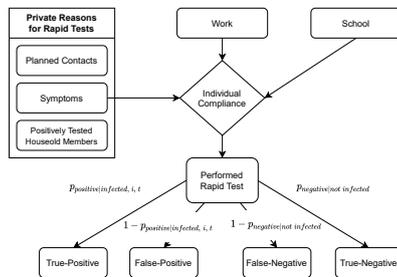
In the model, susceptibility to contracting the SARS-CoV-2 virus is dependent on age (Davies, Klepac, Liu, Prem, Jit, et al., 2020; Goldstein, Lipsitch, and Cevik, 2020). A possible infection progresses as shown in Figure 1b. We differentiate be-



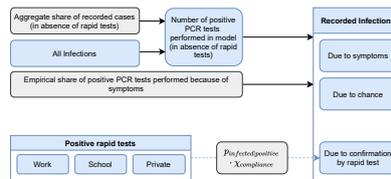
a. Model for contacts and infections



b. Disease progression



c. Model for rapid tests



d. Translating all infections to recorded ones

Figure 1.2.1. A description of the model can be found in Supplementary Material B. Figure 1a shows the influence of an agent’s contacts to other agents on infections. Demographic characteristics set the baseline number of contacts in different networks (η). The agents may reduce the number of contacts due to NPIs, showing symptoms, or testing positively for SARS-CoV-2 (τ). Infections may occur when a susceptible agent meets an infectious agent; the probability depends on the type of contact (β_c), on seasonality (κ_c), and on NPIs ($\rho_{c, t}$). If infected, the infection progresses as depicted in Figure 1b. If rapid tests are available, agents’ demand is modeled as in Figure 1c. All reasons trigger a test only for a fraction of individuals depending on an individual compliance parameter; the thresholds for triggering test demand differ across reasons and they may depend on calendar time ($\pi_{c, t}$ and $\tau_{c, t}$). Figure 1d shows the model of translating all infections in the simulated data to age-specific recorded infections. The model uses data on the aggregate share of recorded cases (ψ), the share of positive PCR tests triggered by symptoms ($\chi_{symptom}$), and the false positive rate of rapid tests ($p_{positive|infected, i, t}$). The lower part of the graph is relevant only for periods where rapid tests are available. All parameters are explained in Supplementary Material A.11.

tween an initial period of infection without being infectious or showing symptoms, being infectious (presymptomatic or asymptomatic), showing symptoms, requiring

intensive care, and recovery or death (similar to Aleta et al. (2020)). The probabilities of transitioning between these states depend on age; their duration is random and calibrated to medical literature (for a detailed description see Supplementary Material A.1). Conditional on the type of contact, infectiousness is independent of age (Jones, Biele, Mühlemann, Veith, Schneider, et al., 2021).

The model includes several other features, which are crucial to describe the evolution of the pandemic in 2020-2021. New virus strains with different infectiousness profiles may appear. Vaccines may become available. During the vaccine roll-out, priority may depend on age and occupation; vaccine hesitancy is modelled by some individuals refusing vaccination offers. With some probability, vaccinated agents become immune and do not transmit the virus (Hunter and Brainard, 2021; Levine-Tiefenbrun, Yelin, Katz, Herzog, Golan, et al., 2021; Petter, Mor, Zuckerman, Oz-Levi, Younger, et al., 2021; Pritchard, Matthews, Stoesser, Eyre, Gethings, et al., 2021).

We include two types of tests. Polymerase chain reaction (PCR) tests reveal whether an individual is infected or not; there is no uncertainty to the result. PCR tests require at least one day to be processed and there are aggregate capacity constraints. In contrast, rapid antigen tests yield immediate results. Specificity and sensitivity of these tests is set according to data analyzed in Brümmer, Katzenschlager, Gaeddert, Erdmann, Schmitz, et al. (2021), Özcürümez, Katsounas, Holdenrieder, Meyer, Renz, et al. (2021), and Scheiblauer, Filomena, Nitsche, Puyskens, Corman, et al. (2021); sensitivity depends on the timing of the test relative to the onset of infectiousness as in Smith, Gibson, Martinez, Ke, Mirza, et al. (2021). We analyse robustness to different assumptions in Supplementary Material B.12. After a phase-in period, all tests that are demanded will be performed. Figure 1c shows our model for rapid test demand. Schools may require staff and students to be tested regularly. Rapid tests may be offered by employers to on-site workers. Individuals may demand tests for private reasons, which include having plans to meet other people, showing symptoms of CoViD-19, and a household member having tested positively for the virus. We endow each agent with an individual compliance parameter. This parameter determines whether she takes up rapid tests. Positive test results or symptoms lead most individuals to reduce their contacts; this is why tests impact the actual contacts in Figure 1.

Modelling a population of agents according to actual demographic characteristics means that we can use a wide array of data to identify and calibrate the model's many parameters (see Supplementary Material A for a complete description). Contact diaries yield pre-pandemic distributions of contacts for different contact types and their assortativity by age group. Mobility data is used to model the evolution of work contacts. School and daycare policies can be incorporated directly from official directives. Administrative records on the number of tests, vaccinations by age and region, and the prevalence of virus strains are generally available. Surveys may ask about test offers, propensities to take them up, and past tests. Other studies' estimates of the seasonality of infections can be incorporated directly. The remaining

parameters—most notably, these include infection probabilities by contact network and the effects of some NPIs, see Supplementary Material A.9—will be chosen numerically so that the model matches features of the data (see McFadden (1989) for the general method). In our application, we keep the number of free parameters low in order to avoid overfitting. The data features to be matched include official case numbers for each age group and region, deaths, and the share of the B.1.1.7 strain.

The main issue with official case numbers is that they will contain only a fraction of all infections. In the German case, this specifically amounts to positive PCR tests. We thus model recorded cases as depicted in Figure 1d. We take mortality-based aggregate estimates of the share of detected cases and use data on the share of PCR tests administered because of CoViD-19 symptoms. As the share of asymptomatic individuals varies by age group, this gives us age-specific shares (see Figure B.11). Our estimates suggest that—in the absence of rapid testing—the detection rate is 80% higher on average for individuals above age 80 compared to school age children. Once rapid test become available, confirmation of a positive result is another reason leading to positive PCR tests.

1.3 Second and Third Waves of the Covid-19 Pandemic in Germany

The model is applied to the second and third waves of the CoViD-19 pandemic in Germany, covering the period mid-September 2020 to the end of May 2021. Figure 2 describes the evolution of the pandemic and of its drivers. The black line in Figure 2a shows officially recorded cases; the black line in Figure 2b the Oxford Response Stringency Index (Hale, Atav, Hallas, Kira, Phillips, et al., 2020), which tracks the tightness of non-pharmaceutical interventions. The index is shown for illustration of the NPIs, we never use it directly. For legibility reasons, we transform the index so that lower values represent higher levels of restrictions. A value of zero means all measures incorporated in the index are turned on. The value one represents the situation in mid-September, with restrictions on gatherings and public events, masking requirements, but open schools and workplaces. In the seven weeks between mid September and early November, cases increased by a factor of ten. Restrictions were somewhat tightened in mid-October and again in early November. New infections remained constant throughout November before rising again in December, prompting the most stringent lockdown to this date. Schools and daycare centers were closed, so were customer-facing businesses except for grocery and drug stores. From the peak of the second wave just before Christmas until the trough in mid-February, newly detected cases decreased by almost three quarters. The third wave in the spring of 2021 is associated with the B.1.1.7 (Alpha) strain, which became dominant in March (Figure 2c). Note that we do not model B.1.617.2 (Delta). That variant was first detected in Germany in April; at the end of our simulation pe-

12 | 1 The Effectiveness of Testing, Vaccinations and Contact Restrictions

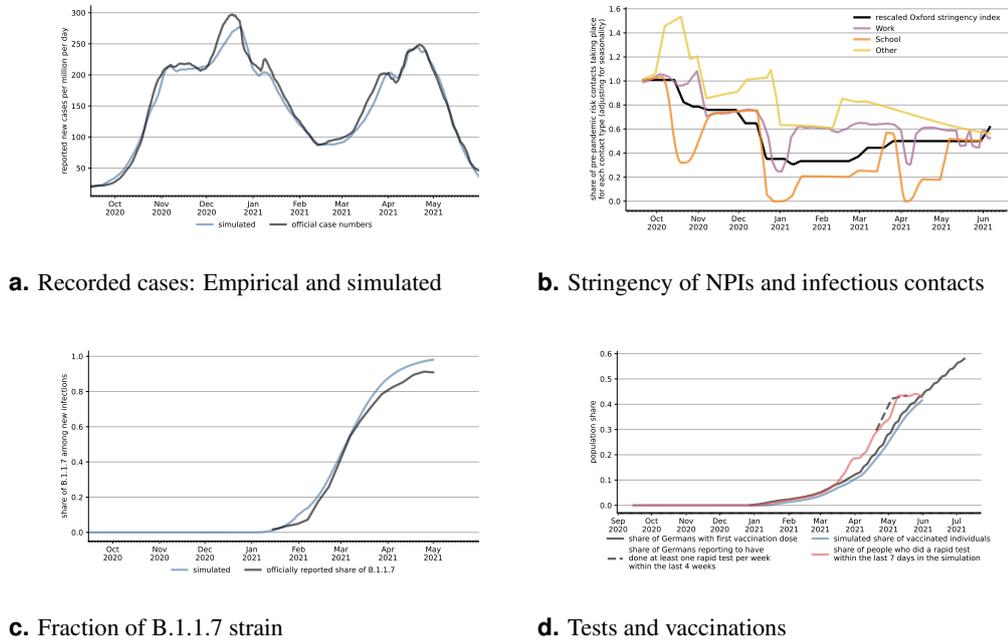


Figure 1.3.1. Evolution of the pandemic, its drivers, and model fit, September 2020 to May 2021: Data sources are described in Supplementary Material A. Age- and region-specific analogues to Figure 2a can be found in Supplementary Material B.10. For legibility reasons, all lines in Figure 2b are rolling 7-day averages. The Oxford Response Stringency Index is scaled as $2 \cdot (1 - x/100)$, so that a value of one refers to the situation at the start of our sample period and zero means that all NPIs included in the index are turned on. The other lines in Figure 2b show the product of the effect of contact reductions, increased hygiene regulations, and seasonality. See Appendix A.5 for separate plots of the three factors by contact type.

riod it accounted for less than 5% of cases. In early March, some NPIs were relaxed; e.g., hairdressers and home improvement stores were allowed to open again to the public. There were many changes in details of regulations afterwards, but they did not change the overall stringency index.

By March 2021, the set of policy instruments had become much more diverse. Around the turn of the year, the first people were vaccinated with a focus on older age groups and medical staff (Figure 2d). Until the end of May, 43% had received at least one dose of a vaccine. In late 2020, rapid tests started to replace regular PCR tests for staff in many medical and nursing facilities. These had to be administered by medical doctors or in pharmacies. At-home tests approved by authorities became available in mid-March. Rapid test centers were opened, and one test per person and week was made available free of charge. In several states, customers were only allowed to enter certain stores with a recent negative rapid test result. These devel-

opments are characteristic of many countries: The initial focus on NPIs to slow the spread of the disease has been accompanied by vaccines and a growing acceptance and use of rapid tests. At broadly similar points in time, novel strains of the virus have started to pose additional challenges.

1.4 Results

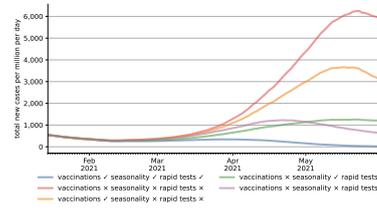
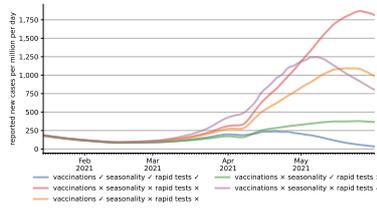
We draw simulated samples of agents from the population structure in September 2020 and use the model to predict recorded infection rates until the end of May 2021. See Supplementary Materials A.2 and B.9 for details. The blue line in Figure 2a shows that our model's predictions are very close to officially recorded cases in the aggregate. This is also true for infections by age and geographical region (see Supplementary Material B.10).

The effects of various mechanisms can be disentangled due to the distinct temporal variation in the drivers of the pandemic. Next to the stringency index, the three lines in Figure 2b summarize how contact reductions, increased hygiene regulations, and seasonality evolved since early September for each of the three broad contact networks. For example, a value of 0.75 for the work multiplier means that if the environment was the same as in September (levels of infection rates, no rapid tests or vaccinations, only the wildtype virus present), infections at the workplace would be reduced by 25%. Two aspects are particularly interesting. First, all lines broadly follow the stringency index and they would do so even more if we left out seasonality and school vacations (roughly the last two weeks of October, two weeks each around Christmas and Easter, and some days in late May). Second, the most stringent regulations coincide with the period of decreasing infection rates between late December 2020 and mid-February 2021. The subsequent reversal of the trend is associated with the spread of the B.1.1.7 variant. During the steep drop in recorded cases during May 2021, for 42% of the population took at least one rapid tests per week, the first-dose vaccination rate rose from 28% to 43%, and seasonality lowered the relative infectiousness of contacts.

In order to better understand the contributions of rapid tests, vaccinations, and seasonality on the evolution of infections in 2021, Figure 1.4.1 considers various scenarios. NPIs are always held constant at their values in the baseline scenario. Figure 3a shows the model fit (the blue line, same as in Figure 2a), a scenario without any of the three factors (red line), and three scenarios turning each of these factors on individually. Figure 3b does the same for total infections in the model. Figure 3c employs Shapley values (Shapley, 2016) to decompose the difference in total infections between the scenario without any of the three factors and our main specification.

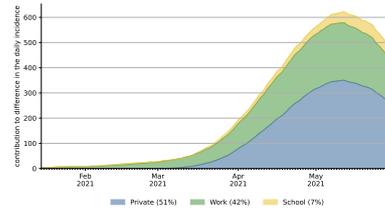
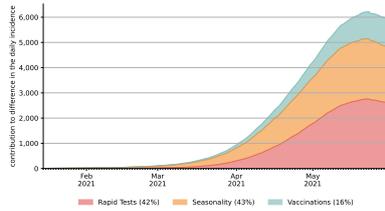
Until mid-March, there is no visible difference between the different scenarios. Seasonality hardly changes, and only few vaccinations and rapid tests were admin-

14 | 1 The Effectiveness of Testing, Vaccinations and Contact Restrictions



a. Recorded cases: 2021 scenarios

b. Total cases: 2021 scenarios



c. Decomposition of the difference between the scenario without any of the three factors and the main scenario in Figure 3b.

d. Decomposition of the difference between the scenario without rapid tests and the main scenario in Figure 3b.

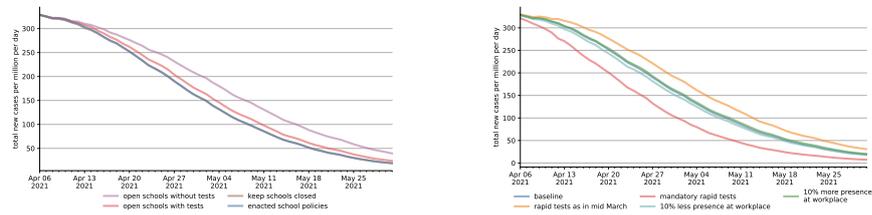
Figure 1.4.1. The effect of different interventions on recorded and actual infections. The blue line in Figure 3a is the same as in Figure 2a and refers to our baseline scenario, so does the blue line in Figure 3b. The red lines refer to a situation where NPIs evolve as in the baseline scenario and the B.1.1.7 variant is introduced in the same way; vaccinations, rapid tests, and seasonality remain at their January levels. The other scenarios turn each of these three factors on individually. The decompositions in Figures 3c and 3d are based on Shapley values, which are explained more thoroughly in Appendix A.10. For legibility reasons, all lines are rolling 7-day averages.

istered. Even thereafter, the effect of the vaccination campaign is surprisingly small at first sight. Whether considering recorded or total infections with only one channel active, the final level is always the highest in case of the vaccination campaign (orange lines). The Shapley value decomposition shows that vaccinations contribute 16% to the cumulative difference between scenarios. Reasons for the low share are the slow start—it took until March 24th until 10% of the population had received their first vaccination, the 20% mark was reached on April 19th—and the focus on older individuals. These groups contribute less to the spread of the disease than others due to a lower number of contacts. By the end of our study period, when first-dose vaccination rates reached 43% of the population, the numbers of new cases would have started to decline. It is important to note that the initial focus of the campaign was to prevent deaths and severe disease. Indeed, the case fatality rate was considerably lower during the third wave when compared to the second (4.4% between October and February and 1.4% between March and the end of May).

Seasonality has a large effect in slowing the spread of SARS-CoV-2. By May 31, both observed and total cases would be reduced by a factor of four if only seasonality mattered. However, in this period, cases would have kept on rising throughout, just at a much lower pace this is in line with results in Gavenčiak, Monrad, Leech, Sharma, Mindermann, et al. (2021), which our seasonality measure is based on. Nevertheless, we estimate seasonality to be a quantitatively important factor determining the evolution of the pandemic, explaining most of the early changes and 43% of the cumulative difference by the end of May.

A similar-sized effect—42% in the decomposition—comes from rapid testing. Here, it is crucial to differentiate between recorded cases and actual cases. Additional testing means that additional infections will be recorded which would otherwise remain undetected. Figure 3a shows that this effect is large and may persist for some time. Until late April, recorded cases are higher in the scenario with rapid testing alone when compared to the setting where none of the three mechanisms are turned on. The effect on total cases, however, is visible immediately in Figure 3b. Despite the fact that only 10% of the population performed weekly rapid tests in March on average, new infections on April 1 would have been reduced by 53% relative to the scenario without vaccinations, rapid tests, or seasonality. In Supplementary Material B.12, we provide a detailed analysis of whether our results are robust regarding the sensitivity parameters we assume for rapid tests. Even if we take a pessimistic stance, the effect is only reduced from 42% to 38%.

So why is rapid testing so effective? In order to shed more light on this question, Figure 3d decomposes the difference in the scenario without rapid tests and the main specification into the three channels for rapid tests. Tests at schools have the smallest effect, which is largely explained by schools not operating at full capacity during our period of study and the relatively small number of students. (18% of our population are in the education sector, e.g., pupils, teachers; 46% are workers outside the education sector.) Almost 40% come from tests at the workplace. Despite the fact that rapid tests for private reasons are phased in only in mid-March, they make up for more than half of the total effect. The reason lies in the fact that a substantial share of these tests is driven by an elevated probability to carry the virus, i.e., showing symptoms of CoViD-19 or following up on a positive test of a household member. The latter is essentially a form of contact tracing, which has been shown to be very effective (Kretzschmar, Rozhnova, Bootsma, Boven, Wiggert, et al., 2020; Contreras et al., 2021; Fetzer and Graeber, 2021). Indeed, a deeper analysis in Supplementary Material B.15 shows that the same amount of rapid tests administered randomly in the population would not have been nearly as effective.



a. Effects of different schooling scenarios

b. Effects of different work scenarios

Figure 1.5.1. Effects of different scenarios for policies regarding schools and workplaces. Blue lines in both figures refer to our baseline scenario; they are the same as in Figure 3b. Interventions start at Easter because there were no capacity constraints for rapid tests afterwards. For legibility reasons, all lines are rolling 7-day averages.

1.5 Discussion and Conclusions

Having quantified the effects of various mechanisms, we now simulate hypothetical scenarios comparing changes in NPIs and testing regimes. Two of the most contentious NPIs concern schools and mandates to work from home. In many countries, schools switched to remote instruction during the first wave, so did Germany. After the summer break, they were operating at full capacity with increased hygiene measures, before being closed again from mid-December onward. Some states started opening them gradually in late February, but operation at normal capacity did not resume until the beginning of June. Figure 4a shows the effects of different policies regarding schools starting after Easter, at which point rapid tests had become widely available. We estimate the realized scenario to have essentially the same effect as a situation with closed schools. Under fully opened schools with mandatory tests, total infections would have been 6% higher; this number rises to 20% without tests. These effect sizes are broadly in line with empirical studies (e.g. Berger, Fritz, and Kauer- mann (2021) and Vlachos, Hertegård, and B. Svaleryd (2021), see Supplementary Material B.11 for a comparison). In light of the large negative effects school closures have on children and parents (Luijten, Muilekom, Teela, Polderman, Terwee, et al., 2021; Melegari, Giallonardo, Sacco, Marcucci, Orecchio, et al., 2021)—and in particular on those with low socio-economic status—these results in conjunction with hindsight bias suggest that opening schools combined with a testing strategy would have been beneficial. In other situations, and in particular when rapid tests are not available at scale, trade-offs may well be different.

Figure 4b shows that with a large fraction of workers receiving tests, testing at the workplace has larger effects than mandating employees to work from home. Whether the share of workers working at the usual workplace is reduced or increased by ten percent changes infection rates by 2.5% or less in either direction. Making

testing mandatory twice a week—assuming independent compliance by employers and workers of 95% each—would have reduced infections by 23%. Reducing rapid tests offers by employers to the level of March would have increased infections by 13%.

Our analysis has shown that during the transition to high levels of vaccination and possibly thereafter, large-scale rapid testing can substitute for some NPIs. This comes at a fraction of the cost. A week of the fairly strict lockdown in early 2021 is estimated to have cost around 50 Euros per capita (Wollmershäuser, 2021); retail prices for rapid tests were below one Euro in early June 2021 and below five Euros for firms. While we do not distinguish between self-administered rapid tests and point of care rapid tests, the former are likely to play a larger role for indication-driven testing. Widespread availability at low prices seems important. However, they rely on purely voluntary participation in a non-public setting. The benefit of point-of-care rapid tests as a precondition to participate in leisure activities as well as mandatory tests at the workplace or at school come from screening the entire population. This is important because disadvantaged groups are less likely to be reached by testing campaigns relying on voluntary participation (e.g. Stillman and Tonin (2022)); at the same time, these groups have a higher risk to contract CoViD-19 (Robert Koch-Institut, 2021a). Mandatory tests at school and at the workplace will extend more into these groups. The same goes for individuals who exhibit a low level of compliance with CoViD-19-related regulations. Compared to vaccinations, rapid testing programmes allow a much quicker roll-out, making it arguably the most effective tool to contain the pandemic in the short run.

Appendix 1.A Materials and Methods

The model is described by a large number of parameters that govern the number of contacts a person has, the reduction in contacts due to NPIs, the demand for rapid tests and PCR tests, the likelihood of becoming infected on each contact, the likelihood of developing light or strong symptoms or even dying from the disease as well as the duration each stage of the disease takes.

1.A.1 Course of Disease

This section discusses the parameters governing the course of disease, their sources and how we arrived at the distributions used in the paper. See Figure 1.2.1 for a summary of our disease progression model.

The first stage of any disease is the infection. As detailed in Equation 1.B.1 the infection probability depends on the contact type, the calendar date to determine the seasonality, the age group of the susceptible person and the variant the infectious person is carrying. The base infection probability of each contact type (β_c) is estimated inside our model (Section 1.A.9). How we model and calibrate the sea-

sonality effect is detailed in Section 1.B.8. For the susceptibility of each age group (ζ_a) we take the estimates of Davies, Klepac, et al. (2020) (Extended Data Fig. 4). Lastly, we calibrate the infectiousness of the B.1.1.7 variant ($\sigma_{B.1.1.7}$) from Davies, Abbott, Barnard, Jarvis, Kucharski, et al. (2021) to 1.67.

We denote the latent period—i.e., the time span between infection and the start of infectiousness—by $\gamma_{infectious}$. Zhao, Tang, Musa, Ma, Zhang, et al. (2021) estimate the latent period to last 3.3 days (95% CI: 0.2, 7.9) on average. In line with this estimate our latent period lasts one to five days.

Once individuals become infectious, a share of them goes on to develop symptoms while others remain asymptomatic. We rely on data by Davies, Klepac, et al. (2020) for the age-dependent probability to develop symptoms. It varies from 25% for children and young adults to nearly 70% for the elderly. Similar to Peak, Kahn, Grad, Childs, Li, et al. (2020) and in line with He, Lau, Wu, Deng, Wang, et al. (2020) we set the length of the presymptomatic stage of age group a , $\gamma_{symptoms,a}$ to be one or two days. The probability to become symptomatic for age group a is split equally between one and two days. This combined with our latency period leads to an incubation period that is in line with the meta analysis by McAloon, Collins, Hunt, Barber, Byrne, et al. (2020).

We assume that the duration of infectiousness ($\gamma_{stop\ infectious}$) is the same for both symptomatic and asymptomatic individuals as evidence suggests little differences in the transmission rates between symptomatic and asymptomatic patients (Yin and Jin (2020)) and that the viral load between symptomatic and asymptomatic individuals are similar (Zou, Ruan, Huang, Liang, Huang, et al. (2020), Byrne, McEvoy, Collins, Hunt, Casey, et al. (2020), Singanayagam, Patel, Charlett, Bernal, Saliba, et al. (2020)). Our distribution of the duration of infectiousness is based on Byrne et al. (2020). For symptomatic cases they arrive at zero to five days before symptom onset (see their figure 2) and three to eight days of infectiousness afterwards.¹ Excluding the most extreme combinations, we arrive at 3 to 11 days as the duration of infectiousness.

We use the duration to recovery of mild and moderate cases reported by Bi, Wu, Mei, Ye, Zou, et al. (2020, Figure S3, Panel 2) for the duration of symptoms for non-ICU requiring symptomatic cases ($\gamma_{stop\ symptoms}$). We only disaggregate by age how likely individuals are to require intensive care.

For the time from symptom onset until need for intensive care we rely on data by Stokes, Zambrano, Anderson, Marder, Raz, et al. (2020)) and Hinch, Probert, Nurtay, Kendall, Wymatt, et al. (2021) ($\gamma_{icu,a}$). For those who will require intensive care we follow Chen, Qi, Liu, Ling, Qian, et al. (2020) who estimate the time from symptom onset to ICU admission as 8.5 ± 4 days. This aligns well with num-

1. Viral loads may be detected much later but eight days seems to be the time after which most people are culture negative, as also reported by Singanayagam et al. (2020).

bers reported for the time from first symptoms to hospitalization: Gaythorpe, Imai, Cuomo-Dannenburg, Baguelin, Bhatia, et al. (2020) report a mean of 5.76 with a standard deviation of four. We assume that the time between symptom onset and ICU takes four, six, eight or ten days with equal probabilities.

We take the survival probabilities and time to death and time until recovery ($\gamma_{stop\ icu\ a}$ and $\gamma_{dead, a}$) from intensive care from Hinch, Probert, Nurtay, Kendall, Wymatt, et al. (2021). They report time until death to have a mean of 11.74 days and a standard deviation of 8.79 days. To match this we discretize that 41% of individuals who will die from Covid-19 do so after one day in intensive care, 22% die after twelve days, 29% after 20 days and 7% after 32 days. Again, we rescale this for every age group among those that will not survive. For survivors Hinch, Probert, Nurtay, Kendall, Wymatt, et al. (2021) reports a mean duration of 18.8 days until recovery and a standard deviation of 12.21 days. We discretize this such that of those who recover in intensive care, 22% do so after one day, 30% after 15 days, 28% after 25 days and 18% after 45 days.

Individuals can become immune either through infection (γ_{immune}) or vaccination ($\gamma_{vacc, d}$). As reinfections are very rare (Abu-Raddad, Chemaitelly, Malek, Ahmed, Mohamoud, et al., 2020), we set the immunity period to one year with probability one, i.e. everyone that has been infected enjoys immunity for the rest of the simulation period.

The second route to immunity is through vaccination. Germany has mostly relied on the Pfizer-BioNTech BNT162b2 and Oxford-AstraZeneca ChAdOx1-S vaccines with smaller shares of the Moderna and Johnson&Johnson vaccines (impf-dashbord.de, 2021). As Pritchard et al. (2021) and Harris, Hall, Zaidi, Andrews, Dunbar, et al. (2021) find no difference in the effectiveness between the two most common vaccines, we do not distinguish between vaccine types.

Immunity is binary in our model, i.e. individuals achieve either sterile immunity or remain susceptible. Thus, we cannot simply use the reported effectiveness but must also include the risk of asymptomatic and sub-clinical reinfection among the vaccinated in our probability to become immune upon vaccination. This is important as there is ample evidence by now that vaccinated individuals can still get infected with SARS-CoV-2 and transmit the disease (Harris, Hall, et al., 2021; Levine-Tiefenbrun et al., 2021; Petter et al., 2021).

The reported effectiveness for BNT162b2 is estimated to be 90% 21 days after the first shot (Hunter and Brainard, 2021). The effectiveness does not increase much through the booster shot as Thompson, Burgess, Naleway, Tyner, Yoon, et al. (2021) report 90% (95% CI = 68%–97%) effectiveness against PCR-confirmed infections after two doses for mRNA vaccines in general. We therefore do not distinguish between the first and the booster shot.

On the other hand, Lipsitch and Kahn (2021) report a lower bound on transmission for the very similar Moderna vaccine of 61%.

To strike a middle ground we assume that 75% of individuals achieve sterile immunity after vaccination. This is split into 35% reaching immunity after 14 days after the first shot and 40% reaching immunity after 21 days.

1.A.2 The Synthetic Population

We build a synthetic population based on the German microcensus (Forschungsdatenzentren Der Statistischen Ämter Des Bundes Und Der Länder, 2018). We only use private households, i.e. exclude living arrangements such as nursing homes as non-private households vary widely in size and it is very difficult to know which contacts take place in such living arrangements.

We sample households to build our synthetic population of over one million households keeping for each of the 2.3 million individuals their age, gender, occupation and whether they work on Saturdays and Sundays. For each household we draw its county and set the corresponding federal state.

We randomly assign 35% of children below three to attend a nursery (Destatis, 2020). For children between three and six years old, we assume all go to preschool (officially 92.5% according to Destatis (2020)). Children that attend a nursery meet in groups of four (Bertelsmann Stiftung, 2019) plus one adult care taker every weekday when there are no school vacations. Preschool children meet in groups of nine (Bertelsmann Stiftung, 2019) with two adult care takers. These groups are mixed with respect to age but all belong to the same state and mostly to the same county.

Every child that goes to school is part of a school class. Each school class meets three times per weekday, each time with a different set of two teachers, unless there are vacations or policies that suspend schools.² Each class consists of approximately 23 students (OECD, 2013). All students in a class are of the same age and live in the same state and mostly also in the same county. In addition, each child gets assigned a value that captures his or her need to attend nursery, preschool or school. This allows us to capture various degrees of emergency care that can be granted while educational facilities are closed or are on some kind of rotating schedule.

Workers are assigned to a daily meeting work group. The group sizes vary to match the number of daily repeating work contacts reported by working individuals in Mossong et al. (2008). These groups only consist of workers that work in the same county. For a distribution of the number of daily recurring work contacts see Figure 1.A.2e. To match the number of weekly work groups we match each worker with up to 14 other workers into pairs to match the number of reported weekly work contacts shown in Figure 1.A.2f. Each pair is assigned a weekday on which they always meet in the absence of work policies. 80% of these contacts are individuals from the same county. In the same way children have an educational priority determining if they are entitled to emergency care workers are assigned a work contact

2. We implement vacations on the federal state level.

priority that captures how necessary their work is and to which degree they can work from home. This means that it's always the same individuals that continue to have work contacts when work from home mandates of a certain strictness are in place.

In addition to creating groups for educational facilities and work we also have other recurring contacts to represent things like groups of friends or sports teams that practice regularly together. Both daily and weekly groups are created analogously to the work groups but matching the numbers in Figure 1.A.2b and Figure 1.A.2c. In addition, since leisure contacts are highly assortative by age all individuals that have a daily leisure contact are matched with a person not only from the same county but also from the same age group.

The individuals in our population can react to events such as developing symptoms that are typical of CoViD-19, a positive PCR test or a positive rapid test by reducing their contacts. To determine who would reduce their contacts in such a situation or demand a rapid test we introduce a quarantine compliance parameter. Similarly, we introduce a rapid test compliance parameter that determines in which order individuals start demanding rapid tests when rapid tests become increasingly available. This makes sure that when for example only 10% of workers get tested, it's the same workers that have access to tests every week.

Lastly, for the distribution of vaccinations every individual is assigned a vaccination group and a vaccination rank from that group that creates a complete vaccination queue over the population including a share that refuses to be vaccinated (ξ) which we calibrate to 15% (Robert Koch-Institut, 2021b). The vaccination groups are created to match the recommendations by the Ständige Impfkommission (Vygen-Bonnet, Koch, Bogdan, Harder, Heininger, et al., 2020).³ To cover that the Pfizer-BioNTech vaccine was later approved for younger age groups we put adolescents and children into two groups that follow after the general population. These groups do not become eligible within our simulation frame until June. The way vaccinations are rolled out in our model is shown in Figure 1.A.1.

3. We cover that teachers were prioritized more than recommended by the commission.

22 | 1 The Effectiveness of Testing, Vaccinations and Contact Restrictions

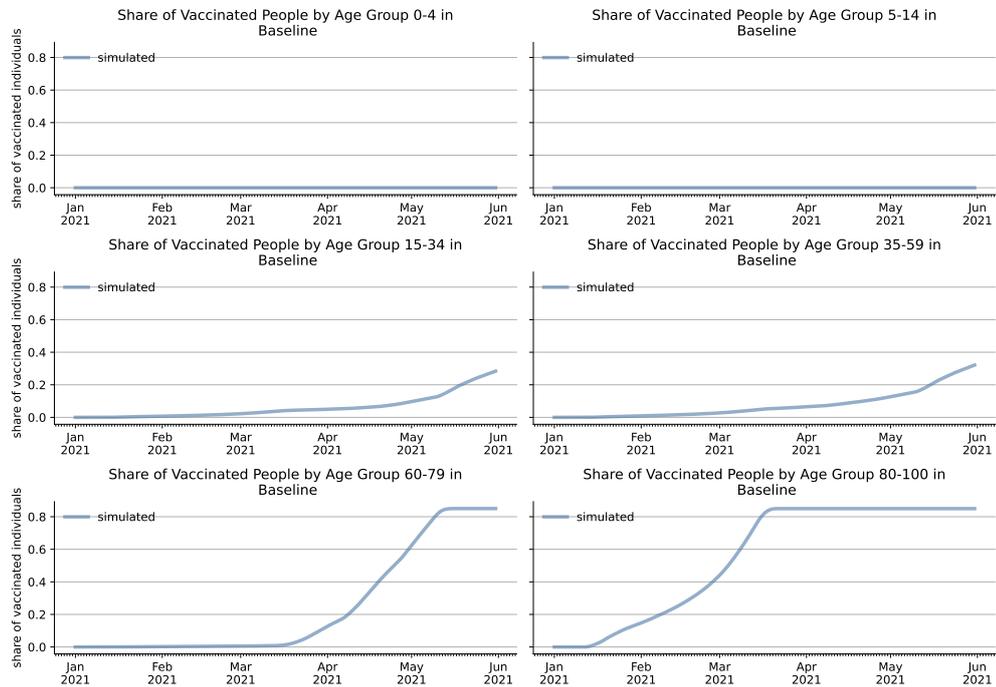


Figure 1.A.1. Vaccination Rates by Age Group

Note: An individual's vaccination priority depends on her work contact priority, her age group and a random component to capture preconditions like diabetes. 15% of the population refuse to be vaccinated (ξ). Adolescents would be vaccinated after the general population and children last. The figure clearly shows that the first vaccinations go to some workers with very high work contact priority and to the 80 to 100 age group followed by the 60 to 79 year olds. Both groups are saturated with vaccinations by mid March and start of May respectively. By June a third of the younger adults have received the vaccination but these groups still remain far from herd immunity thresholds.

1.A.3 Number of Contacts

We calibrate the parameters for the predicted numbers of contacts from contact diaries of over 2000 individuals from Germany, Belgium, the Netherlands and Luxembourg (Mossong et al., 2008). Each contact diary contains all contacts an individual had throughout one day, including information on the other person (such as age and gender) and information on the contact. Importantly, for each contact individuals entered of which type the contact (school, leisure, work etc.) was and how frequent the contact with the other person is. Binning the number of contacts for very high numbers, we arrive at the distributions of the numbers of contacts by type of contact (η_c) as shown in Figure 1.A.2.

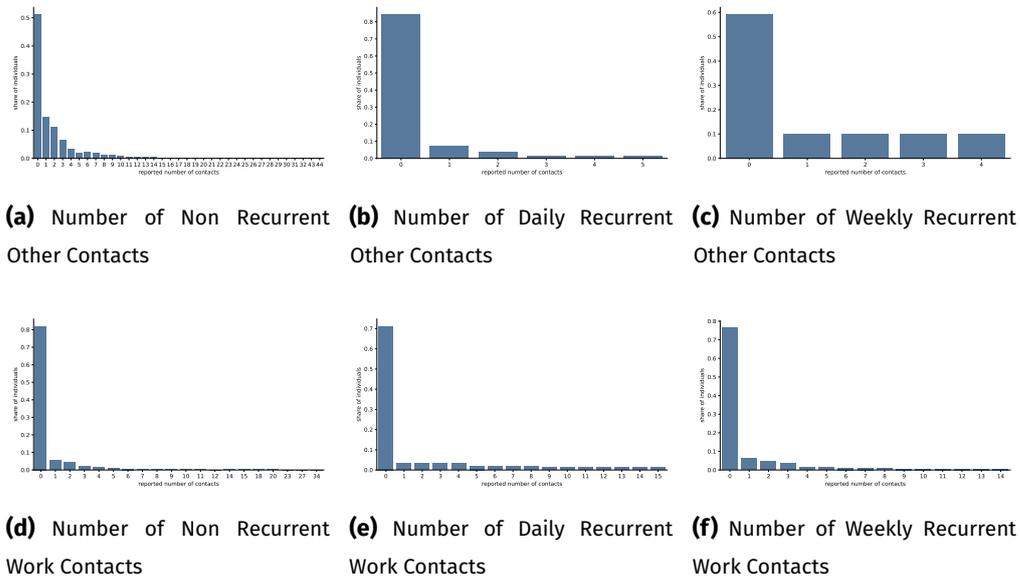


Figure 1.A.2. Number of Contacts of the Different Contact Types

Note: This figure shows the pre-pandemic number of contacts individuals report of different contact types (η_c). In the model it is sampled every day which of the numbers of non recurrent contacts a person is planned to have. Note that the contact diaries include such high values that super spreading events are well possible in our model through non recurrent models. For recurrent contacts individuals are put into groups that meet either every day or on a particular week day every day. The upper row shows the distribution of the number of other contacts individuals report (η_{other}). Other contacts include all contacts that are not household members, school contacts or work contacts, for example leisure contacts. We assume that individuals in households with children or teachers or retired individuals have additional non recurrent other contacts during school vacations to cover things like family visits or travel during vacations. The lower row shows the distribution of the different types of work contacts (η_{work}). Work contacts only take place between working individuals.

An exception where we do not rely on the data by Mossong et al. (2008) are the household contacts. Since households are included in the the German microcensus (Forschungsdatenzentren Der Statistischen Ämter Des Bundes Und Der Länder, 2018) on which we build our synthetic population we simply assume for the household contacts that individuals meet all other household members every day. The number of household contacts that happen every day is shown in Figure 1.A.3.

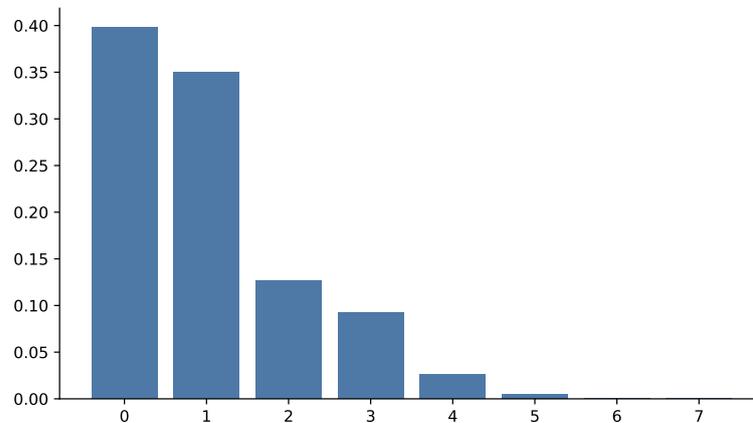


Figure 1.A.3. Number of Household Contacts

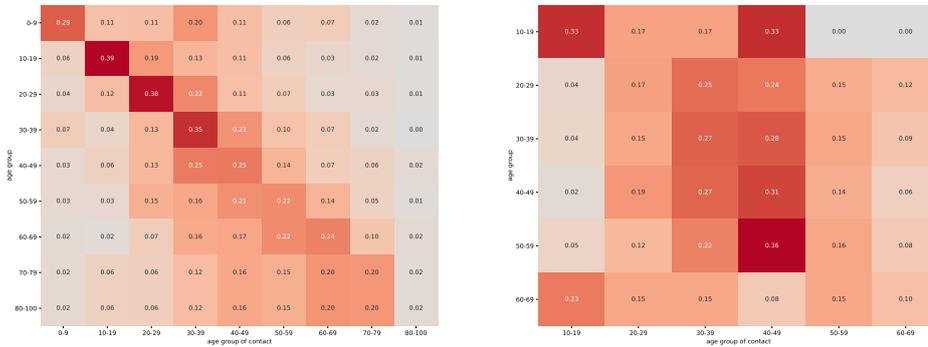
Note: Every individual meets all other household members every day. The German microcensus sampled full households such that our synthetic population automatically fits population characteristics such as size and age distribution.

1.A.4 Assortativity

As explained in section 1.B.5, the probability that two individuals are matched can depend on background characteristics. In particular, we allow this probability to depend on age and county of residence (α). While we do not have good data on geographical assortativity and set it such that 80% of contacts are within the same county, we can calibrate the assortativity by age from Mossong et al. (2008).

Figure 1.A.4a shows that assortativity of the other contacts by age is especially strong for children and adolescents. For older people, the pattern becomes more dispersed around their own age group, but within-age-group contacts are still the most common contacts. Figure 1.A.4b shows that assortativity by age is also important among work contacts.

For recurrent contacts, we constructed groups to have the following features: Recurrent work contacts are not assortative by age. Daily work groups are always of the same county and weekly work contacts are to 80% with workers from the same county. Other recurrent contacts are constructed the same way but we impose for daily contacts that they are always with individuals from the same age group. School classes are groups where the same children of the mostly same age and county meet with teachers every day. Nurseries and preschools mix children by age but match them to come mostly from the same county. Household age composition follows directly from the German microcensus data we use to construct our synthetic population.



(a) Distribution of Non Recurrent Other Contacts by Age Group

(b) Distribution of Non Recurrent Work Contacts by Age Group

Figure 1.A.4. Assortativity by Age Group for Non Recurrent Other and Work Contacts

Note: The figure shows the distribution of non recurrent contacts by age group for other contacts on the left and work contacts on the right. A row shows the share of contacts a certain age group has with all other age groups. Higher values are colored in darker red tones. The diagonal represents the share of contacts with individuals from the same age group. The 80-100 age group for other contacts was so small that we assumed for them to have the same contact distribution as the 70-79 year olds. For work contacts, we only show age groups that have a significant fraction of working individuals.

1.A.5 Policies

Our policies (denoted by ρ) usually affect one of three contact types: education, work and other contacts. Germany had no policies limiting contacts within households so there are no policies on them in our model.⁴

For nurseries, preschools and schools we implement vacations as announced by the German federal states as well as school closures, emergency care and rotating schedules where only one half of students attends every other week or day. An approximation of the share of contacts still taking place with the different school regulations can be found in Figure 1.A.5a. Note that schooling policies differ between states and usually involve rules that depend on local incidences. We simplify these rules to one federal policy from the federal incidence and the policies of the three most populous federal states (North Rhine-Westphalia, Bavaria and Baden-Württemberg). The testing policies for schools are described in Section 1.A.6.

Until November schools were open normally. Starting in November, we assume that increased hygiene measures were taken. Schools stayed open until mid December. From mid December until January 10 schools closed and only offered so

4. Household contacts can, however, be reduced when individuals quarantine themselves after developing symptoms, for example. This happens to a lesser degree than other contacts to capture difficulties in isolation within the home.

called “emergency care” for young children whose parents could credibly demonstrate that both had to work and had no other child care arrangement. Approximately 25% of primary school children and 5% of secondary students attended school as a result. After January 10 when parents had returned to work the rules for emergency care were relaxed and approximately a third of primary school children and 10% of secondary students attended school as a result. In addition, graduating classes (most adolescents between 16 and 18) were allowed to return to school in a rotating scheme where each class was split in two groups. Relying on anecdotal evidence we assume that the groups rotate on a daily basis. Starting on February 22 primary school children were also allowed to return to school on a rotating basis until mid March. We summarize the school policy from mid March until Easter as all students being on a rotating school schedule. In addition, children that qualify for emergency care also attend on days where their group is scheduled to not attend school physically. After the Easter break schools were mostly closed again. Part of this was a federal law, the so called “Bundesnotbremse” (Bundesgesetzblatt, 2021) that set rules for schools based on local incidences that were binding at the time. As a result, most states adjusted their schooling policies and during April most schools were closed with emergency care arrangements as in the time from January 10 to February 21. As cases fell schools were allowed to gradually open. We summarize this as students being on the same rotating schedule as from mid March to Easter starting on May 1 (Bayerisches Staatsministerium für Unterricht und Kultus, 2021; Landesregierung von Baden-Württemberg, 2021a; Landesregierung von Baden-Württemberg, 2021b; Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, 2021a; Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, 2021b; Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, 2021c; Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, 2021d; Bayerisches Staatsministerium für Unterricht und Kultus, no date; Landesregierung von Baden-Württemberg, no date).

The policies for preschools and nurseries are similar to the school policies but simpler. Until November children attended completely normally, starting in November with increased hygiene measures. Nurseries and preschools stayed open until mid December. From mid December until January 10, nurseries and preschools were nearly completely closed. If parents could credibly demonstrate that both parents work in systemically relevant professions and no other child care arrangement was possible, nurseries and preschools offered so called “emergency care”. We assume 10% of children qualified and used emergency care during this time. After January 10 when parents had returned to work the rules for emergency care were relaxed and we assume a third of children attended nursery and preschool. This policy stayed in place until February 20. Afterwards, preschools and nurseries were open normally (maintaining increased hygiene measures) until mid March. Then during the third wave the restrictions February were put back into place until end of April when nurseries and preschools opened again

and stayed open for the rest of our simulation period - maintaining increased hygiene measures. (Bayerisches Staatsministerium für Familie, Arbeit und Soziales, 2021a; Landesregierung von Baden-Württemberg, 2021c; Landesregierung von Baden-Württemberg, 2021d; Bayerisches Staatsministerium für Familie, Arbeit und Soziales, 2021b; Landesregierung von Baden-Württemberg, no date; Ministerium für Kinder, Familie, Flüchtlinge und Integration des Landes Nordrhein-Westfalen, no date).

For work contacts we use the reductions in work mobility reported by the Google Mobility Data (Google, LLC, 2021) to calibrate the reduction in physical work contacts ($\rho_{w, attend, t}$). Reductions in work contacts are not random but governed through a work contact priority where the policy changes the threshold below which workers stay home. Figure 1.A.5b shows the share of workers that go to work over time at the federal German level. We use the data on the state level to account for local holidays and differences in state regulations. In addition, for both work and school contacts we assume that hygiene measures (such as masks, ventilation and hand washing) became more strict and more conscientiously observed in November 2020, leading to a reduction of 33% in the number of contacts with the potential to transmit Covid-19 ($\rho_{hygiene}$).

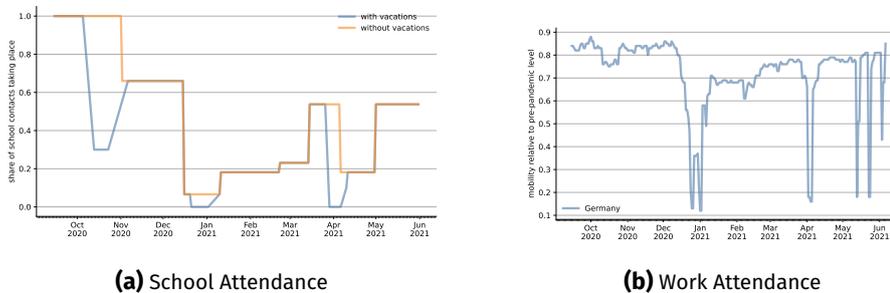


Figure 1.A.5. The Contact Reduction Effects of School and Work Attendance Policies

Note: The left figure shows the approximate share of school contacts taking place with and without vacations factored in. In contrast to other policies, school policies are not implemented via multipliers but as mechanistic models (e.g. split class approaches in with emergency care). For the above plot we assigned approximate multipliers to those policies. The figure is, thus, only an illustration that shows the approximate share of contacts taking place compared to the pre-pandemic level with and without vacations. The right figure shows the work mobility as reported by Google, LLC (2021). We take this as a proxy of the share of workers who still have physical work contacts ($\rho_{w, attend, t}$). The figure interpolates over weekends as we handle weekend effects through information on work on weekends in the German census data we use. The figure shows the share for Germany as a whole. To capture the effect that local policies, school vacations, etc. have on work contacts we use the data on the state level to determine which workers go to work depending on the state they live in.

Lastly, for the other contacts category ($\rho_{other,t}$) we could not calibrate the policies from data but estimated the policy effects. The estimation and values are detailed in Section 1.A.9 and Figure 1.A.10.

1.A.6 Rapid Test Demand

In our model, there are five reasons why rapid tests are done:

1. someone plans to have work contacts
2. someone is an employee or student of an educational facility
3. a household member has tested positive or developed symptoms
4. someone has developed symptoms but has not received a PCR test
5. someone plans to participate in a weekly non-work meeting

For work contacts, we know from the COSMO study (Betsch, Korn, Felgendreff, Eitze, Schmid, et al. (2021), 20th/21st of April) that 60% of workers who receive a test offer by their employer regularly use it ($\pi_{w,d}$). We assume this share to be time constant.

In addition, there are some surveys that allow us to trace the expansion of employers who offer tests to their employees ($\pi_{w,s,t}$). Mid march, 20% of employers offered tests to their employees (Deutscher Industrie- und Handelskammertag, 2021). In the second half of March, 23% of employees reported being offered weekly rapid tests by their employer (Ahlers, Lübker, and Jung, 2021). This share increased to 61% until the first days of April (Bonin, Krause-Pilatus, and Rinne, 2021; Bundesministerium für Wirtschaft und Energie, 2021). Until mid April 72% of workers were expected to receive a weekly test offer (Bonin, Krause-Pilatus, and Rinne, 2021; Bundesministerium für Wirtschaft und Energie, 2021). However, according to surveys conducted in mid April (Betsch, Korn, Felgendreff, Eitze, Schmid, Sprengholz, Wieler, Schmich, Stollorz, Ramharter, Bosnjak, Omer, Thaiss, De Bock, and Von Rüden, 2021), less than two thirds of individuals with work contacts receive a test offer. Starting on April 19th employers were required by law to provide two weekly tests to their employees (Bundesanzeiger, 2021b). We assume that compliance is incomplete and only 80% of employers actually offer tests. We interpolate between these points linearly, arriving at the blue line in Figure 1.A.6. In addition, we increase the frequency of testing ($\theta_{t,work}$) from weekly to twice weekly during April.

We assume that employees in educational facilities start getting tested in 2021 and that by March 1st 30% of them ($\pi_{teacher,t}$) are tested weekly ($\theta_{before\ Easter,educ} = 7$). The share increases to 90% for the week before Easter. At that time both Bavaria (Bayerisches Staatsministerium für Gesundheit und Pflege, 2021) and Baden-Württemberg (Ministerium für Kultus, Jugend und Sport Baden Württemberg, 2021) were offering tests to teachers and North-Rhine Westphalia (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, 2021f) and Lower

Saxony (Niedersächsisches Kultusministerium, 2021) were already testing students and tests for students and teachers were already mandatory in Saxony (Sächsisches Staatsministerium für Kultus, 2021). After Easter we assume that 95% of teachers get tested twice per week ($\theta_{after\ Easter, educ} = 3$).

Tests for students started later (Ministerium für Kultus, Jugend und Sport Baden-Württemberg, 2021; Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, 2021f) so we assume that they only start in February and only 10% of students get tested by March 1st ($\pi_{students, t}$). Relying on the same sources as above we approximate that by the week before Easter this share had increased to 40% (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, 2021f). After Easter the share of students receiving twice weekly tests is set to 75%. This is based on tests becoming mandatory in Bavaria (Bayerische Staatskanzlei, 2021) and North Rhine-Westphalia (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, 2021e) after their Easter breaks and on the 19th in Baden-Württemberg (Ministerium für Kultus, Jugend und Sport Baden-Württemberg, 2021), after which we assume twice weekly rapid tests to be mandatory for all students in Germany. Again, we interpolate linearly between these points and arrive at the purple line for teachers and the red line for school students in Figure 1.A.6.

To limit our degrees of freedom, we only have one parameter that governs how many individuals do a rapid test because of any of the private demand reasons ($\pi_{private, t}$).⁵ We assume that there is no private rapid test demand until March when both the citizens' tests and rapid tests for lay people started to become available (Bundesanzeiger, 2021a; Presse- und Informationsamt der Bundesregierung, 2021) and other access to rapid tests was very limited.

According to the COSMO study (Betsch, Korn, Felgendreff, Eitze, Schmid, et al., 2021) 63% would have been willing to take a test in the round of 23rd of February 2021 when an acquaintance would have tested positive. Since this is only asking for willingness not actual behavior, we take this as the upper bound of private rapid test demand which we estimate in our model to be reached in the beginning of May. To cover that many people are likely to have sought and done their first rapid test before the Easter holidays we add another point that we estimate for the rapid test demand around Easter. Similarly, we estimate one point in mid March when tests started to become available in grocery stores and pharmacies which we estimate in our model. The resulting share of private rapid test demand is shown as the green line in Figure 1.A.6 (also see Section 1.A.9 for details on the estimation).

5. The reasons that can lead to an individual doing a rapid test for private reasons are own symptoms but no PCR test, planned weekly leisure meeting or a symptomatic or positively tested household member

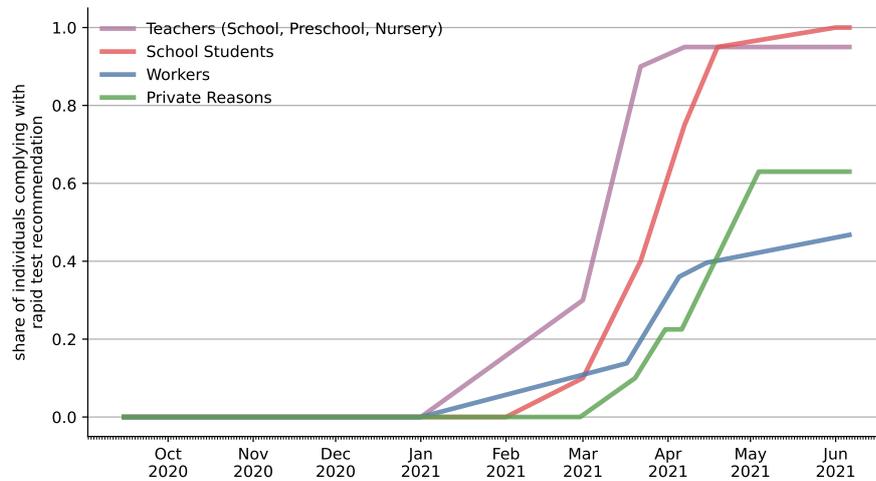


Figure 1.A.6. Share of Individuals Doing a Rapid Test.

Note: Rapid test demand can be triggered by individuals planning to have education contacts ($\pi_{teacher, t}$ or $\pi_{students, t}$), work contacts ($\pi_{w, d}$ and $\pi_{w, s, t}$), developing symptoms without access to a PCR test, having a household member with a positive test or symptoms ($\pi_{private, t}$). In each case whether a rapid test is done depends on how long it has been since the individual's last rapid test and her individual compliance parameter. As an example, take a worker in May. In that time workers are encouraged to test themselves twice weekly (or every three days, i.e. $\theta_{May, work} = 3$) but there is no general requirement to test themselves. If the worker has not done a test within the last four days in our model she will demand a test if her (time-constant) compliance parameter belongs to the upper 60% in the population.

1.A.7 Share of Detected Cases

One important feature of our model is that we distinguish between undetected and detected cases and that we model which cases are detected and which are not (see Section 1.B.7 for a detailed description for how we model both rapid and PCR tests). For our model it is important to have an estimate for the share of cases that is detected in the absence of rapid tests (ψ_t). For this we rely on the Paul, Steinberg, Schaller, Haemmerl, Thum, et al. (Dunkelzifferradar Project 2020) which uses estimates of the case fatality rate to estimate the number of total cases given the number of CoViD-19 deaths which are assumed to be perfectly observable. For 2020, we follow the reported share of detected cases quite closely. One exception is the phase of November 2020 where we interpolate to maintain monotonicity during the fall as there was no reason why the share of detected cases should have risen in that time⁶

Since vaccinations started after Christmas 2020 and these were predominantly given to nursing homes in the beginning and other vulnerable groups in spring, we expect the relationship between deaths and the number of total infections to change rapidly in 2021. This is why we stop using the share of detected cases estimated by the Dunkelzifferradar after Christmas. Instead, we assume that the share of detected cases would have stayed the same in the absence of rapid tests. Thus, we also achieve in our model an increase in the share of detected cases but this is driven from inside our model through increased rapid testing which lead follow-up PCR tests when they are positive (see Section 1.B.13 and 1.B.7).

Lastly, we model reductions in the share of detected cases due to the two major holidays in our simulation period, Christmas and Easter. During both holidays many laboratories did not process tests and most physicians' offices were closed, leading to less PCR tests and short and large drops in the share of known cases. The resulting share of detected cases in the absence of rapid tests is shown in Figure 1.A.7 and was estimated to fit the data.

6. The testing policy changed in November (Robert Koch Institute, 2020). However, this only moved the rare PCR tests more towards vulnerable groups.

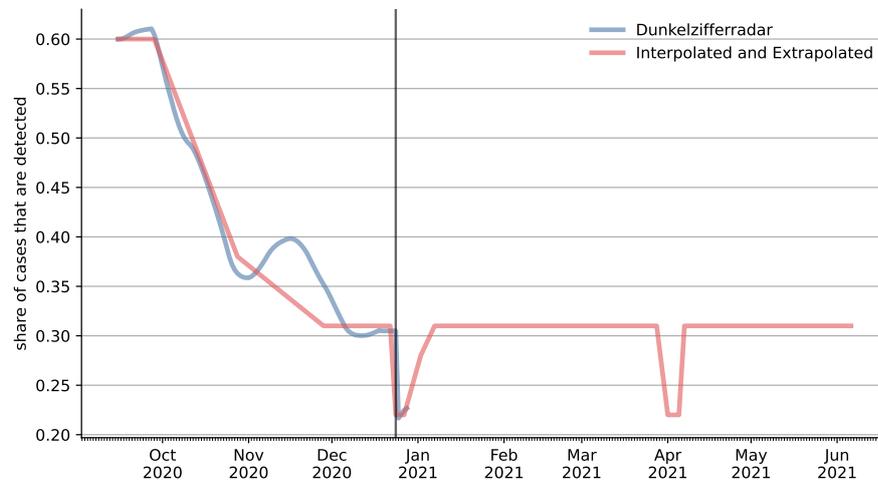


Figure 1.A.7. Share of Detected Cases in the Absence of Rapid Tests

Note: The figure shows the share of cases that is reported as an official case via PCR confirmation. We use the overall share of known cases that was estimated through the case fatality ratio by the Paul et al. (Dunkelzifferradar Project 2020) for all of 2020 and then assume it to be constant as vaccinations of the elderly strongly affect the case fatality rate which the project does not account for. Starting in 2021 in addition to the overall numbers of detected cases through symptoms and a random component, cases are also detected through confirmation of positive rapid tests which happens endogenously inside the model. For the public holidays of Christmas and Easter we lower the share of detected cases as fewer PCR tests are available during public holidays. See Figure 1.B.11 for how the share of detected cases develops in our model for each age group

1.A.8 PCR Testing and Behavioral Response

This section describes the remaining parameters for our testing model. Refer to Section 1.B.7 for a description of the full testing model.

From the share of detected cases and the number of infections we arrive at the number of positive PCR tests in our model. A share of these positive tests goes to symptomatic individuals ($\chi_{symptom,t}$). This share is calibrated from German data on case characteristics (Robert Koch Institut, 2021) and shown in Figure 1.A.8. We keep $\chi_{symptom,t}$ constant after Christmas because the RKI data does not include if a PCR test was done to confirm a positive rapid test and this share is used for PCR test demand without prior rapid test indication.

PCR tests take one to four days until their result is revealed to the individual ($\gamma_{PCR,d}$). Relying on the ARS data (Robert Koch-Institut, 2020) we calculate that 33% of individuals receive the test result after one day, 50% after two days, 10% after three days and 7% after four days.

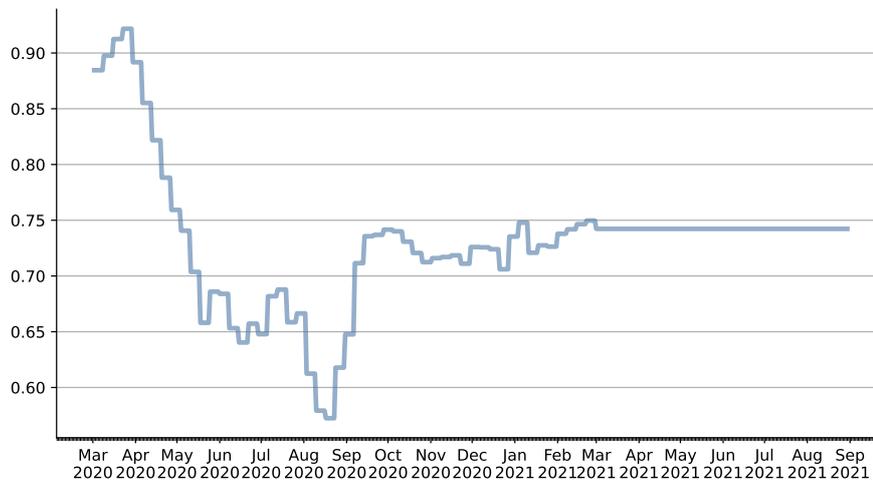


Figure 1.A.8. Share of Positive PCR Tests Administered to Symptomatic Individuals

Note: The share of positive PCR tests that are administered to symptomatic individuals ($\chi_{symptom, t}$). Since it was not recorded for every case if the person was symptomatic or not we take the midpoint between the upper and lower bound. We keep the share constant after Christmas because the RKI data does not include if a PCR test was done to confirm a positive rapid test and this share is used for PCR test demand without prior rapid test indication.

To model the demand for PCR tests through rapid tests, we only need the share of individuals that seek a PCR test to confirm a positive rapid test result ($\chi_{confirmation}$). We calibrate this from Betsch, Korn, Felgendreff, Eitze, Schmid, Sprengholz, Wieler, Schmich, Stollorz, Ramharter, Bosnjak, Omer, Thaiss, De Bock, and Von Rden (2021) who asked this as a hypothetical question in March of 2021. There 82% of Germans reported that they would follow up on a positive rapid test with a PCR test.

Lastly, we need to set the parameters that decide how individuals reduce their contacts after certain events, τ . We distinguish between the reduction in household contacts (which are harder to avoid) and non household contacts. There are three events which trigger potential contact reductions: showing symptoms of CoViD-19, having received a positive rapid test and having received a positive PCR test. The only survey data we are aware of on this is Betsch, Korn, Felgendreff, Eitze, Schmid, Sprengholz, Wieler, Schmich, Stollorz, Ramharter, Bosnjak, Omer, Thaiss, De Bock, and Von Rden (2021) where 85% of individuals claimed they would isolate and restrict their contacts after a positive rapid test. We assume this reduction for non household contacts. As household contacts are much more difficult to avoid, we assume that they are only reduced by 30%. We assume the same behavior for individuals that develop symptoms. Lastly, we assume the response to a positive PCR test

to be stronger than in the other two cases and set the reduction of non household contacts to 95% and the reduction of household contacts to 50%.

1.A.9 Estimated Parameters

We estimate parameters that cannot be calibrated outside of the model with the method of simulated moments (McFadden, 1989) by minimizing the distance between simulated and observed infection rates (disaggregated by region and age groups) and fatality rates. Since our model includes a lot of randomness, we average simulated infection rates over several model runs.

All estimated parameters are described in Table 1.A.1.

Table 1.A.1. Estimated Parameters

notation	estimate	note
Infection Probabilities		
$\beta_{household}$	0.1	base probability of getting infected by an infectious household member
β_{school}	0.012	base probability of getting infected by an infectious classmate or teacher
$\beta_{young\ educ}$	0.005	base probability of getting infected by an infectious classmate or teacher
β_{work}	0.1475	base infection probability for work contacts
β_{other}	0.15875	base infection probability for other contacts
Policy Parameters		
$\rho_{hygiene}$	0.66	reduces infectiousness of work and education contacts from November to end of simulation
$\rho_{other, before\ Oct\ 1}$	0.75	before October
$\rho_{other, Oct\ 1\ to\ Oct\ 20}$	1.00	high activity due to reopenings and fall vacations
$\rho_{other, Oct\ 21\ to\ Nov\ 1}$	0.75	anticipation of a lockdown and precaution due to high incidenes
$\rho_{other, Nov\ 2\ to\ Dec\ 1}$	0.52	“lockdown light”
$\rho_{other, Dec\ 2\ to\ Dec\ 23}$	0.57	“lockdown light” with lockdown fatigue and holiday shopping
$\rho_{other, Dec\ 24\ to\ Dec\ 26}$	0.65	Christmas holidays
$\rho_{other, Dec\ 27\ to\ Feb\ 10}$	0.35	hard lockdown after Christmas
$\rho_{other, Feb\ 11\ to\ Feb\ 28}$	0.50	lower precaution due to low incidences and lockdown fatigue
$\rho_{other, after\ Feb\ 28}$	0.515	many contact reducing policies are lifted
B.1.1.7 Introduction		
$\omega_{B.1.1.7, Jan\ 31}$	0.986	number of B.1.1.7 cases per 100 000 individuals to import on January 31st. Imported B.1.1.7 cases gradually rise from January 1st where 0 cases are imported. No cases are imported in other months.
Rapid Test Introduction		
$\pi_{private,t}$	Figure 1.A.6	the private rapid tests levels in mid March and at Easter as well as the date at which full availability of private rapid tests is reached are fit to the data. Values between those levels are interpolated linearly. The remaining rapid tests demands are calibrated from surveys. See Section 1.A.6

We fit our model to data for Germany from mid September 2020 until June 2021. We do not use earlier periods for three reasons. Firstly, in the beginning PCR tests were very scarce and the reported case numbers unreliable. Secondly, during the summer the case numbers were extremely low. This could lead to the epidemic going extinct in our simulation. Thirdly, over the summer, imported cases from touristic travel were likely important for the infection dynamic but there is not enough data to include them into our model.

To avoid over-fitting and simplify the numerical optimization problem, we only allow for five different infection probabilities: 1) for contacts in schools 2) for contacts in preschools and nurseries. 3) for work contacts. 4) for households. 5) for other contacts.

Since the infectiousness of a contact between an infectious and a susceptible person depends on many things, the numerical values of the infection probabilities in Table 1.A.1 only reflect a base probability. This base probability is modified by a seasonality factor, an age specific susceptibility factor and an infectiousness factor that depends on the virus strand of the infected person. The base infection probability is only equal to the actual infection probability when all of those factors are 1. This would be the case for a contact between an 80+ year old susceptible person with a person who is infected with the B.1.1.7 strand of the virus on January first.

It is not possible to rank different types of contacts according to their infectiousness just from the numerical values of the infection probabilities. There are two reasons for this: Firstly, for computational reasons the seasonality factor is normalized such that it reaches 1 at its peak. It has thus a lower average for contact types with strong seasonality (e.g. other contacts) than for contact types with weak seasonality (e.g. work contacts). Secondly, for household and school contacts we do not have data on whether people actually have physical contact. Thus the infection probabilities for those contact types are actually the product of the probability to actually have physical contact on a given day and the infection probability of that contact.

In order to get a feeling for the infectiousness of each contact type it is more intuitive to look at how many infections were actually caused by each contact type. This is depicted in Figure 1.A.9. We can see that work and other contacts are the main drivers of the pandemic, followed by infections in households. Schools and preschools contribute fewer infections which is to be expected given that there are much fewer students than working adults in the German population. Nevertheless, Figure 1.B.16 shows that schools do have a notable effect on the infection dynamic in the long run.

We also estimate a parameter that reflects the effect of hygiene measures at work and in educational facilities. This parameter becomes active in November 2020 when stricter mask mandates and distancing rules were introduced. It is estimated to reduce infectiousness of contacts by one third.

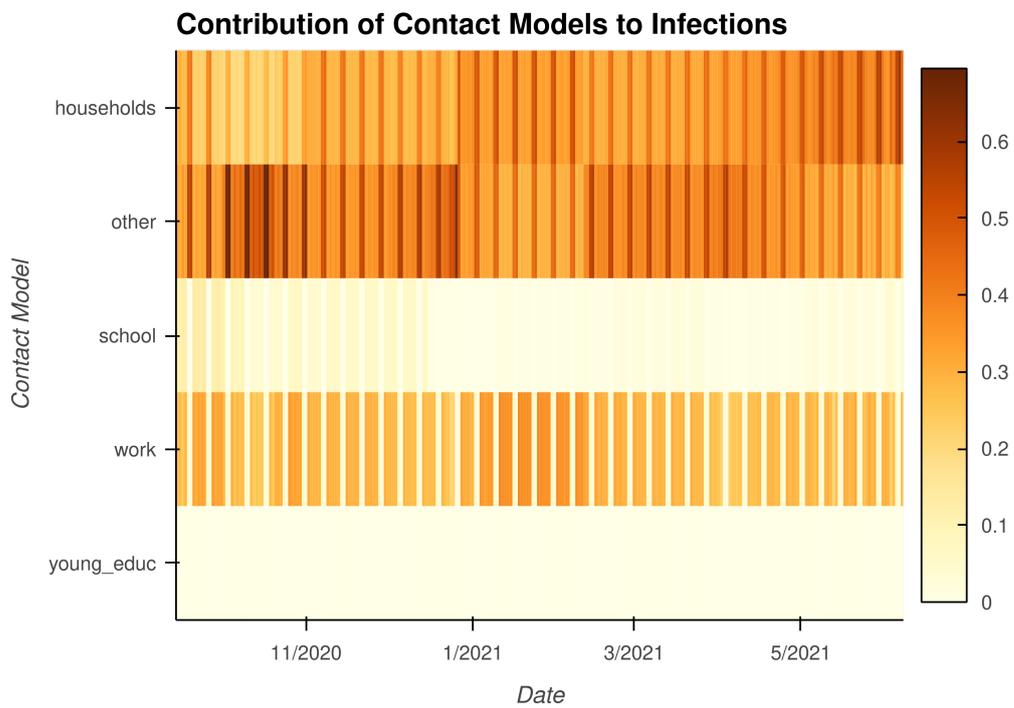


Figure 1.A.9. Daily share of infections by contact type

Note: Daily share of infections that were contributed by each contact type. Darker colors mean that a larger share of infections were contributed by that contact type. The majority of infections take place in the workplace, in households and via other contacts. Schools and preschools contribute less infections, especially after hygiene measures have been introduced.

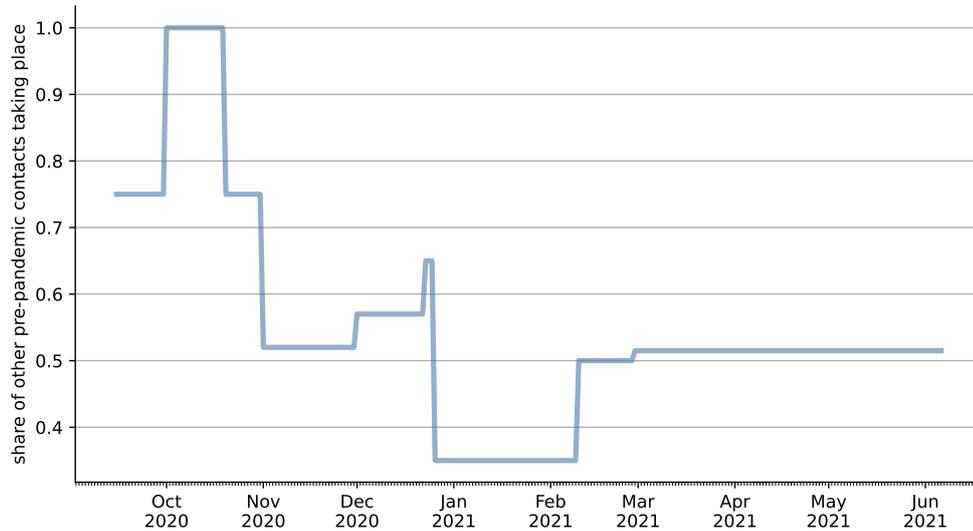


Figure 1.A.10. Share of Pre-Pandemic Other Contacts Taking Place with Infection Potential

Note: Values of the other multiplier. All values are estimated via the method of simulated moments. The rationale behind each switching point is described in Table 1.A.1

Moreover, we estimate nine different multipliers that reflect how strongly other contacts are reduced over time. The dates at which we switch between the multipliers usually coincide with policy changes and is not determined from the case numbers. The only exception to this are slight adjustments to parameters to incorporate lockdown fatigue (towards the end of a lockdown period) or precautionary contact reductions (in times of high incidences right before a lockdown is enacted). The estimated other multipliers are also depicted in Figure 1.A.10.

While we estimate nine different values for the other contact multiplier, they are not estimated completely freely. In particular we ensure that the ordering of the parameter values is consistent with the stringency of policies. For example, the strongest contact reduction was estimated for January 2021 during where very strict measures and curfews were in place, whereas the weakest contact reduction was in October 2020 where policies were very lenient.

Since we do not have good data on the reduction of other contacts, it is not possible to separately estimate parameters for contact reduction and the effect of hygiene measures. The reported other multipliers in Table 1.A.1 are thus a combination of contact reduction and hygiene measures.

Finally we estimate one parameter that governs the introduction of the B.1.1.7 virus variant in January 2021. This parameter implies that at the end of January roughly one case per 100 000 individuals per day is imported. After January we do not model imported cases of B.1.1.7 anymore because they are negligible compared to the endogenous growth of that virus variant.

While a formal identification argument is beyond the scope of this paper, below we give a rough intuition which features of the data help us to estimate each parameter.

The different infection probabilities can be separately identified because the degree to which each contact type is active varies over time (e.g. school closures, vacations and different work from home policies) and they affect different subgroups of the population differently (e.g. β_{school} most strongly affects kids whereas β_{work} has the strongest effect on adults in working age and β_{other} affects all age groups equally). The hygiene and other multipliers can be identified because they are only active in certain time periods. However, it is necessary to normalize one other multiplier to 1 because there is no period without any contact reduction in our data. The introduction parameter for the B.1.1.7 mutation can be identified from the share of that virus strand in the population. The rapid test demand parameters are identified because rapid tests first lead to a very steep increase in observed cases and then to a sudden decrease – in a time where almost all other things in the model would not cause a change in trend.

1.A.10 Shapley Values

We decompose the effects of different NPIs and seasonality on the infection rates with Shapley values. Shapley values (Shapley, 2016) are a concept in game theory to divide payoffs between a coalition of players. It allows to assign a single value to the contribution of an NPI or seasonality which takes into account substitutional and complementary effects with other factors.

More formally, define a coalitional game with N players and a super-additive function ν which maps subsets of N to the real numbers. The function ν is also called the characteristic function and assigns a value to a coalition. Then, the Shapley value ϕ for player i is

$$\phi_i(\nu) = \frac{1}{|N|!} \sum_{S \subseteq N \setminus \{i\}} |S|!(|N| - |S| - 1)!(\nu(S \cup \{i\}) - \nu(S))$$

The last term ($\nu(S \cup \{i\}) - \nu(S)$) is the marginal contribution of player i minus the coalition without player i . Then, compute the sum of marginal contributions over all subsets S of N which do not include player i . Each marginal contribution has to be multiplied by all combinations of other players in S which precede i and all possible combinations of remaining players which follow player i in the coalition. To arrive at the Shapley value for player i , divide the sum by the total number of combinations.

The Shapley value has some properties.

Efficiency The sum of Shapley values is equal to the value of a coalition formed by all players.

Symmetry The Shapley does not depend on the label of a player but only on its position in the characteristic function.

Linearity The Shapley value depends linearly on the values from the characteristic function v .

Dummy Axiom The Shapley value of a player who contributes nothing to any coalition is 0.

To produce panels c and b of Figure 1.4.1 we calculate the Shapley values of each factor in the comparison on the cumulative number of saved infections between the main scenario and the scenario without any of the factors for every day. Then, we divide up the saved infections on a particular day according to the Shapley values for the same day which yields the daily saved infections for each factor.

1.A.11 Overview Model Parameters

Table 1.A.2. Contacts, Matching and Policies

name	notation	time active	source	description
n_contacts	$\eta_{c,n}$	always	Mossong et al. (2008)	probability to have n contacts in contact type c . For non recurrent contacts the number of contacts is drawn for each individual every day. For recurrent contacts individuals are assigned group identifiers that determine the groups in which they meet either daily or weekly. The number of groups individuals belong to are calibrated to arrive at the number of recurrent contacts reported in Mossong et al. (2008). See Section 1.A.3.
degree_of_assortativity	$\alpha_{c, a_i, a_j, county_i, county_j}$	always	Mossong et al. (2008)	probability in contact type c that an individual of age group a_i and county $county_i$ meets an individual of age group a_j and county $county_j$. Contacts can be assortative by age group, state and county. The degree of assortativity depends on the contact type c . See Section 1.A.4 for details.
work_attend_multiplier	$\rho_{w, attend, t}$	always	Google, LLC (2021)	share of workers that continue to have work contacts, i.e. do not work from home on date t . We use the reduction in work mobility reported by Google, LLC (2021) as a proxy for the share of workers that work from home. See Section 1.A.5 for details.
hygiene_multiplier	$\rho_{hygiene}$	since November 2020	estimated	reduction in transmission during work and educational contacts due to stricter hygiene measures such as wearing face masks. See Section 1.A.9.
other_multiplier	$\rho_{other, t}$	always	estimated	reduction in the transmission during other contacts, such as leisure. This incorporates both hygiene measures as well as the reduction of physical meetings. There are nine breakpoints for the whole estimation period. See Section 1.A.9 for details.

Table 1.A.3. Infection Probabilities and Virus Variants

name	notation	time active	source	description
infection_prob	β_c	always	estimated	Base infection probability of contact type c . For each contact, this base probability is further adjusted by the seasonality, susceptibility of the contact and the virus strain. See Section 1.A.9 for details.
susceptibility	ζ_a	always	Davies, Klepac, et al. (2020)	Susceptibility to Covid-19 depends on a person's age group. The higher the age the more easily people become infected. The susceptibility of the oldest age group is normalized to one.
seasonality	κ_c	always	Gavenčiak et al. (2021)	The probability to contract Covid-19 when exposed depends on the seasonality. Since different contact types are more or less subject to seasonal variation (e.g. by moving contacts outdoors) the seasonality also depends on the contact type. Refer to Section 1.B.8 for an explanation.
variant_infectiousness	σ_v	always	Davies, Abbott, et al. (2021)	Variant v 's infectiousness relative to the wild type.
variant_introduction	$\omega_{v,t}$	time dependent	estimated	Number of infections per 100,000 individuals of variant v to be introduced on day t . See Section 1.A.9 for details.

Table 1.A.4. Disease and Vaccination Model

name	notation	time active	source	description
p_duration_immune	$Y_{immune, d}$	always	see Section 1.A.1	probability to stay immune for d days after having contracted CoViD-19
p_duration_until_infectious	$Y_{infectious, d}$	always	see Section 1.A.1	probability to become infectious d days after infection
p_duration_of_infectiousness	$Y_{stop\ infectious, d}$	always	see Section 1.A.1	probability that infectiousness lasts d days
p_duration_until_symptoms	$Y_{symptoms, a, d}$	always	see Section 1.A.1	probability for individuals of age group a to develop symptoms d days (possibly infinite) after becoming infectious
p_duration_of_symptoms	$Y_{stop\ symptoms, d}$	always	see Section 1.A.1	probability for individuals of age group a that symptoms last d days
p_duration_until_icu	$Y_{icu, a, d}$	always	see Section 1.A.1	probability for symptomatic individuals of age group a to require intensive care d days (possibly infinite) after symptom onset
p_duration_of_icu	$Y_{stop\ icu, a, d}$	always	see Section 1.A.1	probability to recover after d days from requiring intensive care
p_duration_until_death	$Y_{dead, a, d}$	always	see Section 1.A.1	probability for individuals of age group a in intensive care to die after d days (possibly infinite)
p_until_immune_by_vaccine	$Y_{vacc, d}$	2021	see Section 1.A.1	probability to develop immunity d days (possibly infinite) after being vaccinated
share_vaccine_refusers	ξ	always	Frisch (2021)	share of individuals refusing to be vaccinated. See Section 1.A.2

Table 1.A.5. Rapid Testing

name	notation	time active	source	description
rapid_test_specificity	$P_{negative not\ infected}$	2021	Brümmer et al. (2021)	the probability of an uninfected person to receive a negative rapid test result. See Section 1.B.7.
rapid_test_sensitivity	$P_{positive infected, i, t}$	2021	see 1.B.7	the probability of an infected person to receive a positive rapid test result. This depends on the timing of the test relative to the individual's onset of infectiousness. See Section 1.B.7.
share_accepting_work_rapid_test_offer	$\alpha_{w, d}$	2021	Betsch, Korn, Felgendreff, Eitze, Schmid, Sprengholz, Wieler, Schmich, Stollorz, Ramharter, Bosnjak, Omer, Thaiss, De Bock, and Von Rügen (2021)	share of workers that regularly pick up rapid test offers by their employers when they do not work from home. In our baseline specification this is time constant. See Section 1.A.6.
share_workers_receiving_rapid_test_offer	$\pi_{w, s, t}$	2021	see 1.A.6	share of workers that get a regular rapid test offer by their employer when they do not work from home on date t .
share_educ_workers_with_rapid_test	$\pi_{teacher, t}$	2021	see 1.A.6	share of educational workers who test themselves as part of their work on date t
share_students_with_rapid_test	$\pi_{students, t}$	2021	see 1.A.6	share of school pupils that do rapid tests at school on date t .
share_private_rapid_test_demand	$\theta_{private, t}$	2021	see 1.A.6	share of individuals that do a rapid test when any of the private reason events such as a household member testing positive occur on date t .
rapid_test_educ_freq	$\theta_{t, educ}$	2021	see 1.A.6	Frequency with which individuals test themselves in educational settings at time t . See Section 1.A.6.
rapid_test_work_freq	$\theta_{t, work}$	2021	see 1.A.6	Frequency with which complying workers test themselves at time t . See Section 1.A.6.

Table 1.A.6. PCR Testing, Case Detection and Behavioral Response

name	notation	time active	source	description
p_duration_until_test_result	$\gamma_{PCR, d}$	always	Robert Koch-Institut (2020)	probability that it takes d days between the performance of a PCR test and receiving the result. See Section 1.A.8.
share_of_tests_for_symptomatics	$\chi_{symptom, t}$	always	calibrated from RKI	share of positive PCR tests that are performed on individuals because of Covid-19 symptoms. See Section 1.A.8.
share_w_positive_rapid_test_requesting_pcr	$\chi_{confirmation}$	2021	Betsch, Korn, Felgendreff, Eitze, Schmid, Sprengholz, Wieler, Schmich, Stollorz, Ramharter, Bosnjak, Omer, Thaiss, De Bock, and Von Rügen (2021)	share of individuals with positive rapid test that seek a PCR test. See Section 1.A.8.
share_known_cases_without_rapid_tests	ψ_t	always	Paul et al. (2020)	share of cases that would be detected in the absence of rapid tests (see Section 1.B.7)
symptomatic_multiplier	$\tau_{symptoms, c}$	always	see 1.A.8	share of symptomatic individuals that still have contacts of type c .
positive_pcr_multiplier	$\tau_{positive PCR, c}$	always	see 1.A.8	share of individuals with a positive PCR test that still have contacts of type c .
positive_rapid_test_multiplier	$\tau_{positive rapid test, c}$	2021	Betsch, Korn, Felgendreff, Eitze, Schmid, Sprengholz, Wieler, Schmich, Stollorz, Ramharter, Bosnjak, Omer, Thaiss, De Bock, and Von Rügen (2021)	share of individuals with a recent positive rapid test that still have contacts of type c . See Section 1.A.8

1.A.12 Reproducibility

The source code used for this paper is open source and available under the MIT License. It is split into two parts

- The source code for the model can be found at <https://github.com/covid-19-impact-lab/sid/> and its documentation at <https://sid-dev.readthedocs.io>.
- The source code for the application to Germany can be found at <https://github.com/covid-19-impact-lab/sid-germany/> with a shorter documentation at <https://sid-germany.readthedocs.io>.

We are grateful to the authors and contributors of the following software packages upon which our software is built: conda Anaconda (2016), conda-forge conda-forge community (2015) dask Rocklin (2015), estimagic Gabler (2020), holoviews Stevens, Rudiger, and Bednar (2015), matplotlib Hunter (2007), numba Lam, Pitrou, and Seibert (2015), numpy Harris, Millman, Walt, Gommers, Virtanen, et al. (2020), pandas McKinney (2010) and The pandas development team (2020), pytask Raabe (2020), Python Van Rossum and Drake Jr (1995), scipy Virtanen, Gommers, Oliphant, Haberland, Reddy, et al. (2020), and seaborn Waskom (2021).

Appendix 1.B Supplementary Text

1.B.1 Literature Review

A commonly used model class in epidemiology are agent-based simulation models. In a prototypical agent-based simulation model, individuals are simulated as moving particles. Infections take place when two particles come closer than a certain contact radius (e.g. Silva, Batista, Lima, Alves, Guimarães, et al. (2020) and Cuevas (2020)). While the simulation approach makes it easy to incorporate heterogeneity in disease progression, it is hard to incorporate heterogeneity in meeting patterns. Moreover, policies are modeled as changes in the contact radius or momentum equation of the particles. The translation from real policies to corresponding model parameters is a hard task.

These shortcomings have motivated variations of agent-based simulation models where moving particles have been replaced by contact networks for households, work and random contacts. The OpenABM-Covid-19 model by Hinch, Probert, Nurtay, Kendall, Wymant, et al. (2021) and the model by Aleta et al. (2020) are the closest in spirit to ours.

The OpenABM-Covid-19 model by Hinch, Probert, Nurtay, Kendall, Wymant, et al. (2021) also uses detailed contact networks for workplaces, schools and households and can evaluate the effect of several NPIs. The main focus of their application are contact tracing policies (Abueg, Hinch, Wu, Liu, Probert, et al., 2021). Recently they have also added support for multiple virus strains and vaccinations.

Aleta et al. (2020) develop an agent-based simulation model with a very high geographical resolution by estimating contact networks from fine grained mobility data for the Boston metropolitan area. They use this model to show how NPIs, contact tracing and PCR testing can influence the infection dynamics. However, they do not calibrate their model to match actual infection numbers which makes it more suitable to explore the general mechanics of different disease mitigation measures than for their quantitative evaluation.

Bicher, Rippinger, Urach, Brunmeir, Siebert, et al. (2021) simulate the entire Austrian population. They use data from the first wave (February 21 to April 9, 2020) to calibrate their model and predict the effect of different NPIs and contact tracing policies until November 2020. They use the same data provided by Mossong et al. (2008) as we to calibrate contact networks for households, workplaces and schools. The model focuses on analyzing the effect of different contact tracing strategies and not on modelling enacted Austrian policies over a long period of time.

Moreover, there are several working papers that develop agent-based simulation models with contact networks in conjunction with economic models. Examples are Basurto, Dawid, Harting, Hepp, and Kohlweyer (2020), Delli Gatti and Reissl (2020) and Mellacher (2020).

Our model combines elements from the above models and adds several others. To the best of our knowledge, our model is the only one with the following features:

1. The free model parameters have been estimated with the method of simulated moments (McFadden, 1989). Despite having few free parameters our model does an excellent job in explaining observed case numbers and the spread of the B.1.1.7 mutation over more than nine months of data.

2. We have an extremely fine grained representation of schools and preschools. We can thus easily model all schooling policies that have been implemented in Germany in the past nine months. This includes complete school closures, phases where only those students whose parents could not find any private childcare arrangement could attend, split class approaches for some or all age groups and combinations thereof. Moreover, we can account for additional hygiene measures whose effect is estimated inside the model.

3. We model the evolution of the pandemic and all enacted policies since the start of the second wave. Since the vast majority of cases has occurred in that time period and we also model unobserved infections our simulations take into account that many people are already immune because they have recovered from an infection and that this immunity is not spread randomly across the population.

4. We have an extremely detailed model of PCR and rapid tests with a share of detected cases that varies over time and across age groups.

5. Our model is designed to combine information from many different data sources. Examples are surveys on rapid test demand (Betsch, Korn, Felgendreff, Eitze, Schmid, Sprengholz, Wieler, Schmich, Stollorz, Ramharter, Bosnjak, Omer, Thaiss, De Bock, and Von Rden, 2021), reaction to test results (Betsch, Korn, Fel-

gendreff, Eitze, Schmid, Sprengholz, Wieler, Schmich, Stollorz, Ramharter, Bosnjak, Omer, Thaiss, De Bock, and Von Räden, 2021), contact diaries (Mossong et al., 2008), share of detected cases (Paul et al., 2020) and many more.

1.B.2 Summary

We use an agent-based simulation model with detailed contact networks. The model structure is depicted in panel a of Figure 1.2.1.

We distinguish between eight types of contacts which are all listed in Figure 1.2.1: households, recurrent and random work contacts, recurrent and random leisure contacts, as well as nursery, preschool, and school contacts.

The number of contacts is translated into infections by a matching algorithm. There are different matching algorithms for recurrent contacts (e.g. classmates, family members) and non-recurrent contacts (e.g. clients, contacts in supermarkets). All types of contacts can be assortative with respect to geographic and demographic characteristics.

The infection probabilities of contacts vary with contact type, age of the susceptible person, and the virus strain of the infected person. Moreover, they follow a seasonal pattern. The strength of the seasonality effect is higher for contacts that are easy to be moved to an outside location in summer (such as leisure contacts) and smaller for contacts that take place inside even in summer (e.g. work contacts).

Once a person is infected, the disease progresses in a fairly standard way which is depicted in panel b of Figure 1.2.1. Asymptomatic cases and cases with mild symptoms are infectious for some time and recover eventually. Cases with severe symptoms additionally require hospitalization and lead to either recovery or death.

After rapid tests become available, people who work or go to school can receive rapid tests there. Moreover, people can decide to do a rapid test if they develop symptoms, have many planned contacts or have a sick or positively tested household member. People who have a positive rapid test demand a confirmatory PCR test with a certain probability. Moreover, PCR tests can be demanded because of symptoms or randomly.

This rich model of PCR and rapid tests leads to a share of detected cases that varies over time and across age groups. It also allows to quantify the effect of changes in testing policies on the dynamic of infections.

People who have symptoms or received a positive test can reduce their number of contacts across all contact types endogenously. The extent to which this is done is calibrated from survey data.

The model makes it very simple to translate policies into model quantities. For example, school closures imply the complete suspension of school contacts. A strict lockdown implies shutting down work contacts of all people who are not employed in a systemically relevant sector. It is also possible to have more sophisticated policies

that condition the number of contacts on observable characteristics, risk contacts or health states.

An important feature of the model is that the number of contacts an individual has of each contact type can be calibrated from publicly available data (Mossong et al., 2008). This in turn allows us to estimate policy-invariant infection probabilities from time series of infection and death rates using the method of simulated moments (McFadden, 1989). Since the infection probabilities are time-invariant, data collected since the beginning of the pandemic can be used for estimation. Moreover, since we model the testing strategies that were in place at each point in time, we can correct the estimates for the fact that not all infections are observed.

The model has a very modular structure and can easily be extended to distinguish more contact types, add more stages to the disease progression, implement new policies or test demand models. The main bottleneck is not complexity or computational cost but the availability of data to calibrate additional model features.

1.B.3 Modeling Numbers of Contacts

Consider a hypothetical population of 1,000 individuals in which 50 were infected with a novel infectious disease. From this alone, it is impossible to say whether only those 50 people had contact with an infectious person and the disease has an infection probability per contact (β) of one or whether everyone met one infectious person but the disease has an infection probability of only 5 percent per contact. SEIR models do not distinguish between the number of contacts (η) and the infectiousness of each contact (β). Instead, they combine the two into one parameter that is not invariant to social distancing policies.

To model social distancing policies, we need to disentangle the effects of the number of contacts of each individual and the effect of mostly policy-invariant infection probabilities specific to each contact type.

The number and type of contacts in our model can be easily extended. Each type of contacts is described by a function that maps individual characteristics, health states and the date into a number of planned contacts for each individual. This allows to model a wide range of contact types.

In our empirical application we distinguish the following contact types that are depicted in panel a of Figure 1.2.1 and can be further grouped in the categories household, work, education and others:

- Households: Each household member meets all other household members every day.
- Recurrent work contacts: These capture contacts with coworkers, repeating clients and superiors. Some of these recurrent contacts take place on every work-day, others just once per week. The contacts are assortative in geographical location and age.

- Non recurrent work contacts: Working adults have contacts with randomly drawn other people, which are assortative in geographical location and age.
- Schools: Each student meets all of his classmates every day. Class sizes are calibrated to be representative for Germany and students have the same age and mostly live in the same county. Schools are closed on weekends and during vacations, which vary by states. School classes also meet six teachers every day and some of the teachers meet each other.
- Preschools: Children who are between three and six years old attend preschool. Each group consists of nine children of mixed ages and two adults who live mostly in the same county. They all meet each other every work day when there are no vacations.
- Nurseries: Children younger than three years may attend a nursery and interact with one adult. The age of the children varies within groups but all live in the same county. They all meet each other every work day when there are no vacations.
- Non recurrent other contacts: Contacts with randomly drawn other people, which are assortative with respect to geographic location and age group. This contact type reflects contacts during leisure activities, grocery shopping, medical appointments, etc..
- Recurrent other contacts representing contacts with friends neighbours or family members who do not live in the same household. Some of these contacts happen daily, others only once per week. They are assortative in geographic location and age.

The number of random and recurrent contacts at the workplace, during leisure activities and at home is calibrated with data provided by Mossong et al. (2008). For details see Section 1.A.3. In particular, we sample the number of contacts or group sizes from empirical distributions. It would also be possible to use economic or other behavioral models to predict the number of contacts.

1.B.4 Reducing Numbers of Contacts via NPIs

Our model makes it very easy to model a wide range of NPIs, either in isolation or simultaneously. This is important for two reasons: Firstly, it allows to predict and quantify the effect of novel NPIs. Secondly, it allows to model the actually implemented policy environment in great detail, which is necessary to use the full time series of infections and fatality rates to estimate the model parameters.⁷

7. See Avery, Bossert, Clark, Ellison, and Ellison (2020) for an explanation why it can be harmful to use too long time series to estimate simple SEIR type models.

Instead of thinking of policies as completely replacing how many contacts people have, it is often more helpful to think of them as adjusting the pre-pandemic number of contacts. Therefore, we implement policies as a step that happens after the number of contacts is calculated but before individuals are matched.

On an abstract level, a policy is a function that modifies the number of contacts of one contact type. This function can be random or deterministic. For example, school closures simply set all school contacts to zero. A work from home mandate leads to a share of workers staying home every day whereas those who cannot work from home are unaffected. Hygiene measures at work randomly reduce the number of infectious contacts for all workers who still go to work.

Policies can also interact. For example, school vacations are temporally reducing school contacts to zero while at the same time increasing other contacts to account for increased leisure activities and family visits during this time. This is important to reproduce the finding that school vacations do not reduce infection numbers even though schools lead to infections when open (Isphording, Lipfert, and Pestel, 2021).

The most complex policies are typically found in the education sector. Since the beginning of 2021 schools have switched back and forth between full closures, split class approaches with alternating schedules for some or all age groups and reopening while maintaining hygiene measures. On top of that there are different policies for allowing young students whose parents work full time to attend school even on days where they normally would not. For details on how we calibrate these policies see Section 1.A.5.

Importantly, policies can depend on the health states of participating individuals. For example children rarely go to school when they have symptoms. It would even be possible to quarantine entire school classes if one student tested positive and many other forms of contact tracing. For an application of our model showcasing private contact tracing in the context of the Christmas holidays see Gabler, Raabe, Röhrli, and Gaudecker (2020).

Not all things that reduce contacts compared to the pre-pandemic level are driven by NPIs. Therefore, we also model endogenous contact reductions that depend on the health state of individuals. Other possible factors could include things such as the local incidence. The extent to which contacts are reduced can be calibrated from surveys.

1.B.5 Matching Individuals

The empirical data described above only allows to estimate the number of contacts each person has. In order to simulate transmissions of Covid-19, the numbers of contacts have to be translated into actual meetings between people. This is achieved by a matching algorithm:

As described in section 1.B.3, some contact types are recurrent (i.e. the same people meet regularly), others are non-recurrent (i.e. it would only be by accident

that two people meet twice). The matching process is different for recurrent and non recurrent contact models.

Recurrent contacts are described by two components: 1) A set of time invariant groups, such as school classes or groups of co-workers. Those groups are generated once at the beginning of the simulation. The groups can be sampled from empirical data or created by randomly matching simulated individuals into groups. 2) A deterministic or random function that takes the value 0 (non-participating) and 1 (participating) and can depend on the weekday, date and health states of the entire population. This can be used to model things like vacations, weekends or symptomatic people who stay home (see section 1.B.4 for details).

Given those two components, the disease transmission for recurrent contacts is extremely simple: On each simulated day, every person who does not stay home meets all other group members who do not stay home. If there is a contact between individual i who is infected with virus variant v and infectious and individual j who is in age group a and susceptible, then j becomes infected with the following probability

$$P(\text{infection}) = \beta_c \cdot s_{c,t} \cdot \sigma_v \cdot \zeta_a \quad (1.B.1)$$

where β_c denotes the base infection probability of contact type c , $s_{c,t}$ is a seasonality factor between zero and one that depends on the contact type c and time t (see Equation 1.B.5), σ_v is the infectiousness factor of virus variant v and ζ_a is an age dependent susceptibility factor.

The assumption that all group members have contacts with all other group members is not fully realistic, but a good approximation to reality, especially in light of the suspected role of aerosol transmission for Covid-19 (Anderson, Turnham, Griffin, and Clarke, 2020; Morawska, Tang, Bahnfleth, Bluysen, Boerstra, et al., 2020). Alternatively, the infection probability of recurrent contact types can be interpreted as being the product of a true infection probability and the probability that an actual contact takes place.

The matching of non-recurrent contact types is more difficult because the contact network is resampled randomly every day. Moreover, it needs to allow for assortative matching. In our application, all random contacts are assortative with respect to age group a (it is usually more likely to meet people from the same age group) and county (it is more likely to meet people from the same county) but in principle any set of discrete variables can be used. This set of variables that influence matching probabilities induce a discrete partition of the population into groups.

Below we first describe one iteration of a simplified matching algorithm that illustrates what we want to achieve. In practice, we approximate the result of this matching algorithm by a two stage sampling procedure that is computationally more efficient. The matching is done for each non-recurrent contact type c . The following step is repeated until no individual has unmatched contacts left. Let z be an iteration

counter for the matching algorithm and i denote the individual whose unmatched contacts we are trying to match.

Let $K_{z,i,c}$ denote the number of unmatched contacts of individual i of contact type c before iteration z is completed. Note that $K_{z,i,c} \leq n_{ic}$ which is the total number of contacts individual i has of type c .

Let a_i denote i 's age group and $county_i$ her county of residence.

We first draw individual j from the distribution defined by probability mass function F_z over individuals $j \neq i$ in the synthetic population where the probability f_{zj} is calculated as follows:

$$f_{zj} = \underbrace{\alpha_{c,a_i,a_j,county_i,county_j}}_{\text{Group Probability}} \cdot \underbrace{\frac{K_{z,j,c}}{\sum_{l=1}^N K_{z,l,c} \cdot \mathbb{I}_{county_l=county_j \wedge a_l=a_j}}}_{\text{Individual Probability}} \quad (1.B.2)$$

We then draw an individual j . If one of the two participants is susceptible and the other one is infectious, we sample whether an infection takes place. The success probability for this event is calculated as in Equation 1.B.1. Finally we update the remaining numbers of unmatched contacts by setting:

$$K_{z+1,i,c} = K_{z,i,c} - 1 \quad (1.B.3)$$

$$K_{z+1,j,c} = K_{z,j,c} - 1 \quad (1.B.4)$$

The runtime of this algorithm scales roughly cubic in the number N of simulated individuals. This is because the number of iterations is proportional to N , in each iteration we have evaluate Equation 1.B.2 N times and each evaluation of that equation entails a sum over N individuals.

This makes it prohibitively expensive. We therefore replace the above algorithm by a two stage sampling procedure, where we first sample the group from which individual j will be drawn according to the group probabilities defined in Equation 1.B.2. Next we sample an individual from this group with the Individual probabilities defined in Equation 1.B.2.

Thus, while the calculation of any given second stage probability entails exactly the same number of calculations as before we do not have to calculate a second stage probability for all simulated individuals but only for those who are members of the group that was sampled in the first stage.

It is easy to see that ex-ante the probability of being sampled are identical between the two stage sampling and the one stage sampling. The only drawback is that towards the end of the matching process it becomes possible to sample a group in which no unmatched contacts are left. In our empirical application this happens extremely rarely. This is so for two reasons: Firstly, the first stage sampling probabilities have been estimated from the same dataset as the numbers of contacts so

there cannot be any mismatches such as for example a group that has a low probability of being sampled in the first stage but where all members have many contacts. Secondly, the group sizes are relatively large and we go over individuals in random order. Therefore, groups where no unmatched contacts remain only occur very late in the matching process.⁸

1.B.6 Course of Disease

The disease progression in the model is fairly standard. It is depicted in panel b of Figure 1.2.1 and the values and source of the relevant parameters are described in Section 1.A.1.

First, infected individuals will become infectious after one to five days. Overall, about one third of people remain asymptomatic. The rest develop symptoms about one to two days after they become infectious. Modeling asymptomatic and pre-symptomatic cases is important because those people do not reduce their contacts nor do they have an elevated probability to demand a test. Thus they can potentially infect many other people (Donsimoni, Glawion, Plachter, and Wälde, 2020). The probability to develop symptoms with Covid-19 is highly age dependent with 75% of children not developing clinical symptoms (Davies, Klepac, et al., 2020).

A small share of symptomatic people will develop strong symptoms that require intensive care. The exact share and time span is age-dependent. An age-dependent share of intensive care unit (ICU) patients will die after spending up to 32 days in intensive care. Moreover, if the ICU capacity was reached, all patients who require intensive care but do not receive it die.

We allow the progression of the disease to be stochastic in two ways: Firstly, state changes only occur with a certain probability (e.g. only a fraction of infected individuals develops symptoms). Secondly, the number of periods for which an individual remains in a state is drawn randomly. The parameters that govern these processes are taken from the literature and detailed in Section 1.A.1. For an overview of our disease progression parameters see Table 1.A.4.

1.B.7 Testing

Having a realistic model of PCR and rapid tests is crucial for two reasons: Firstly, only via a testing model can the simulated infections from the model be made comparable to official case numbers. Secondly, individuals with undetected or not yet detected infections are an important driver of the pandemic.

8. If unmatched contacts were a concern one could simply use the fast two stage sampling process for a first pass over contacts and then match all remaining contacts with the slow algorithm.

In principle, our modeling approach is flexible enough to incorporate mechanistic test demand, allocation and processing models. However, there is not enough data available to calibrate such a mechanistic model.

Therefore, we build a simpler model of PCR and rapid tests that can be calibrated with available data on test demand and availability and – nevertheless – can produce a share of undetected cases that varies over time and across age groups and agrees with other estimates over the time periods where they are available.

PCR tests are modeled since the beginning of the simulation and determine whether an infection is officially recorded. Rapid tests are only added at the beginning of 2021. Positive rapid tests do not enter official case numbers directly, but most people with a positive rapid test demand a confirmatory PCR test. However, positive rapid tests can have a strong effect on the infection dynamics because they trigger contact reductions and additional rapid tests.

During 2020 people can demand PCR tests either because they have symptoms or randomly. The probability that a PCR test is performed in each of the two situations depends on the number of new infections and the number of available tests. Thus, it varies strongly over time and is unknown.

To distribute the correct number of PCR tests among symptomatic and asymptomatic infections without knowing explicit test demand probabilities, we use the following approach: First, we calculate the total number of positive PCR tests by multiplying the number of newly infected individuals with an estimate of the share of detected cases from the Dunkelzifferradar project (Paul et al., 2020). Next, we determine how many of these tests should go to symptomatic and asymptomatic individuals from data by the RKI (Robert Koch-Institut, 2020). Then, we sample the individuals to which those tests are allocated from the pools of symptomatic and asymptomatic infected but not yet tested individuals.

Sampling uniformly from the pool of symptomatic individuals ensures that age groups who are more likely to develop symptoms are also more likely to receive tests. Thus, the share of detected cases is much higher for the elderly than for children in time periods where many tests are done because of symptoms.

At the beginning of 2021, two challenges arise: Firstly, the externally estimated share of detected cases from Dunkelzifferradar project (Paul et al., 2020) can no longer be used because it is based on the case fatality rate which changes drastically due to vaccinations. Secondly, rapid tests become available at a large scale.

We solve the first challenge by assuming that the share of detected cases would have remained at the level it reached before Christmas if rapid tests had not become available. While this is only an approximation to reality, changes in the share of detected cases that would have happened without rapid tests are very likely to be small compared to the changes caused by rapid tests.

The second challenge is solved by mechanistic rapid test demand models for the workplace, schools and by private individuals. The calibration of these models is

described in Section 1.A.6. Figure 1.B.6 shows that the number of performed rapid tests in the model fits the empirical data well (where empirical data is available).

In contrast to PCR tests, rapid tests are not perfect and can be falsely positive or falsely negative. While the specificity of rapid tests is calibrated at 99.4% (Brümmer et al., 2021), their sensitivity strongly depends on the timing of the rapid test relative to the start of infectiousness. We follow Smith et al. (2021) for our main results: Before the onset of infectiousness the sensitivity is very low (35%). On the first day of infectiousness it is much higher (88%) but still lower than during the remaining infectious period (92%). After infectiousness stops, the sensitivity drops to 50%. We show that our results are robust to more conservative assumptions in Section 1.B.12.

Modeling the diagnostic gap before and at the beginning of infectiousness is very important to address concerns that rapid tests are too unreliable to serve as screening devices.

We do not distinguish between self administered rapid tests and those that are administered by medical personnel. While there were concerns that self administered tests are less reliable, a recent study has found no basis for this concern (Lindner, Nikolai, Kausch, Wintel, Hommes, et al., 2020).

While rapid tests do not directly enter official case numbers, 82% ($\chi_{confirmation}$) of positively tested individuals seek a PCR test (Betsch, Korn, Felgendreff, Eitze, Schmid, Sprengholz, Wieler, Schmich, Stollorz, Ramharter, Bosnjak, Omer, Thaiss, De Bock, and Von Rügen, 2021). Importantly, those PCR tests are made in addition to the tests that would have been done otherwise. Section 1.B.13 discusses the effect of rapid tests on the share of detected cases.

1.B.8 Seasonality

It is widely acknowledged that the transmission of SARS-CoV-2 is subject to seasonal influences. Infectiousness is increased in winter when most contacts take place inside and the immune system is weakened by low levels of vitamin D, dry air and large temperature swings. For a detailed overview of possible drivers see Kronfeld-Schor, Stevenson, Nickbakhsh, Schernhammer, Dopico, et al. (2021).

We follow Kühn, Abele, Mitra, Koslow, Abedi, et al. (2020) and Gavenčiak et al. (2021) in modeling seasonality in the transmission of SARS-CoV-2 as a multiplicative factor on infection probabilities. The factor follows a sine curve that reaches its maximum at January 1 and its minimum on June 30.

For simplicity we normalize the factor to reach one at its maximum. Thus, the formula of the seasonality factor is given by:

$$s_{c,t} = 1 + 0.5\kappa_c \sin\left(\pi\left(\frac{1}{2} + \frac{t}{182.5}\right)\right) - 0.5\kappa_c \quad (1.B.5)$$

Where κ_c is difference in the seasonality factor between peak infectiousness and lowest infectiousness.

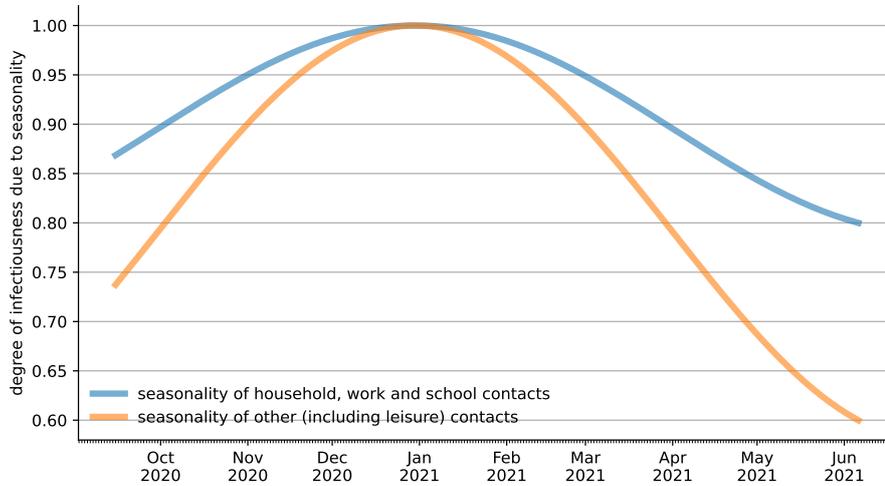


Figure 1.B.1. Seasonality by Type of Contact

Note: We model seasonality as a factor that reduces the probability of infection of all encounters. The factor depends on the day and is calculated from a sinus shaped function with its maximum on January 1. Since seasonality can affect the transmission both through physical conditions such as temperature and humidity as well as through the numbers of contacts that take place outside we assume two seasonality factors. One for other contacts which we expect to be strongly affected by fairer weather with a maximum reduction of 42% in the infection probability. The other seasonality only makes contacts up to 21% less infectious and is applied to household, work and school contacts.

The subscript c is needed because the strength of the seasonality effect differs across contact types: Work, household and school contacts are likely to take place inside even in summer. Thus they are only subject to seasonality due to factors that influence the immune system. Other contacts (for example meeting friends and while doing leisure activities) are mostly happening outside in the summer. Therefore, transmission via those contacts should have a stronger seasonal pattern.

We calibrate κ_{strong} to 0.42 and κ_{weak} to 0.21. This is in line with Gavenčiak et al. (2021) and Kühn et al. (2020).

The two seasonality curves are shown in Figure 1.B.1.

1.B.9 Initial Conditions

Consider a situation where you want to start a simulation with the beginning set amidst the pandemic. It means that several thousands of individuals should already have recovered from the disease, be infectious, symptomatic or in intensive care at the start of your simulation. Additionally, the sample of infectious people who will determine the course of the pandemic in the following periods is likely not representative of the whole population because of differences in behavior (number of contacts, assortativity), past policies (school closures), etc.. The distribution of

health states in the population at the beginning of the simulation is called initial conditions.

To come up with realistic initial conditions, we match reported infections from official data to simulated individuals by age group and county. We use one month of data to generate initial conditions with in all possible health states. Meanwhile health states evolve until the beginning of the simulation period without simulating infections by contacts. We also correct reported infections for a reporting lag and scale them up by the share of detected cases to arrive at the true number of infections.

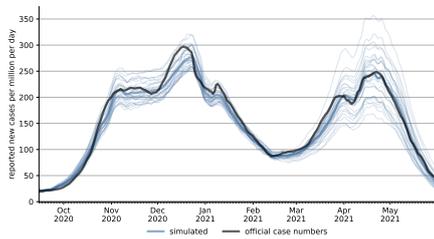
1.B.10 Model Fit

This section compares simulated data from our model with empirical data from Germany. We look at observed infections (overall as well as by age group and federal state), the effective replication number, the spread of B.1.1.7 and vaccinations.

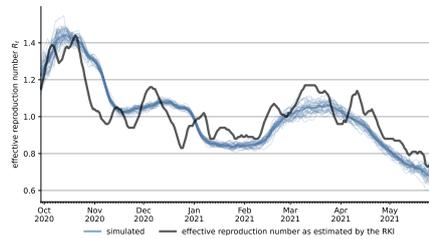
Overall, our model achieves an excellent fit of the two waves of infections with few free parameters (Figure 1.B.2a). As a result the effective replication number R_t also closely follows that reported by the RKI (see Figure 1.B.2b). We also achieve an excellent fit for most age groups in Germany. The fit is also good for many German federal states. Despite the fact that the number of performed rapid tests and their distribution in the population are determined endogenously in our model, we fit the share of the population with at least a weekly rapid test very well. For the share of individuals who have ever done a rapid test we err on the side of too few test.

Our fit of the infection rates in Germany between October 2020 and June 2021 is excellent. The incidence in our model matches both the levels and the shape of the reported incidence almost perfectly. When the prevalence of the virus is high and especially after explosive growth phases, the effect of random events on the incidence is large. Therefore all reported simulations average over at least 30 simulation runs which is enough to reduce the sampling uncertainty to a negligible level.

Our fit of the effective replication number R_t closely follows the values reported by the RKI (see Figure 1.B.2b) even though we calculate R_t on all infected individuals not just the detected cases. This explains why the R_t in our simulations is higher during phases where the share of detected cases (ψ_t) falls. This is the case in the fall of 2020 (see Figure 1.A.7) where the RKI underestimated the effective replication number due to observing a falling share of cases. Analogously, the R_t in our simulations is lower than the R_t reported by the RKI in spring where the share of known cases increased due to increased rapid testing.



(a) Observed Incidence in the Model and as Reported by the RKI



(b) Effective Replication Number R_t in the Model and as Reported by the RKI

Figure 1.B.2. Model Fit of the Reported Cases and the Effective Replication Number

Note: Both figures show averages and single runs. The average is the thick line. Single runs are shown as lighter and thinner lines. We averaged and show 30 simulation runs. The left figure shows the daily incidence rate per million for the simulated reported infection rates. The official case numbers as reported by the RKI are plotted in black. The fit is overall very good. The right figure shows the effective replication number (R_t) as reported by the RKI and as calculated in our model. The R_t gives the average number of new infections caused by one infected individual. The R_t in our model broadly follows the R_t reported by the RKI. Two differences stand out. Firstly, the RKI's R_t drops faster in November. This is likely due to a decline in the estimated overall share of detected cases (ψ_t) when the second wave hit Germany. The second difference is from mid February to mid March where the RKI's reported R_t increased more rapidly than that in our model. Here the opposite effect can be expected. During this time rapid tests increased strongly leading to more cases being detected. In the short term this leads an R_t estimation that is based on detected cases to overestimate the replication number. For legibility reasons, all lines are rolling 7-day averages.

Zooming into the different age groups in Figure 1.B.3, we can see that our model is also able to reproduce the infection rates on this level. The only major deviation from this pattern is that our model predicts too few infections for the 80 to 100 year olds. This was to be expected because our synthetic population does not include inhabitants of nursing homes. Outbreaks in nursing homes led to a large number of infections among the oldest during the second wave of the pandemic in Germany. Moreover, the model predicts too few observed infections for the 15 to 34 years old at the end of 2020 and the 5 to 14 years old in April and May 2021. The former is likely due to the fact that this age group has a very active social life which is not fully captured by our contact networks. The latter probably comes from a too conservative model of school reopenings.

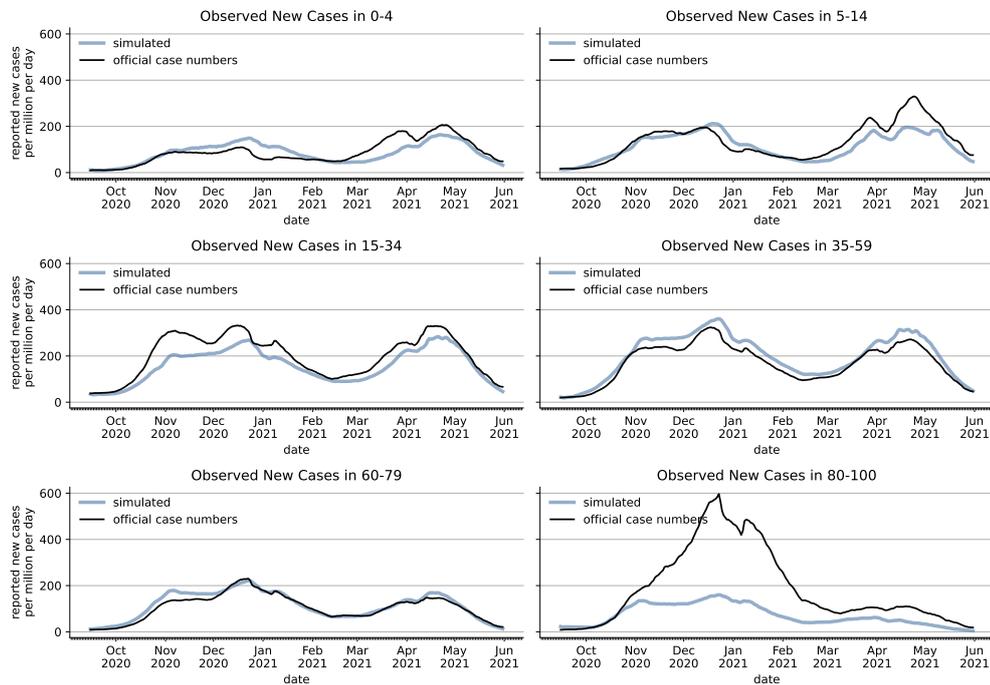


Figure 1.B.3. Simulated and Empirical Infections by Age Group

Note: The figure shows the number of reported versus simulated cases per one million people per day for different age groups. The age group of individuals above 80 needs to be interpreted with caution because our synthetic population only includes private households, i.e. nursing homes are not represented in our model. They accounted for many cases and deaths in the winter of 2020 and many 80 to 100 year olds live in these facilities. We average over 30 simulation runs. For legibility reasons, all lines are rolling 7-day averages.

Our model fit is also very good for the different German federal states. This holds not only for the large states such as North Rhine-Westphalia or Bavaria but also for many smaller states such as Hessen or Rhineland-Palatinate. This shows that using school vacations dates and work mobility reductions by Google, LLC (2021) at the state level combined with county and age group specific initial conditions (see Section 1.B.9) and county level assortativity of contacts is sufficient to represent many local differences. The fit is especially good given that our model does not aim to have a high local resolution. For example we abstract from population density and cross-border travel. It is, thus, unsurprising that there are states that we do not match well, such as very thinly populated Mecklenburg-Vorpommern and Schleswig-Holstein or Saxony with its large border to the Czech Republic that had a much higher incidence than Germany.

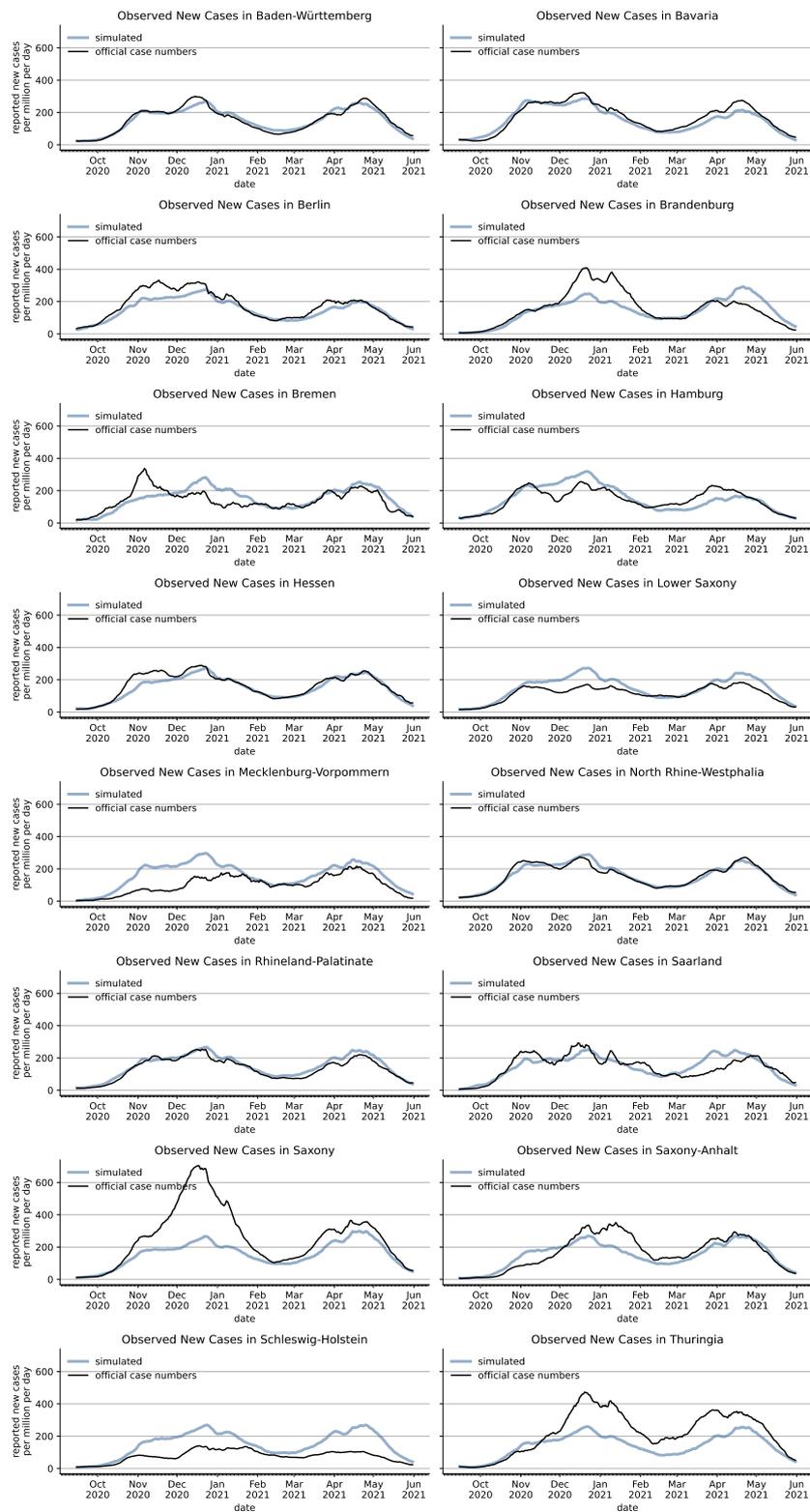


Figure 1.B.4. Simulated and Empirical Infections by Federal State

Note: The figure shows the number of reported versus simulated cases per one million people per day for different federal states. We averaged over 30 simulation runs. For legibility reasons, all lines are rolling 7-day averages.

We fit the proliferation of the B.1.1.7 variant quite exactly despite only introducing a few cases in January ($\omega_{B.1.1.7,t}$) as can be seen in Figure 1.B.5a. Since we only model B.1.1.7 and do not include other variants, B.1.1.7 reaches a share of nearly 100% by May while the true rate plateaued at 90%. By the end of May B.1.617.2 gained traction in Germany. However, given that B.1.617.2 made up less than 5% even at the end of our simulation period, we did not include it in our model.

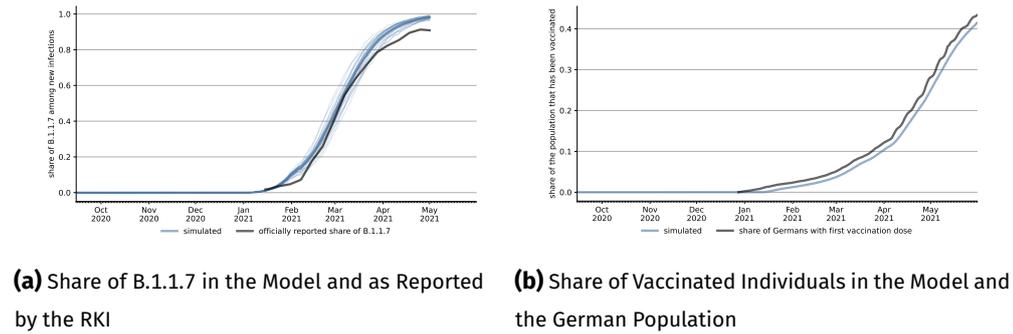


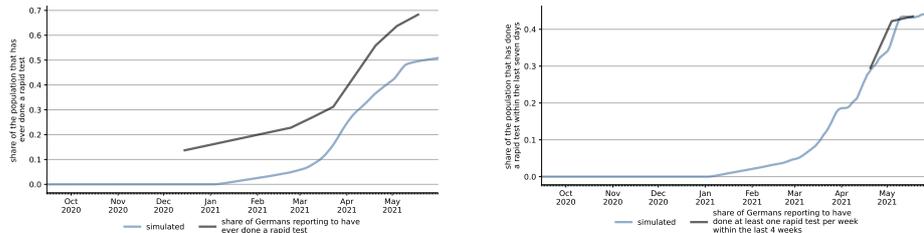
Figure 1.B.5. Model Fit of the Share of B.1.1.7 and Vaccinations

Note: The left figure shows the share of B.1.1.7 as reported by the RKI and as calculated in our model. We only introduce a few cases over the course of January. From then B.1.1.7 takes over endogenously through its increased infectiousness ($\sigma_{B.1.1.7}$). The right figure shows the rate of individuals that are vaccinated in our synthetic population versus in the general German population. For legibility reasons, all lines are rolling 7-day averages.

The fit of the share of vaccinated individuals can be seen in Figure 1.B.5b. In Germany, vaccines were rolled out according to four priority groups. The first vaccines were mostly reserved for nursing homes and some selected professions such as first responders. Since we do not have nursing home inhabitants in our model, we subtract the first percent of vaccinations which is equivalent to the share of Germans living in nursing homes. Afterwards, the share of vaccinated individuals in the population follows the German increase exactly. We took great care to model the prioritization of older individuals and professions that cannot reduce physical contact easily such as teachers or medical staff (see Section 1.A.2 and Figure 1.A.1 for the vaccination rates in our model by age group).

The most difficult moment to match in our model is the rapid test demand. This is because we have five different channels through which individuals demand rapid tests and many of the demand curves are at least partially calibrated through survey data. It is therefore very reassuring that we fit the share of individuals that do weekly rapid tests almost perfectly. For the share of individuals that have ever done a rapid test our model is conservative. There are two reasons for this: Firstly, we do not model people who have done rapid tests out of curiosity once they became available. Secondly, in the model, the decision to take a rapid test is based on a time invariant individual specific compliance factor without any additional random components.

While this captures important features of rapid test demand it abstracts from people who turn down rapid tests most of the time but accept them sometimes. Fortunately, Section 1.B.11 shows that our main results are robust to changes in the exact shares of individuals demanding rapid tests.



(a) Share of the Population That Has Ever Done a Rapid Test

(b) Share of the Population That Did a Rapid Test in the Last Week

Figure 1.B.6. Share of Individuals With Rapid Tests

Note: Both panels compare empirical and simulated rapid test demands. The empirical data comes from Betsch, Wieler, et al. (2021). The left panel compares the share of individuals who have ever done a rapid test. The right panel compares the share of individuals who have done a rapid test within the last seven days in our simulation compared to the share reporting to have done at least weekly rapid tests in the last four weeks in the COSMO survey. Overall our calibration of rapid tests are slightly conservative. The overall share is below that in the study. We fit the share of weekly tests quite exactly. For legibility reasons, all simulated lines are rolling 7-day averages.

1.B.11 Model Validation

Achieving a good in-sample fit does not necessarily guarantee that our model will also be able to make out of sample predictions. For example, it could be that the results are very sensitive to the exact number of vaccinations, the work mobility multiplier ($\rho_{w, attend, t}$) or the number of performed rapid tests (governed by the π parameters) – all of which are things that cannot be known exactly ex-ante.

In this section we compare simulated infections that use all available data with out of sample predictions that only use data that was available at March 1 2021.

For the out of sample predictions we predict the number of vaccinations between March and June with a simple linear regression model that was fitted on vaccine data from February. This prediction model is pessimistic compared to the actual number of vaccinations. The work mobility multiplier ($\rho_{w, attend, t}$) is predicted to be constant at a value of 0.75, which is an approximate average of the second half of February. This turned out to be optimistic.

The area that is fraught with the most uncertainty is the introduction of rapid tests, because it comprises both supply and demand factors. Moreover, accurately

predicting the number of rapid tests is expected to be important because rapid tests play a large role for the transmission dynamic.

We therefore make a scenario analysis with different assumptions on the availability of rapid tests. The number of rapid tests performed in each scenario can be seen in Figure 1.B.7. All scenarios are the same until March 1 and have the same level of rapid tests when all supply constraints are resolved. They differ in the date at which the full number of tests is reached. For students ($\pi_{students,t}$) and teachers ($\pi_{teacher,t}$) the full number of rapid tests is reached after the Easter holidays in all scenarios. For rapid tests in the workplace ($\pi_{w,s,t}$) and private rapid tests ($\pi_{private,t}$) it is reached between May 1 and June 10, depending on the scenario.

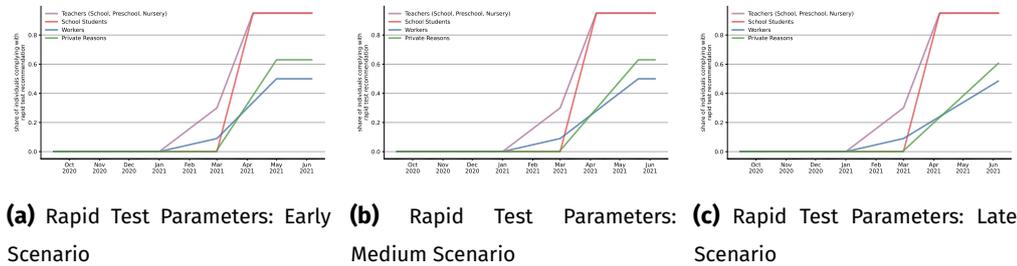


Figure 1.B.7. Rapid Test Introduction in the Three Scenarios

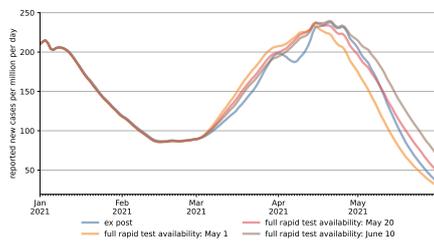
Note: Number of rapid tests performed in the different prediction scenarios. All scenarios are the same until March 1 and have the same level of rapid tests when all supply constraints are resolved. They differ in the date at which the full number of tests is reached. For students ($\pi_{students,t}$) and teachers ($\pi_{teacher,t}$) the full number of rapid tests is reached after the Easter holidays in all scenarios. For rapid tests in the workplace ($\pi_{w,s,t}$) and private rapid tests ($\pi_{private,t}$) it is reached between May 1 and June 10, depending on the scenario.

Moreover, the out of sample predictions assume that the share of detected cases (ψ_t) that would have been obtained without rapid tests is not affected by the Easter holidays because the extent to which this was the case was estimated from case numbers in April.

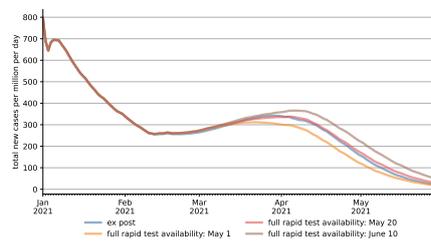
The results of the out of sample prediction are displayed in Figure 1.B.8. While all scenarios considerably deviate from the ex-post scenario, they all reproduce the steep increase of cases until the end of April, followed by a decline until June. We can therefore conclude that our main results are not sensitive to measurement errors in the number of rapid tests, vaccinations or mobility data.

Another form of validating our model is to see how well our main results align with other studies that evaluate the effect of large scale rapid testing. Of course, this has to be taken with a grain of salt as the effect of any rapid testing policy depends on the incidence of the disease in the population, how well other testing policies such as PCR tests are working, the effect of seasonality and NPIs that are in place.

Nevertheless, it is reassuring that other studies find effect sizes in the same order of magnitude.



(a) Reported Cases



(b) Total Cases

Figure 1.B.8. Out of Sample Prediction for Reported and Total Cases from March to June 2021.

Note: The ex-post scenario is an in-sample prediction that uses all available information and is very close to actual case numbers. For the other scenarios data on vaccinations, work mobility and rapid tests that became available after March 1 have been replaced by prediction models that are calibrated with data from February. Moreover, they do not model a lower number of detected cases over the Easter holidays. The different scenarios make different assumptions on the date at which full availability of rapid tests is reached. While the out of sample predictions differ substantially for the exact case numbers at the beginning of June (between 20 and 70 cases per million), they can all reproduce the decline in case numbers that is jointly driven by seasonality, large scale rapid tests and vaccinations. For legibility reasons, all lines are rolling 7-day averages. Each line is the average over 30 simulation runs.

Pavelka, Van-Zandvoort, Abbott, Sherratt, Majdan, et al. (2021) estimate that a mass testing campaign in Slovakia in October and November 2020 where approximately 65 % of the population took a rapid test within a two week period lead to a reduction in case numbers of 70 % three weeks after the start of the intervention. Moreover, they find that this strong reduction in cases cannot be explained by isolation of people who tested positive alone but only when they took into account that household members of people who tested positive reduced their contacts.

While we do not model the exact scenario of Pavelka et al. (2021), we can roughly compare their estimates with our predictions for the difference between the baseline scenario and a scenario without rapid tests. In May about 45% of people do at least one rapid test in every week. Taking into account that there are many repeated testers the number of people who do a test within a two week period is probably slightly less than the 65% from the intervention in Slovakia. On the other hand, we have many people who do more than one rapid test in that time which also leads to the detection of cases. Our model predicts that the observed incidence with tests is approximately 65% lower than without tests after three weeks. Thus we have an effect size in the same order of magnitude but are slightly less optimistic regarding the efficiency of rapid tests.

Berger, Fritz, and Kauermann (2021) analyse the effect of twice weekly rapid testing in schools. They have two main findings: Firstly, rapid tests reduced the share of undetected cases among students by a factor between two and four. Secondly, open schools with mandatory testing might lead to the same or even lower numbers

of infections than closed schools. The estimates are based on infection numbers after the Easter holiday.

Again, we do not directly simulate their scenarios but can roughly compare our results to theirs. We estimate a share of undetected cases of approximately 75% among school age children (five to 14 years) at the beginning of April, see Figure 1.B.11. This drops to slightly less than 40% at the end of our simulation period. Thus in the long run, mandatory tests at schools led to a reduction of the share of undetected cases by a factor of more than 1.8 which is just slightly below the factor of two to four predicted by Berger, Fritz, and Kauermann (2021).

Similarly we are slightly less optimistic for the effect of opening schools with testing compared to closing schools. While they predict that opening schools could even be beneficial we estimate that it would lead to a slight increase in case numbers see Figure 1.B.16).

1.B.12 Robustness to assumptions about rapid test sensitivity

Our main results are based on rapid test sensitivities read from clinical trials. Recent studies showing that the actual sensitivity of rapid tests may be lower than that (e.g., Scheiblauer et al., 2021).

This section shows that our results are robust to making less favorable assumptions on rapid test sensitivity. We proceed by describing several possible ways of calibrating rapid test sensitivity profiles based on recent studies. Since none of these methods is inherently better than the others, we make simulations with two sensitivity profiles: The average over all methods and the lower envelope over all methods using recent studies.

Both profiles imply lower sensitivities of rapid tests than used in our main results. This is especially true during the later stage of an infection. However, the main results stay very robust. The original result was that rapid tests, seasonality and vaccinations are responsible for 42%, 43% and 16%, respectively. With the average profile, the effect of rapid tests decreases to 41 %. With the lower envelope profile, which is an extremely unfavorable assumption, it becomes 38%.

The effect of rapid tests on infection dynamics strongly depends on *when* an infection is detected. Earlier detection means that it is more likely that the infection has not yet been discovered for a different reason (e.g. due to the onset of symptoms) and that the infected person can be isolated before spreading the disease to others.

The sensitivity of rapid tests depends on the viral load in the respiratory tract. It is low at the beginning of an infection (especially before the onset of infectiousness), high in the first few days of infectiousness and then gradually decreasing towards the end of infectiousness.

We thus need to calibrate a profile of rapid test sensitivities based on the number of days until or since the onset of infectiousness.

Unfortunately, such sensitivity profiles are not usually reported in studies. We thus need to create them by combining two types of studies: 1. Studies that report the viral load in terms of threshold cycle (C_t) values determined by PCR test (e.g. Jang, Rhee, Wi, and Jung, 2021; Ong, Chiew, Ang, Mak, Cui, et al., 2021; Zuin, Gentili, Cervellati, Rizzo, and Zuliani, 2021; Bonnet, Masse, Benamar, Vilcu, Swital, et al., 2022; Cosentino, Bernard, Ambroise, Giannoli, Guedj, et al., 2022). 2. Studies that report the sensitivity of rapid tests for different C_t values (e.g. Brümmer et al., 2021; Scheiblaue et al., 2021).

It is natural to assume that the evolution of C_t values over time as well as the effect of C_t values on rapid test sensitivity are continuous functions. However, the results of the available studies are usually reported in a discretized way. This leads to multiple ways of calculating the sensitivity profiles. Some try to recover the underlying continuous functions using interpolation or regression, others simply use the discretized values.

For the calibration of C_t values over time we can either use discretized values for several time bins from Ong et al. (2021) and Jang et al. (2021). Alternatively, we can use linearized formulas for calculating sensitivities over time Cosentino et al. (2022) and complement it with interpolations of data points from Jang et al. (2021) in the pre-infectious stage. Throughout we assume that the C_t values of individuals who eventually develop symptoms and those who do not follow the same trajectory. This is in line with results by Zuin et al. (2021) and recent evidence that rapid tests excel at discovering asymptomatic cases Rosella, Agrawal, Gans, Goldfarb, Sennik, et al. (2022).

For the mapping of C_t values to rapid test sensitivities, again we have two options. First, we can simply look up the discretized values for the three C_t bins provided in Scheiblaue et al. (2021) (below 20, 25 to 30 and above 35). Secondly, we can use linear regression to estimate a continuous mapping for the relationship by assuming that the C_t values of each bin are achieved exactly at the bin midpoints and that the relationship is linear.

In general, using discretized values can lead to an underestimation of peak sensitivities and an overestimation of very low sensitivities. This is because discretization is essentially a smoothing device. On the other hand, it has the advantage of simply working with published results, without introducing any tuning parameters or other assumptions.

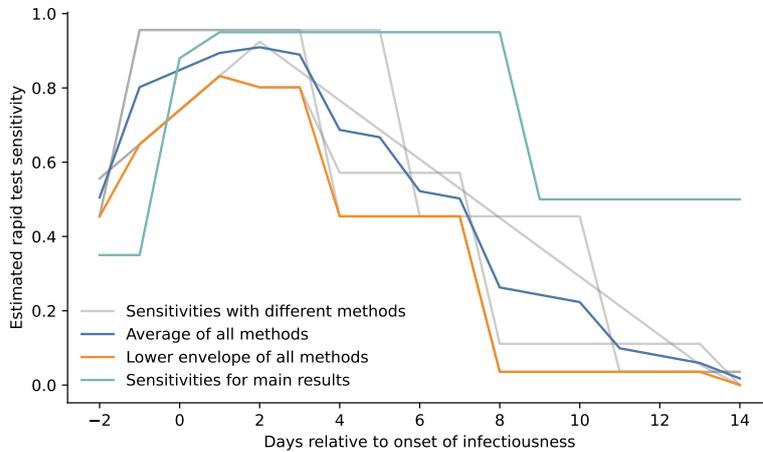
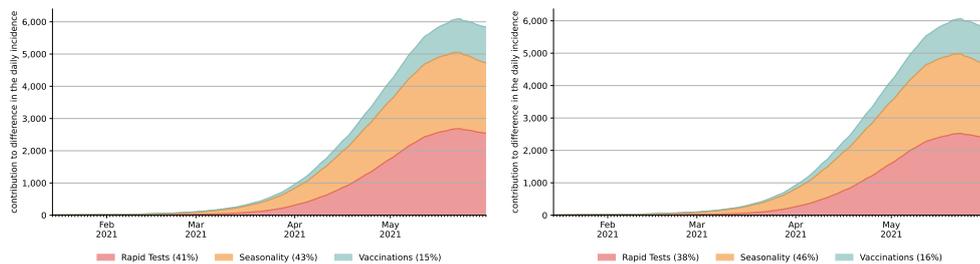


Figure 1.B.9. Rapid test sensitivity profiles

Note: The figure shows estimated sensitivities of rapid tests over the course of an infection. The x-axis shows days relative to the onset of infectiousness. The y-axis shows the estimated rapid test sensitivities. The grey lines are the raw sensitivity estimates obtained with different methods of dealing with discretized data. The blue line shows their average and the yellow line their lower envelope. The turquoise line are the test sensitivities used for the main results of the paper.

Figure 1.B.9 shows that the updated sensitivity estimates are lower than the ones used for our original results, especially towards the end of an infection. However, the main results barely change. This is due to the fact that the differences are largest towards the later stage of an infection. Uncovering an infection that was previously undetected at that stage does not have a large effect on infection dynamics.



(a) Shapley decomposition with average of sensitivity estimates **(b)** Shapley decomposition with lower envelope of sensitivity estimates

Figure 1.B.10. Shapley decompositions for different values of rapid test sensitivity

Note: The figure shows updated versions of the Shapley decomposition in figure 1.4.1. The decompositions are based on 20 model runs with different random seeds. The share attributed to each channel is rounded to the next full percentage point to acknowledge the remaining sampling uncertainty.

1.B.13 Share of Detected Cases

This section shows the share of detected cases for different age groups. See Section 1.B.7 for an explanation of how we model the detection of cases and Section 1.A.7 for the calibration of the relevant parameters.

The share of detected cases fall drastically from October to December when the incidence of CoViD-19 skyrocketed, PCR tests were still scarce and official contact tracing became impossible due to the sheer amount of cases.

As rapid tests become available and more and more individuals receive positive rapid tests and seek PCR tests, the share of detected cases starts to increase. While first rapid tests are available since the beginning of 2021 the effect only becomes substantial after March when access to rapid tests was greatly expanded.

Overall, the share of detected cases is much higher in older age groups. This is because the likelihood to develop symptoms increases with age and symptomatic cases are more likely to be detected.

A notable exception is that school age children (5-14, green line) overtake the next age group in May 2021. This comes from a particularly strong increase in their share of detected cases after Easter, when weekly rapid tests become mandatory in schools.

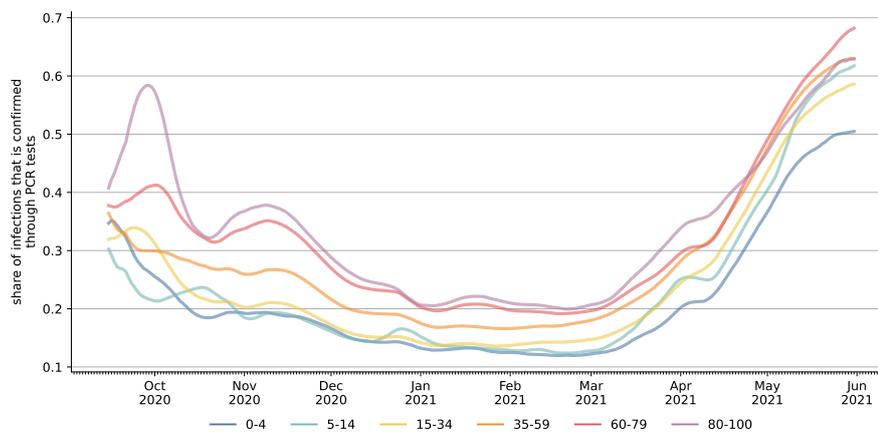


Figure 1.B.11. Share of Detected Cases by Age Group

Note: The figure shows the share of cases that is reported as an official case for each age group in our simulated population. For legibility reasons, all lines are rolling 7-day averages of the average of 30 simulation runs.

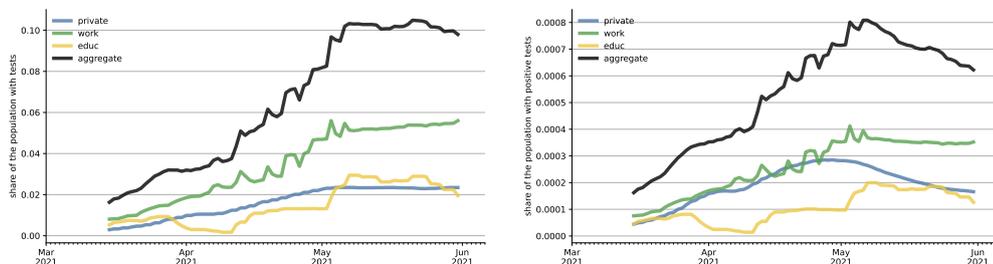
1.B.14 Simulated Rapid Tests

In order to make the most use out of limited data sources on rapid test usage, we model the number of performed rapid tests as a result of time invariant willingness to do rapid tests and time varying supply side factors and events that trigger rapid

tests. Thus, the π parameters governing when individuals do rapid tests described in Section 1.A.6 are only indirectly related to the number of rapid tests that are actually performed in the model. When it comes to positive and negative rapid tests, there is even an additional layer because rapid tests are imperfectly sensitive and specific.

In this section we look at how rapid tests expanded in our simulations over time and to what degree they are useful as a screening device despite their imperfections.

We start with the share of the population doing a rapid test and receiving a positive rapid test over time by the channel through which the test was demanded in Figures 1.B.12a, 1.B.12b, respectively. Overall, the share of the population getting a rapid test on a given day increases from 2% in mid March to over 10% by May. The work rapid tests are a little ragged because of public holidays. For education rapid tests both vacations (first half of April) as well as the opening of schools in May are very visible in the rapid test demand. Overall, work tests make up the largest fraction of rapid tests. The image is very similar for the share of positive tests, except that the overall number of positive tests starts decreasing in May as rapid test expansion comes to a halt and cases fall, especially the positive share of private rapid tests falls as less and less individuals are triggered to seek a rapid test because of a risk contact in their household.



(a) Share of the Population Doing a Rapid Test on a Given Day, by Channels (b) Share of the Population Testing Positive on a Given Day, by Channels

Figure 1.B.12. Rapid Test Shares in the Population by Channel

Note: Rapid tests in the education setting are demanded by teachers (nursery, preschool and school) as well as pupils. After Easter the required frequency of tests is increased from once per week to twice per week. Work rapid tests are demanded by individuals that still have work contacts, i.e. do not work from home. The share of employers offering rapid tests increases over the time frame and the frequency of testing is also increased. Private tests are demanded by individuals for one of three reasons: having developed symptoms without access to a PCR test, having a household member that has tested positive or developed symptoms or having planned a weekly meeting with friends. Panel a shows the share of the population doing a rapid test on a given day. Panel b shows the share of the population testing positive on a given day (true and false positives).

Next, we show the tests split by whether they are true positive, false positive, true negative or false negative (see Figure 1.B.13) in numbers per million individuals to make the metric comparable to incidences.

The number of true positives (Figure 1.B.13a) rapidly increases and peaks at the end of April with over 200 cases per million detected through rapid tests per day. This means that our model suggests that Germany was able to detect up to 16,600 cases per day that would have likely gone undetected otherwise. The most powerful tool for detecting cases are the private rapid tests. This is because a large share of them are targeted, i.e. triggered by events in the household. However, this does not mean that rapid tests in the workplace or at school are less important. It is rather the combination of large scale screening at work and in schools and very efficient follow up tests whenever those screening tests detected a case. Shapley values (Figure 1.4.1) take this into account and assign about 50% of the overall reduction of case numbers via rapid tests to private rapid tests with work and school rapid tests accounting for 40% and 7%, respectively.

Such a large effect of rapid tests seems to be at odds with the general perception that they are not very reliable. However, one has to differentiate between the reliability of one test in isolation and the effect imperfect tests can have when employed at a large scale. On average our tests have a sensitivity of slightly more than 70%. This means they miss almost 30% of infections among the tested. Of course perfect tests would have an even larger effect but the relevant number to compare is that up to 200 cases per million are detected by rapid tests every day which would have otherwise gone undetected.

This clearly shows that the large effect of rapid tests on the infection dynamic is not driven by unrealistic assumptions about their sensitivity but rather by the fact that there was a very large number of infected individuals who did not know they are infected. Detecting and isolating some of them is enough to slow down the overall infection dynamic.

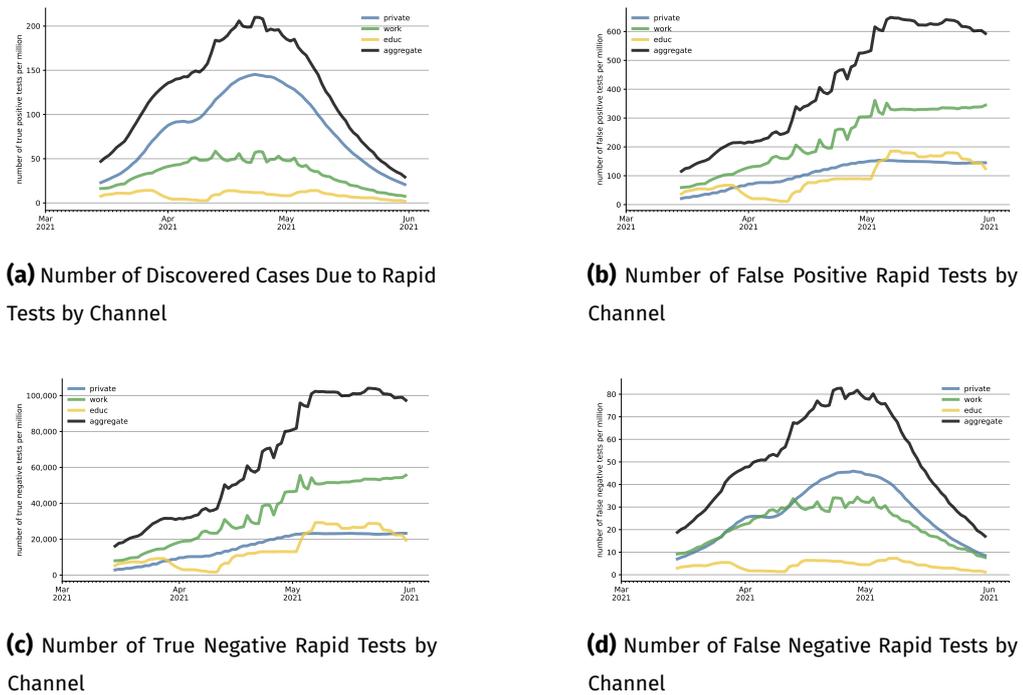
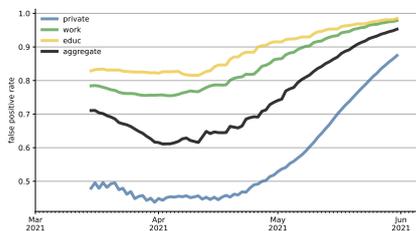


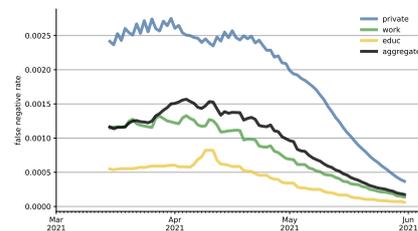
Figure 1.B.13. Simulated Rapid Test Statistics

Note: Each panel shows the number of rapid tests per million inhabitants that fall into the respective category. Private rapid tests are especially good at detecting cases but since they are often triggered by rapid tests from other channels, the other groups of tests, especially rapid tests at the workplace, also play an important role for containing the pandemic. All results are averaged over 30 simulation runs. For legibility reasons, all lines are rolling 7-day averages.

A similar picture arises, when looking at the false positive rate, i.e. the share of positive tests that go to people who are not infected. Figure 1.B.14a shows that the false positive rate is very high. On average 60% to 93% of positive tests are received by individuals that are not infected. The false positive rate increases over time. This is due to the low prevalence of infections in the population, which falls over time. Again, private rapid tests are an exception with a much lower false positive rate because those tests are primarily demanded when there is a high likelihood of being infected. The false negative rate of 0.2% looks very low. As discussed above this is deceiving and just a mechanical consequence of a very low prevalence of the disease and the many rapid tests done by non-infected people.



(a) Rate of False Positive Rapid Tests by Channel



(b) Rate of False Negative Rapid Tests by Channel

Figure 1.B.14. False Positive and False Negative Rates by Channel

Note: The left panel shows the share of positive tests that are given to people who are not infected. This share is large as can be expected with a very low baseline rate of positive individuals. As the incidence in the population drops, the false positive rate increases. An exception are the private rapid tests because they are – especially when the incidence is high – often triggered by events that make it likely that the test taker is infected and therefore their false positive rate is much lower. The right panel shows the false negative rate in the population, i.e. the share of negative tests done by infected individuals. This is very low because there are many truly negative tests in times of low incidences and large scale screening tests.

1.B.15 Scenarios

Here we complement our analysis of the effectiveness of vaccinations and rapid tests by showing the effects of rapid test policies vis-à-vis the more traditional NPIs, work from home mandates and school closures. All scenarios start after Easter (April 6). Our analyses show that many socially costly NPIs can be avoided through strong rapid testing policies.

Figure 1.B.15 shows the effects of different work policies on the infections in the general population. We compare four scenarios with our baseline scenario: Keeping the share of workers having physical work contacts the same as in our baseline scenario the orange line shows what would have happened with rapid testing in firms at the level of mid March (orange line) where only 14% of workers regularly

did rapid tests. We also include a scenario what would have happened if rapid tests had become truly mandatory after Easter⁹, assuming a 95% compliance rate on both the employer and the employee side. On the work from home dimension we compare our baseline scenario with 10% more or less work from home compared to the baseline scenario. For the total cases, the picture is very clear. Given the testing policy Germany had in place during that time (twice weekly tests done by 35% to 50% of workers over that time frame) whether 70% (10% below the actual mobility) or 85% (10% above the actual mobility) of workers attend work physically makes little difference for the incidence. On the other hand, the effect of a laxer or more ambitious testing policy for firms is sizable: As can be seen in Figure 1.B.15b the gap between the two scenarios grows to over 80 incidence points around May 1. As in other scenarios, the observed cases can be misleading because more testing leads to more detected cases. It takes two to three weeks for the reduction in new infections to dominate the increased detection. Furthermore, the two opposing effects lead to a smaller effect size than is actually the case.

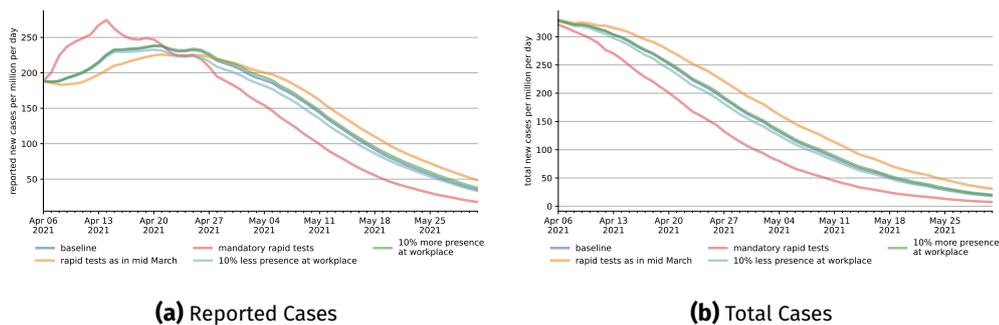


Figure 1.B.15. The Effect of Different Work Scenarios on Reported and Total Cases

Note: The figure shows the development of cases after different hypothetical work policy changes take place at Easter until the end of our simulation period. We vary the share of workers that have physical work contacts (10% more or less compared to the share in the baseline scenario, 85% or 70% of workers, respectively) and how many tests are performed at work relative to our baseline scenario. As an ambitious scenario we implement mandatory tests for all employees that do not work from home, assuming 95% compliance on both the employer and the employee side. On the other hand, we show what would have happened if the test offers had fallen back to the level of mid March (only 14% of workers are tested regularly). The observed cases can be misleading because more testing leads to more detected cases. It takes two to three weeks for the reduction in new infections to dominate the increased detection. Furthermore, the two opposing effects lead to a smaller effect size than is actually the case.

9. Starting on April 19th employers were required by law to provide two weekly tests to their employees (Bundesanzeiger, 2021b). However, voluntarily only 60% of workers regularly test themselves when offered tests (Betsch, Korn, Felgendreff, Eitze, Schmid, Sprengholz, Wieler, Schmich, Stollorz, Ramharter, Bosnjak, Omer, Thaiss, De Bock, and Von Rden (2021), 20th/21st of April).

The second commonly employed and also very contentious NPI we look at are school closures. Due to the very high incidence we model the German schooling policy as generous emergency care with rotating on-site schooling for graduating classes for April. In May where cases fall and schools gradually opened, we model the policy as rotating on-site schooling for most students (except for children eligible for emergency care and graduating classes who attend in full). We compare this baseline scenario to simply keeping schools completely closed (the brown line) and opening schools normally (but maintaining our hygiene multiplier to account for mask wearing, ventilation etc.) with and without tests.

As can be seen, the transmission potential in schools is very low both in the generous emergency setting as well as the rotating operation. The difference to keeping schools completely closed is very small. Also, consistent testing reduces the transmission potential at schools strongly. Had schools opened directly after Easter given the testing rates Germany managed at schools during that time, the total incidence would have been only 9 incidence points higher on average. Tests, however, are crucial here. Had schools opened completely without any testing of students and staff, schools would have added up to 50 incidence points.

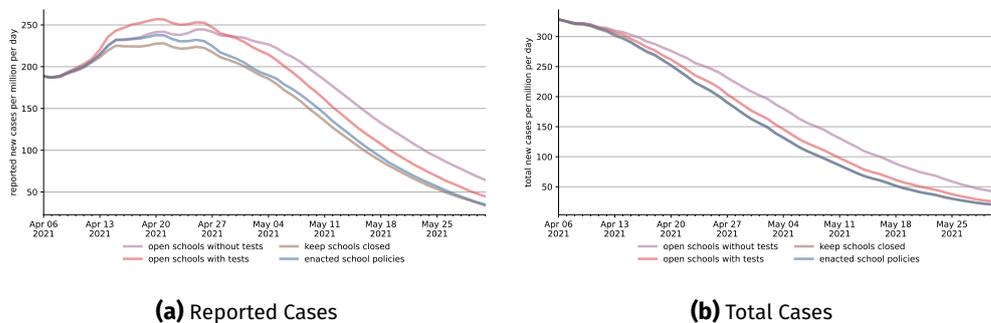


Figure 1.B.16. The Effect of Different School Scenarios on Reported and Total Cases

Note: The figure shows the development of cases after different hypothetical school policy changes take place at Easter until the end of our simulation period. Apart from the enacted school policies as our baseline we simulate how cases would have developed if schools had been closed completely as the strictest possible counterfactual scenario and two opening models: One where schools open normally (with hygiene measures) without any testing in the education sector and one where schools open normally but testing develops as in the baseline scenario.

Lastly, we shed some light on the role our rapid test demand channels play for the effect of rapid tests on case numbers. To do so we ran two scenarios where we allocated rapid tests either completely randomly in the entire population or among

70% of the population to account for the fact that a share of the population might refuse or be very hard to reach with rapid tests.¹⁰

Figure 1.B.17 shows how the incidence of detected and total cases develops in the two random scenarios (red and purple line) relative to our baseline scenario (blue line). Two things stand out: Firstly, the total number of cases falls much faster in our baseline scenario compared to the two random scenarios. Secondly, this is not because the share of detected cases is higher in the baseline scenario; in fact, it is even slightly lower until end of April.

There are two mechanisms that can explain these surprising facts: Firstly, tests at the workplace predominantly target a group that has many contacts. Thus, catching infections in this group prevents more infections than in the general population. Secondly, rapid tests that are done because of private contact tracing are more effective at interrupting infection chains because they catch many infections in an early stage. Isolating infected individuals early on means that there are fewer days on which they can infect others. The difference between the two random scenarios are small. This is likely due to only a small fraction of the population being tested on any given day.

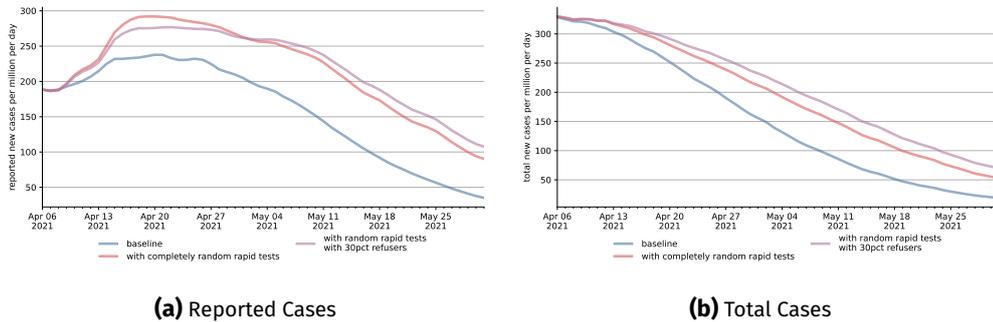


Figure 1.B.17. The Role of Targeted and Compliance Driven Rapid Test Demand

Note: The figure shows the development of cases in two scenarios where rapid tests are distributed randomly in the population compared to our baseline scenario after Easter. In the baseline scenario rapid tests are targeted to workers, students, teachers and individuals at high risk of being infected including a weekly or twice weekly spacing between rapid tests. In the scenario with 30% refusers tests are randomly distributed among 70% of the population who are identified as compliers.

10. We calculate the number of rapid tests as in our baseline model. This leads to similar numbers of rapid tests. However given the higher incidence in our random scenarios these scenarios have a slightly higher number of rapid tests.

References

- Abu-Raddad, Laith J, Hiam Chemaitelly, Joel A Malek, Ayeda A Ahmed, Yasmin A Mohamoud, Shameem Younuskuju, Houssein H Ayoub, Zaina Al Kanaani, Abdullatif Al Khal, Einas Al Kuwari, Adeel A Butt, Peter Coyle, Andrew Jeremijenko, Anvar Hassan Kaleeckal, Ali Nizar Latif, Riyazuddin Mohammad Shaik, Hanan F Abdul Rahim, Hadi M Yassine, Mohamed G Al Kuwari, Hamad Eid Al Romaihi, Mohamed H Al-Thani, and Roberto Bertollini.** 2020. "Assessment of the Risk of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Reinfection in an Intense Reexposure Setting." *Clinical Infectious Diseases*, [19]
- Abueg, Matthew, Robert Hinch, Neo Wu, Luyang Liu, William Probert, Austin Wu, Paul Eastham, Yusef Shafi, Matt Rosencrantz, Michael Dikovsky, Zhao Cheng, Anel Nurtay, Lucie Abeler-Dörner, David Bonsall, Michael V. McConnell, Shawn O'Banion, and Christophe Fraser.** 2021. "Modeling the effect of exposure notification and non-pharmaceutical interventions on COVID-19 transmission in Washington state." *npj Digital Medicine* 4 (1): [46]
- Ahlers, Elke, Malte Lübker, and Rainer Jung.** 2021. "Schleppender Start für wöchentliche Corona-Schnelltests am Arbeitsplatz: Nur 23 Prozent der Beschäftigten haben schon Zugang." Working paper. Wirtschafts- und Sozialwissenschaftliches Institut. [28]
- Aleta, Alberto, David Martín-Corral, Ana Pastore y Piontti, Marco Ajelli, Maria Litvinova, Matteo Chinazzi, Natalie E. Dean, M. Elizabeth Halloran, Ira M. Longini Jr, Stefano Merler, Alex Pentland, Alessandro Vespignani, Esteban Moro, and Yamir Moreno.** 2020. "Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19." *Nature Human Behaviour* 4 (9): 964–71. [8, 10, 46, 47]
- Anaconda.** 2016. "Anaconda Software Distribution." Version Vers. 2-2.4.0. [46]
- Anderson, Elizabeth L., Paul Turnham, John R. Griffin, and Chester C. Clarke.** 2020. "Consideration of the Aerosol Transmission for COVID-19 and Public Health." *Risk Analysis* 40 (5): 902–7. [52]
- Avery, Christopher, William Bossert, Adam Clark, Glenn Ellison, and Sara Fisher Ellison.** 2020. "An Economist's Guide to Epidemiology Models of Infectious Disease." *Journal of Economic Perspectives* 34 (4): 79–104. [50]
- Basurto, Alessandro, Herbert Dawid, Philipp Harting, Jasper Hepp, and Dirk Kohlweyer.** 2020. "Economic and Epidemic Implications of Virus Containment Policies: Insights from Agent-Based Simulations." *SSRN Electronic Journal*, [47]
- Bayerische Staatskanzlei.** 7, 2021. *Bericht aus der Kabinettsitzung vom 7. April 2021*. Pressemitteilung. URL: <https://www.bayern.de/bericht-aus-der-kabinettsitzung-vom-7-april-2021/?seite=2453> (visited on 06/24/2021). [29]
- Bayerisches Staatsministerium für Familie, Arbeit und Soziales.** 14, 2021a. *Kindertagesbetreuung: Aktuelle Informationen zum Coronavirus (SARS-CoV-2)*. URL: <https://www.stmas.bayern.de/coronavirus-info/corona-kindertagesbetreuung.php> (visited on 07/14/2021). [27]
- Bayerisches Staatsministerium für Familie, Arbeit und Soziales.** 2021b. 404. *Newsletter: Allgemeine Informationen zur Kindertagesbetreuung: Informationen zum Coronavirus (SARS-CoV-*

2). URL: https://www.stmas.bayern.de/imperia/md/content/stmas/stmas_inet/service-kinder/newsletter/404-newsletter.pdf (visited on 03/08/2021). [27]

Bayerisches Staatsministerium für Gesundheit und Pflege. 26, 2021. *Holetschek: Bayern hat frühzeitig Selbsttests gesichert - Schulen und Kitas bekommen schon nächste Woche 1,3 Millionen Selbsttests.* Pressemitteilung Nr. 44/GP. URL: <https://www.stmgp.bayern.de/presse/holetschek-bayern-hat-fruehzeitig-selbsttests-gesichert-schulen-und-kitas-bekommen-schon/> (visited on 06/24/2021). [28]

Bayerisches Staatsministerium für Unterricht und Kultus. 22, 2021. *Informationen zum Unterrichtsbetrieb nach den Osterferien.* URL: <https://www.km.bayern.de/lehrer/meldung/7244/informationen-zum-unterrichtsbetrieb-nach-den-osterferien.html>. [26]

Bayerisches Staatsministerium für Unterricht und Kultus. No date. *FAQ zum Unterrichtsbetrieb an Bayerns Schulen.* URL: <https://www.km.bayern.de/allgemein/meldung/7047/faq-zum-unterrichtsbetrieb-an-bayerns-schulen.html>. Accessed on 2021-04-30. [26]

Berger, Ursula, Cornelius Fritz, and Göran Kauermann. 30, 2021. "CODAG Bericht Nr. 14: 2. Schulschließungen oder Schulöffnung mit Testpflicht? Epidemiologisch-statistische Aspekte sprechen für Schulöffnungen mit verpflichtenden Tests." Working paper. Ludwig-Maximilians-Universität München. [16, 65, 66]

Bertelsmann Stiftung. 2019. "Ländermonitor Frühkindliche Bildungssysteme 2019." Working paper. Bertelsmann Stiftung. [20]

Betsch, Cornelia, Lars Korn, Lisa Felgendreff, Sarah Eitze, Philipp Schmid, Philipp Sprengholz, Lothar Wieler, Patrick Schmich, Volker Stollorz, Michael Ramharter, Michael Bosnjak, Saad B. Omer, Heidrun Thaiss, Freia De Bock, and Ursula von Rügen. 26, 2021. *Zusammenfassung und Empfehlungen Welle 37.* URL: <https://projekte.uni-erfurt.de/cosmo2020/web/summary/37/> (visited on 06/27/2021). [29]

Betsch, Cornelia, Lars Korn, Lisa Felgendreff, Sarah Eitze, Philipp Schmid, Philipp Sprengholz, Lothar Wieler, Patrick Schmich, Volker Stollorz, Michael Ramharter, Michael Bosnjak, Saad B. Omer, Heidrun Thaiss, Freia De Bock, and Ursula Von Rügen. 2021. "COVID-19 Snapshot Monitoring (COSMO Germany) - Wave 41." [28, 33, 44, 45, 47, 56, 74]

Betsch, Cornelia, Lothar Wieler, Michael Bosnjak, Michael Ramharter, Volker Stollorz, Saad Omer, Lars Korn, Philipp Sprengholz, Lisa Felgendreff, Sarah Eitze, et al. 2021. "Germany COVID-19 Snapshot Monitoring (COSMO Germany): Rapid Tests." [63]

Bi, Qifang, Yongsheng Wu, Shujiang Mei, Chenfei Ye, Xuan Zou, Zhen Zhang, Xiaojian Liu, Lan Wei, Shaun A Truelove, Tong Zhang, Wei Gao, Cong Cheng, Xiujuan Tang, Xiaoliang Wu, Yu Wu, Binbin Sun, Suli Huang, Yu Sun, Juncen Zhang, Ting Ma, Justin Lessler, and Tiejian Feng. 2020. "Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study." *The Lancet Infectious Diseases* 20 (8): 911–19. [18]

Bicher, Martin, Claire Rippinger, Christoph Urach, Dominik Brunmeir, Uwe Siebert, and Niki Popper. 2021. "Evaluation of Contact-Tracing Policies against the Spread of SARS-CoV-2 in Austria: An Agent-Based Simulation." *Medical Decision Making*, 0272989X2110133. [47]

- Bonin, Holger, Annabelle Krause-Pilatus, and Ulf Rinne.** 15, 2021. "Verbreitung von Homeoffice und Corona-Tests der Arbeitgeber Ende März-Anfang April 2021. Ergebnisse einer repräsentativen Befragung von abhängig Beschäftigten." *Forschungsbericht* 570/3. [28]
- Bonnet, Camille, Shirley Masse, Hayat Benamar, Ana-Maria Vilcu, Morgane Swital, Thomas Hanslik, Sylvie van der Werf, Xavier Duval, Fabrice Carrat, Alessandra Falchi, and Thierry Blanchon.** 2022. "Is the Alpha Variant of SARS-CoV-2 Associated with a Higher Viral Load than the Historical Strain in Saliva Samples in Patients with Mild to Moderate Symptoms?" *Life* 12(2): [67]
- Brümmer, Lukas E, Stephan Katzenschlager, Mary Gaeddert, Christian Erdmann, Stephani Schmitz, Marc Bota, Maurizio Grilli, Jan Larmann, Markus A Weigand, Nira R Pollock, et al.** 2021. "Accuracy of novel antigen rapid diagnostics for SARS-CoV-2: A living systematic review and meta-analysis." *PLoS medicine* 18(8): e1003735. [10, 44, 56, 67]
- Bundesanzeiger.** 8, 2021a. "Verordnung zum Anspruch auf Testung in Bezug auf einen direkten Erregernachweis des Coronavirus SARS-CoV-2 (Coronavirus-Testverordnung - TestV)." [29]
- Bundesanzeiger.** 2021b. "Zweite Verordnung zur Änderung der SARS-CoV-2-Arbeitsschutzverordnung." [28, 74]
- Bundesgesetzblatt.** 22, 2021. "Viertes Gesetz zum Schutz der Bevölkerung bei einer epidemischen Lage von nationaler Tragweite vom 22. April 2021." [26]
- Bundesministerium für Wirtschaft und Energie.** 8, 2021. *Corona-Tests in der Wirtschaft: Ergebnisse der Umfragen bei Beschäftigten und Unternehmen.* Pressemitteilung. URL: <https://www.bmwi.de/Redaktion/DE/Pressemitteilungen/2021/04/20210408-corona-tests-in-der-wirtschaft-ergebnisse-umfragen.html> (visited on 06/24/2021). [28]
- Byrne, Andrew William, David McEvoy, Aine B Collins, Kevin Hunt, Miriam Casey, Ann Barber, Francis Butler, John Griffin, Elizabeth A Lane, Conor McAloon, Kirsty O'Brien, Patrick Wall, Kieran A Walsh, and Simon J More.** 2020. "Inferred duration of infectious period of SARS-CoV-2: rapid scoping review and analysis of available evidence for asymptomatic and symptomatic COVID-19 cases." *BMJ Open* 10(8): e039856. [18]
- Chen, Jun, Tangkai Qi, Li Liu, Yun Ling, Zhiping Qian, Tao Li, Feng Li, Qingnian Xu, Yuyi Zhang, Shuibao Xu, Zhigang Song, Yigang Zeng, Yinzhong Shen, Yuxin Shi, Tongyu Zhu, and Hongzhou Lu.** 2020. "Clinical progression of patients with COVID-19 in Shanghai, China." *Journal of Infection* 80(5): e1–e6. [18]
- Cheng, Yafang, Nan Ma, Christian Witt, Steffen Rapp, Philipp S. Wild, Meinrat O. Andreae, Ulrich Pöschl, and Hang Su.** 2021. "Face masks effectively limit the probability of SARS-CoV-2 transmission." *Science*, [8]
- conda-forge community.** 2015. "The conda-forge Project: Community-based Software Distribution Built on the conda Package Format and Ecosystem." [46]
- Contreras, Sebastian, Jonas Dehning, Sebastian B Mohr, Simon Bauer, F Paul Spitzner, and Viola Priesemann.** 2021. "Low case numbers enable long-term stable pandemic control without lockdowns." *Science advances* 7(41): [7, 15]

- Cosentino, Gina, Mathieu Bernard, Joëvin Ambroise, Jean-Marc Giannoli, Jérémie Guedj, Florence Débarre, and François Blanquart.** 2022. "SARS-CoV-2 viral dynamics in infections with Alpha and Beta variants of concern in the French community." *Journal of Infection* 84(1): 94–118. [67]
- Cuevas, Erik.** 2020. "An agent-based model to evaluate the COVID-19 transmission risks in facilities." *Computers in Biology and Medicine* 121(6): 103827. [46]
- Davies, Nicholas G., Sam Abbott, Rosanna C. Barnard, Christopher I. Jarvis, Adam J. Kucharski, James D. Munday, Carl A. B. Pearson, Timothy W. Russell, Damien C. Tully, Alex D. Washburne, Tom Wenseleers, Amy Gimma, William Waites, Kerry L. M. Wong, Kevin van Zandvoort, Justin D. Silverman, Karla Diaz-Ordaz, Ruth Keogh, Rosalind M. Eggo, Sebastian Funk, Mark Jit, Katherine E. Atkins, and W. John Edmunds.** 2021. "Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England." *Science*, (3): [18, 42]
- Davies, Nicholas G., Petra Klepac, Yang Liu, Kiesha Prem, Mark Jit, and Rosalind M. Eggo.** 2020. "Age-dependent effects in the transmission and control of COVID-19 epidemics." *Nature Medicine* 26(8): 1205–11. [8, 18, 42, 54]
- Delli Gatti, Domenico, and Severin Reissl.** 2020. "ABC: An Agent Based Exploration of the Macroeconomic Effects of Covid-19." *CESifo Working Paper Series*, [47]
- Destatis.** 2020. "Betreuungsquote von Kindern unter 6 Jahren nach Bundesländern." Working paper. Statistisches Bundesamt. [20]
- Deutscher Industrie- und Handelskammertag.** 2021. "Umfrage der IHK-Organisation zu Corona-Tests in den Unternehmen." Working paper. [28]
- Donsimoni, Jean Roch, René Glawion, Bodo Plachter, and Klaus Wälde.** 2020. "Projecting the spread of COVID-19 for Germany." *German Economic Review* 21(2): 181–216. [54]
- Fetzer, Thimo, and Thomas Graeber.** 2021. "Measuring the scientific effectiveness of contact tracing: Evidence from a natural experiment." *Proceedings of the National Academy of Sciences* 118(33): [15]
- Forschungsdatenzentren Der Statistischen Ämter Des Bundes Und Der Länder.** 2018. "Mikrozensus 2010, CF, Version 0." de. [20, 23]
- Frisch, Julia.** 2, 2021. *Deutsche mit hoher Corona-Impfbereitschaft – aber Wissenslücken.* URL: <https://www.aerztezeitung.de/Politik/Deutsche-mit-hoher-Corona-Impfbereitschaft-aber-Wissensluecken-417569.html>. [43]
- Gabler, Janos.** 2020. "A Python Tool for the Estimation of (Structural) Econometric Models." [46]
- Gabler, Janoš, Tobias Raabe, Klara Röhr, and Hans-Martin von Gaudecker.** 2020. "Die Bedeutung individuellen Verhaltens über den Jahreswechsel für die Weiterentwicklung der Covid-19-Pandemie in Deutschland." Working paper. Institute of Labor Economics (IZA). [51]
- Gavenciak, Tomáš, Joshua Teperowski Monrad, Gavin Leech, Mrinank Sharma, Sören Mindermann, Jan Marcus Brauner, Samir Bhatt, and Jan Kulveit.** 2021. "Seasonal variation in SARS-CoV-2 transmission in temperate climates." *medRxiv*, [15, 42, 56, 57]
- Gaythorpe, K, N Imai, G Cuomo-Dannenburg, M Baguelin, S Bhatia, A Boonyasiri, A Cori, Z Cucunuba Perez, A Dighe, I Dorigatti, R Fitzjohn, H Fu, W Green, J Griffin, A Hamlet, W Hinsley,**

- N Hong, M Kwun, D Laydon, G Nedjati Gilani, L Okell, S Riley, H Thompson, S Van Elsland, R Verity, E Volz, P Walker, H Wang, Y Wang, C Walters, C Whittaker, P Winskill, X Xi, C Donnelly, A Ghani, and N Ferguson.** 2020. “Report 8: Symptom progression of COVID-19.” [19]
- Goldstein, Edward, Marc Lipsitch, and Muge Cevik.** 2020. “On the Effect of Age on the Transmission of SARS-CoV-2 in Households, Schools, and the Community.” *Journal of Infectious Diseases* 223 (3): 362–69. [8]
- Google, LLC.** 2021. “Google COVID-19 Community Mobility Reports.” Working paper. [27, 41, 60]
- Hale, Thomas, Tilbe Atav, Laura Hallas, Beatriz Kira, Toby Phillips, Anna Petherick, and Annalena Pott.** 2020. “Variation in US states responses to COVID-19.” *Blavatnik School of Government*, [11]
- Harris, Charles R., K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant.** 2020. “Array programming with NumPy.” *Nature* 585 (7825): 357–62. [46]
- Harris, Ross, Jennifer Hall, Asad Zaidi, Nick Andrews, J Kevin Dunbar, and Gavin Dabrera.** 2021. “Impact of vaccination on household transmission of SARS-COV-2 in England.” Working paper. Public Health England. [19]
- He, Xi, Eric HY Lau, Peng Wu, Xilong Deng, Jian Wang, Xinxin Hao, Yiu Chung Lau, Jessica Y Wong, Yujuan Guan, Xinghua Tan, et al.** 2020. “Temporal dynamics in viral shedding and transmissibility of COVID-19.” *Nature medicine* 26 (5): 672–75. [18]
- Hinch, Robert, William J M Probert, Anel Nurtay, Michelle Kendall, Chris Wymatt, Matthew Hall, Katrina Lythgoe, Ana Bulas Cruz, Lele Zhao, Andrea Stewart, Luca Ferritti, Daniel Montero, James Warren, Nicole Mather, Matthew Abueg, Neo Wu, Anthony Finkelstein, David G Bonsall, Lucie Abeler-Dorner, and Christophe Fraser.** 21, 2021. *OpenABM-Covid19 - Parameter Dictionary*. URL: https://github.com/BDI-pathogens/OpenABM-Covid19/blob/14e24ce298ad523eba5e6f33ee9c70a90c527b3e/documentation/parameters/parameter_dictionary.md. [18, 19]
- Hinch, Robert, William J. M. Probert, Anel Nurtay, Michelle Kendall, Chris Wymant, Matthew Hall, Katrina Lythgoe, Ana Bulas Cruz, Lele Zhao, Andrea Stewart, Luca Ferretti, Daniel Montero, James Warren, Nicole Mather, Matthew Abueg, Neo Wu, Olivier Legat, Katie Bentley, Thomas Mead, Kelvin Van-Vuuren, Dylan Feldner-Busztin, Tommaso Ristori, Anthony Finkelstein, David G. Bonsall, Lucie Abeler-Dörner, and Christophe Fraser.** 2021. “OpenABM-Covid19—An agent-based model for non-pharmaceutical interventions against COVID-19 including contact tracing.” *PLOS Computational Biology* 17 (7): 1–26. [8, 46]
- Hoang, Thang, Pietro Coletti, Alessia Melegaro, Jacco Wallinga, Carlos G Grijalva, John W Edmunds, Philippe Beutels, and Niel Hens.** 2019. “A systematic review of social contact surveys to inform transmission models of close-contact infections.” *Epidemiology* 30 (5): 723. [8]

- Hunter, John D.** 2007. "Matplotlib: A 2D graphics environment." *Computing in science & engineering* 9 (03): 90–95. [46]
- Hunter, Paul R., and Julii Brainard.** 2021. "Estimating the effectiveness of the Pfizer COVID-19 BNT162b2 vaccine after a single dose. A reanalysis of a study of real-world vaccination outcomes from Israel." *medRxiv*, (2): [10, 19]
- impfdashboard.de.** 15, 2021. *Download der Datensätze*. URL: <https://impfdashboard.de/daten>. [19]
- Isphording, Ingo E., Marc Lipfert, and Nico Pestel.** 2021. "Does re-opening schools contribute to the spread of SARS-CoV-2? Evidence from staggered summer breaks in Germany." *Journal of Public Economics* 198 (6): 104426. [51]
- Jang, Sukbin, Ji-Young Rhee, Yu Mi Wi, and Bo Kyeong Jung.** 2021. "Viral kinetics of SARS-CoV-2 over the preclinical, clinical, and postclinical period." *International Journal of Infectious Diseases* 102 (1): 561–65. [67]
- Jones, Terry C., Guido Biele, Barbara Mühlemann, Talitha Veith, Julia Schneider, Jörn Beheim-Schwarzbach, Tobias Bleicker, Julia Tesch, Marie Luisa Schmidt, Leif Erik Sander, Florian Kurth, Peter Menzel, Rolf Schwarzer, Marta Zuchowski, Jörg Hofmann, Andi Krumbholz, Angela Stein, Anke Edelmann, Victor Max Corman, and Christian Drosten.** 2021. "Estimating infectiousness throughout SARS-CoV-2 infection course." *Science*, [10]
- Kretzschmar, Mirjam E, Ganna Rozhnova, Martin C J Bootsma, Michiel van Boven, Janneke H H M van de Wijgert, and Marc J M Bonten.** 2020. "Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study." *Lancet Public Health* 5 (8): e452–e459. [15]
- Kronfeld-Schor, N., T. J. Stevenson, S. Nickbakhsh, E. S. Schernhammer, X. C. Dopico, T. Dayan, M. Martinez, and B. Helm.** 2021. "Drivers of Infectious Disease Seasonality: Potential Implications for COVID-19." *Journal of Biological Rhythms* 36 (1): 35–54. [56]
- Kühn, Martin J., Daniel Abele, Tanmay Mitra, Wadim Koslow, Majid Abedi, Kathrin Rack, Martin Siggel, Sahamoddin Khailaie, Margrit Klitz, Sebastian Binder, Luca Spataro, Jonas Gilg, Jan Kleinert, Matthias Häberle, Lena Plötzke, Christoph D. Spinner, Melanie Stecher, Xiao Xiang Zhu, Achim Basermann, and Michael Meyer-Hermann.** 2020. "Assessment of effective mitigation and prediction of the spread of SARS-CoV-2 in Germany using demographic information and spatial resolution." (12): [56, 57]
- Lam, Siu Kwan, Antoine Pitrou, and Stanley Seibert.** 2015. "Numba: A llvm-based python jit compiler." In *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC*, 1–6. [46]
- Landesregierung von Baden-Württemberg.** 25, 2021a. *Coronavirus: Aktuelle Informationen für Schulen und Kindertageseinrichtungen*. URL: <https://km-bw.de/Coronavirus> (visited on 03/25/2021). Accessed 2021-03-25. [26]
- Landesregierung von Baden-Württemberg.** 15, 2021b. *Präsenzbetrieb und inzidenzunabhängige Testpflicht ab dem 19. April*. URL: <https://km-bw.de/,Lde/startseite/service/2021-04-15-schulbetrieb-ab-dem-19april> (visited on 04/15/2021). [26]

- Landesregierung von Baden-Württemberg.** 14, 2021c. *Pressemitteilung: Landesweite Schließung von Schulen und Kindergärten.* URL: <https://www.baden-wuerttemberg.de/de/service/presse/pressemitteilung/pid/landesweite-schliessung-von-schulen-und-kindergaerten/> (visited on 03/13/2021). [27]
- Landesregierung von Baden-Württemberg.** 14, 2021d. *Pressemitteilung: Schnelltests in Kitas und Kindergärten.* URL: <https://www.baden-wuerttemberg.de/de/service/presse/pressemitteilung/pid/schnelltests-in-kitas-und-kindergaerten/> (visited on 04/08/2021). [27]
- Landesregierung von Baden-Württemberg.** No date. *Auswirkungen der Bundesnotbremse auf den Schul- und Kitabetrieb in Baden-Württemberg.* URL: <https://km-bw.de/Lde/startseite/service/2021-04-23+Auswirkungen+der+Bundesnotbremse+auf+den+Schul-+und+Kitabetrieb+in+Baden-Wuerttemberg.> [26, 27]
- Lessler, Justin, M. Kate Grabowski, Kyra H. Grantz, Elena Badillo-Goicoechea, C. Jessica E. Metcalf, Carly Lupton-Smith, Andrew S. Azman, and Elizabeth A. Stuart.** 2021. "Household COVID-19 risk and in-person schooling." *Science* 372 (6546): 1092–97. [8]
- Levine-Tiefenbrun, Matan, Idan Yelin, Rachel Katz, Esmá Herzal, Ziv Golan, Licita Schreiber, Tamar Wolf, Varda Nadler, Amir Ben-Tov, Jacob Kuint, Sivan Gazit, Tal Patalon, Gabriel Chodick, and Roy Kishony.** 2021. "Initial report of decreased SARS-CoV-2 viral load after inoculation with the BNT162b2 vaccine." *Nature medicine* 27 (5): 790–92. [10, 19]
- Lindner, Andreas K., Olga Nikolai, Franka Kausch, Mia Wintel, Franziska Hommes, Maximilian Gertler, Lisa J. Krüger, Mary Gaeddert, Frank Tobian, Federica Lainati, Lisa Köppel, Joachim Seybold, Victor M. Corman, Christian Drosten, Jörg Hofmann, Jilian A. Sacks, Frank P. Mockenhaupt, and Claudia M. Denkinger.** 2020. "Head-to-head comparison of SARS-CoV-2 antigen-detecting rapid test with self-collected nasal swab versus professional-collected nasopharyngeal swab." *European Respiratory Journal* 57 (4): 2003961. [56]
- Lipsitch, Marc, and Rebecca Kahn.** 2021. "Interpreting vaccine efficacy trial results for infection and transmission." *Vaccine* 39 (30): 4082–88. [19]
- Luijten, Michiel A. J., Maud M. van Muilekom, Lorynn Teela, Tinca J. C. Polderman, Caroline B. Terwee, Josjan Zijlmans, Leonie Klaufus, Arne Popma, Kim J. Oostrom, Hedy A. van Oers, and Lotte Haverman.** 2021. "The impact of lockdown during the COVID-19 pandemic on mental and social health of children and adolescents." *Quality of Life Research*, (5): [16]
- Mathieu, Edouard, Hannah Ritchie, Esteban Ortiz-Ospina, Max Roser, Joe Hasell, Cameron Appel, Charlie Giattino, and Lucas Rodés-Guirao.** 2021. "A global database of COVID-19 vaccinations." *Nature Human Behaviour*, [7]
- McAloon, Conor, Áine Collins, Kevin Hunt, Ann Barber, Andrew W Byrne, Francis Butler, Miriam Casey, John Griffin, Elizabeth Lane, David McEvoy, Patrick Wall, Martin Green, Luke O'Grady, and Simon J More.** 2020. "Incubation period of COVID-19: a rapid systematic review and meta-analysis of observational research." *BMJ Open* 10 (8): e039652. [18]

- McFadden, Daniel.** 1989. "A method of simulated moments for estimation of discrete response models without numerical integration." *Econometrica: Journal of the Econometric Society*, 995–1026. [11, 34, 47, 49]
- McKinney, Wes.** 2010. "Data Structures for Statistical Computing in Python." In *Proceedings of the 9th Python in Science Conference*. Edited by Stéfan van der Walt and Jarrod Millman, 56–61. [46]
- Melegari, M.G., M. Giallonardo, R. Sacco, L. Marcucci, S. Orecchio, and O. Bruni.** 2021. "Identifying the impact of the confinement of Covid-19 on emotional-mood and behavioural dimensions in children and adolescents with attention deficit hyperactivity disorder (ADHD)." *Psychiatry Research* 296 (2): 113692. [16]
- Mellacher, Patrick.** 2020. "COVID-Town: An Integrated Economic-Epidemiological Agent-Based Model." MPRA Paper 103661. University Library of Munich, Germany. [47]
- Mina, Michael J., and Kristian G. Andersen.** 2021. "COVID-19 testing: One size does not fit all." *Science* 371 (6525): 126–27. [8]
- Ministerium für Kinder, Familie, Flüchtlinge und Integration des Landes Nordrhein-Westfalen.** No date. "Verordnung zum Schutz vor Neuinfizierungen mit dem Coronavirus SARS-CoV-2im Bereich der Betreuungsinfrastruktur (Coronabetreuungsverordnung – CoronaBetrVO)Vom 7. Januar 2021: In der ab dem 8. März 2021 gültigen Fassung." [27]
- Ministerium für Kultus, Jugend und Sport Baden Württemberg.** 2021. "Erweiterte Teststrategie - Selbsttests an Schulen." *Pressemitteilungen*, (4): [28, 29]
- Ministerium für Kultus, Jugend und Sport Baden-Württemberg.** 2, 2021. *Informationen zum Schulbetrieb nach den Osterferien*. Pressemitteilung. URL: <https://km-bw.de/,Lde/9112691> (visited on 06/24/2021). [29]
- Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen.** 23, 2021a. „Notbremse“ im Infektionsschutzgesetz des Bundes. URL: <https://schulministerium.nrw/presse/pressemitteilungen/notbremse-im-infektionsschutzgesetz-des-bundes-23-04-2021> (visited on 04/23/2021). [26]
- Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen.** 8, 2021b. [08.04.2021] *Informationen zum Schulbetrieb in NRW*. URL: <https://schulministerium.nrw/08042021-informationen-zum-schulbetrieb-nrw> (visited on 04/08/2021). [26]
- Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen.** 22, 2021c. [22.04.2021] *Informationen zum Schulbetrieb ab 26. April 2021*. URL: <https://www.schulministerium.nrw/22042021-informationen-zum-schulbetrieb-ab-26-april-2021> (visited on 04/22/2021). [26]
- Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen.** 14, 2021d. *Pressemitteilung: Ministerin Gebauer: Bleiben auf dem Weg der Vorsicht: Wechselunterricht mit klaren Testvorgaben ab Montag*. URL: <https://schulministerium.nrw/presse/pressemitteilungen/ministerin-gebauer-bleiben-auf-dem-weg-der-vorsicht-14-04-2021> (visited on 04/14/2021). [26]

- Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen.** 14, 2021e. *Schulbetrieb im Wechselunterricht ab Montag, 19.April 2021 // Coronaselbsttests an Schulen - Testpflicht.* URL: <https://www.schulministerium.nrw/14042021-schulbetrieb-im-wechselunterricht-ab-montag-19-april-2021-coronaselbsttests-schulen> (visited on 06/27/2021). [29]
- Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen.** 15, 2021f. *Umfassende Informationen für die Schulen zu Corona-Selbsttests für Schülerinnen und Schüler.* Pressemitteilung. URL: <https://www.land.nrw/de/pressemitteilung/umfassende-informationen-fuer-die-schulen-zu-corona-selbsttests-fuer-schuelerinnen> (visited on 06/27/2021). [28, 29]
- Morawska, Lidia, Julian W. Tang, William Bahnfleth, Philomena M. Bluysen, Atze Boerstra, Giorgio Buonanno, Junji Cao, Stephanie Dancer, Andres Floto, Francesco Franchimon, Charles Haworth, Jaap Hogeling, Christina Isaxon, Jose L. Jimenez, Jarek Kurnitski, Yuguo Li, Marcel Loomans, Guy Marks, Linsey C. Marr, Livio Mazzarella, Arsen Krikor Melikov, Shelly Miller, Donald K. Milton, William Nazaroff, Peter V. Nielsen, Catherine Noakes, Jordan Peccia, Xavier Querol, Chandra Sekhar, Olli Seppänen, Shin-ichi Tanabe, Raymond Tellier, Kwok Wai Tham, Pawel Wargocki, Aneta Wierzbicka, and Maosheng Yao.** 2020. "How can airborne transmission of COVID-19 indoors be minimised?" *Environment International* 142 (9): 105832. [52]
- Mossong, Joël, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael Mikolajczyk, Marco Mas-sari, Stefania Salmaso, Gianpaolo Scalia Tomba, Jacco Wallinga, et al.** 2008. "Social contacts and mixing patterns relevant to the spread of infectious diseases." *PLoS medicine* 5 (3): [8, 20, 22–24, 41, 47–50]
- Niedersächsisches Kultusministerium.** 18, 2021. *Selbsttests an Schulen laufen an - Auslieferung der ersten 400.000 Tests beginnt heute, Testwoche startet Montag.* Pressemitteilung. URL: <https://www.mk.niedersachsen.de/startseite/aktuelles/presseinformationen/selbsttests-an-schulen-laufen-an-erste-400-000-tests-ab-heute-in-der-auslieferung-testwoche-startet-montag-198555.html> (visited on 06/24/2021). [29]
- OECD.** 2013. *Bildung auf einen Blick 2013: OECD-Indikatoren*, 530. [20]
- Ong, Sean Wei Xiang, Calvin J Chiew, Li Wei Ang, Tze-Minn Mak, Lin Cui, Matthias Paul H S Toh, Yi Ding Lim, Pei Hua Lee, Tau Hong Lee, Po Ying Chia, Sebastian Maurer-Stroh, Raymond T P Lin, Yee-Sin Leo, Vernon J Lee, David Chien Lye, and Barnaby Edward Young.** 2021. "Clinical and Virological Features of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Variants of Concern: A Retrospective Cohort Study Comparing B.1.1.7 (Alpha), B.1.351 (Beta), and B.1.617.2 (Delta)." *Clinical Infectious Diseases*, (08): ciab721. [67]
- Özcürümez, Mustafa, Antonios Katsounas, Stefan Holdenrieder, Alexander von Meyer, Harald Renz, and Roman Wölfel.** 2021. "Assessment of SARS-CoV-2 rapid antigen tests." *Journal of Laboratory Medicine* 45 (3): 143–48. [10]
- Paul, Christian, Anna Steinberg, Christoph Schaller, Kathy Haemmerl, Leonie Thum, Marcel Parciak, Michael Knaf, Sarina Meyer, Tina Parciak, and Urs Grupp.** 2020. *Dunkelzifferradar.* Edited by Prototype Fund. URL: <https://prototypefund.de/project/dunkelzifferradar/>. [31, 32, 45, 48, 55]

- Pavelka, Martin, Kevin Van-Zandvoort, Sam Abbott, Katharine Sherratt, Marek Majdan, Pavol Jarčuška, Marek Krajčí, Stefan Flasche, and Sebastian Funk.** 2021. “The impact of population-wide rapid antigen testing on SARS-CoV-2 prevalence in Slovakia.” *Science* 372 (6542): 635–41. [65]
- Peak, Corey M, Rebecca Kahn, Yonatan H Grad, Lauren M Childs, Ruoran Li, Marc Lipsitch, and Caroline O Buckee.** 2020. “Individual quarantine versus active monitoring of contacts for the mitigation of COVID-19: a modelling study.” *Lancet Infectious Diseases* 20 (9): 1025–33. [18]
- Petter, Ella, Orna Mor, Neta Zuckerman, Danit Oz-Levi, Asaf Younger, Dvir Aran, and Yaniv Erlich.** 2021. “Initial real world evidence for lower viral load of individuals who have been vaccinated by BNT162b2.” *medRxiv*, (2): [10, 19]
- Presse- und Informationsamt der Bundesregierung.** 24, 2021. *Erste Zulassungen für Selbsttests*. URL: <https://www.bundesregierung.de/breg-de/themen/coronavirus/zulassung-schnell-test-1861354> (visited on 06/24/2021). [29]
- Pritchard, Emma, Philippa C. Matthews, Nicole Stoesser, David W. Eyre, Owen Gethings, Karina-Doris Vihta, Joel Jones, Thomas House, Harper VanSteenHouse, Iain Bell, John I. Bell, John N. Newton, Jeremy Farrar, Ian Diamond, Emma Rourke, Ruth Studley, Derrick Crook, Tim E. A. Peto, A. Sarah Walker, and Koen B. Pouwels.** 2021. “Impact of vaccination on new SARS-CoV-2 infections in the United Kingdom.” *Nature Medicine*, [10, 19]
- Raabe, Tobias.** 2020. “A Python tool for managing scientific workflows.” [46]
- Robert Koch Institutue.** 7, 2021. *COVID-19-Fälle nach Meldewoche und Geschlecht sowie Anteile mit für COVID-19 relevanten Symptomen, Anteile Hospitalisierter/Verstorbener und Altersmittelwert/-median*. URL: https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Daten/Klinische_Aspekte.xlsx. [32]
- Robert Koch Institute.** 6, 2020. *Testkriterien für die SARS-CoV-2 Diagnostik: Anpassungen für die Herbst- und Wintersaison 2020/2021*. URL: https://web.archive.org/web/20201103122248/https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Teststrategie/Testkriterien_Herbst_Winter.html. [31]
- Robert Koch-Institut.** 2020. *Laborbasierte Surveillance SARS-CoV-2*. URL: <https://ars.rki.de/Content/COVID19/Main.aspx>. [32, 45, 55]
- Robert Koch-Institut.** 2021a. “Corona-Monitoring bundesweit (RKI-SOEP-Studie).” de. Working paper. Robert Koch-Institut. [17]
- Robert Koch-Institut.** 29, 2021b. “COVID-19 Impfquoten-Monitoring in Deutschland (COVIMO): Report 5 - Fokuserhebung Impfquoten.” Working paper. [21]
- Rocklin, Matthew.** 2015. “Dask: Parallel computation with blocked algorithms and task scheduling.” In *Proceedings of the 14th python in science conference*. Vol. 130, Citeseer, 136. [46]
- Rosella, Laura C., Ajay Agrawal, Joshua Gans, Avi Goldfarb, Sonia Sennik, and Janice Stein.** 2022. “Large-scale implementation of rapid antigen testing system for COVID-19 in workplaces.” *Science Advances* 8 (8): eabm3608. [67]

- Sächsisches Staatsministerium für Kultus.** 16, 2021. *Selbsttests für Schüler und Lehrer ab Mittwoch*. URL: <https://www.medienservice.sachsen.de/medien/news/248791> (visited on 06/24/2021). [29]
- Scheiblaue, Heinrich, Angela Filomena, Andreas Nitsche, Andreas Puyskens, Víctor M Corman, Christian Drosten, Karin Zwirgmaier, Constanze Lange, Petra Emmerich, Michael Müller, Olivia Knauer, and C Micha Nübling.** 2021. "Comparative sensitivity evaluation for 122 CE-marked rapid diagnostic tests for SARS-CoV-2 antigen, Germany, September 2020 to April 2021." *Eurosurveillance* 26(44): 2100441. [10, 66, 67]
- Shapley, Lloyd S.** 2016. *A value for n-person games*. Princeton University Press. [13, 39]
- Silva, Petrônio C.L., Paulo V.C. Batista, Hélder S. Lima, Marcos A. Alves, Frederico G. Guimarães, and Rodrigo C.P. Silva.** 2020. "COVID-ABS: An agent-based model of COVID-19 epidemic to simulate health and economic effects of social distancing interventions." *Chaos, Solitons & Fractals* 139(10): 110088. [46]
- Singanayagam, Anika, Monika Patel, Andre Charlett, Jamie Lopez Bernal, Vanessa Saliba, Joanna Ellis, Shamez Ladhani, Maria Zamboni, and Robin Gopal.** 2020. "Duration of infectiousness and correlation with RT-PCR cycle threshold values in cases of COVID-19, England, January to May 2020." *Eurosurveillance* 25(32): [18]
- Smith, Rebecca L, Laura L Gibson, Pamela P Martinez, Ruian Ke, Agha Mirza, Madison Conte, Nicholas Gallagher, Abigail Conte, Leyi Wang, Richard Fredrickson, Darci C Edmonson, Melinda E Baughman, Karen K Chiu, Hannah Choi, Tor W Jensen, Kevin R Scardina, Shannon Bradley, Stacy L Gloss, Crystal Reinhart, Jagadeesh Yedetore, Alyssa N Owens, John Broach, Bruce Barton, Peter Lazar, Darcy Henness, Todd Young, Alastair Dunnett, Matthew L Robinson, Heba H Mostafa, Andrew Pekosz, Yukari C Manabe, William J Heetderks, David D McManus, and Christopher B Brooke.** 2021. "Longitudinal assessment of diagnostic test performance over the course of acute SARS-CoV-2 infection." *Journal of Infectious Diseases*, (6): jjab337. [10, 56]
- Stevens, Jean-Luc R, Philipp Rudiger, and J Bednar.** 2015. "HoloViews: Building complex visualizations easily for reproducible science." In *Proceedings of the 14th Python in Science Conference*. Citeseer, 61–69. [46]
- Stillman, Steven, and Mirco Tonin.** 2022. "Communities and Testing for COVID-19." *European Journal of Health Economics*, 1–9. [17]
- Stokes, Erin K., Laura D. Zambrano, Kayla N. Anderson, Ellyn P. Marder, Kala M. Raz, Suad El Burai Felix, Yunfeng Tie, and Kathleen E. Fullerton.** 2020. "Coronavirus Disease 2019 Case Surveillance — United States, January 22–May 30, 2020." *MMWR. Morbidity and Mortality Weekly Report* 69(24): 759–65. [18]
- The pandas development team.** 2020. *pandas-dev/pandas: Pandas*. Version latest. [46]
- Thompson, Mark G, Jefferey L Burgess, Allison L Naleway, Harmony L Tyner, Sarang K Yoon, Jennifer Meece, Lauren EW Olsho, Alberto J Caban-Martinez, Ashley Fowlkes, Karen Lutrick, et al.** 2021. "Interim estimates of vaccine effectiveness of BNT162b2 and mRNA-1273 COVID-19 vaccines in preventing SARS-CoV-2 infection among health care personnel, first responders,

and other essential and frontline workers—eight US locations, December 2020–March 2021.” *Morbidity and Mortality Weekly Report* 70(13): 495. [19]

Van Rossum, Guido, and Fred L Drake Jr. 1995. *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam. [46]

Virtanen, Pauli, Ralf Gommers, Travis E Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, et al. 2020. “SciPy 1.0: fundamental algorithms for scientific computing in Python.” *Nature methods* 17(3): 261–72. [46]

Vlachos, Jonas, Edvin Hertegård, and Helena B. Svaleryd. 2021. “The effects of school closures on SARS-CoV-2 among parents and teachers.” *Proceedings of the National Academy of Sciences* 118(9): [16]

Vygen-Bonnet, Sabine, Judith Koch, Christian Bogdan, Thomas Harder, Ulrich Heininger, Kerstin Kling, Martina Littmann, Joerg Meerpohl, Heidi Meyer, Thomas Mertens, Nora Schmid-Küpke, Stefan Scholz, Martin Terhardt, Marina Treskova-Schwarzbach, Klaus Überla, Marianne van der Sande, Ole Wichmann, Sabine Wicker, Ursula Wiedermann, Verina Wild, and Rüdiger von Kries. 2020. “Beschluss und Wissenschaftliche Begründung der Ständigen Impfkommision (STIKO) für die COVID-19-Impfempfehlung.” (2): 3–63. [21]

Waskom, Michael L. 2021. “Seaborn: statistical data visualization.” *Journal of Open Source Software* 6(60): 3021. [46]

Wollmershäuser, Timo. 2021. “Konjunkturelle Auswirkungen der zweiten Coronawelle auf ausgewählte Wirtschaftsbereiche.” *ifo Schnelldienst Digital* 2(5): [17]

Yin, Guosheng, and Huaqing Jin. 2020. “Comparison of Transmissibility of Coronavirus Between Symptomatic and Asymptomatic Patients: Reanalysis of the Ningbo COVID-19 Data.” *JMIR Public Health and Surveillance* 6(2): e19464. [18]

Zhao, Shi, Biao Tang, Salihu S Musa, Shujuan Ma, Jiayue Zhang, Minyan Zeng, Qingping Yun, Wei Guo, Yixiang Zheng, Zuyao Yang, Zhihang Peng, Marc KC Chong, Mohammad Javanbakht, Daihai He, and Maggie H. Wang. 2021. “Estimating the generation interval and inferring the latent period of COVID-19 from the contact tracing data.” *Epidemics* 36: 100482. [18]

Zou, Lirong, Feng Ruan, Mingxing Huang, Lijun Liang, Huitao Huang, Zhongsi Hong, Jianxiang Yu, Min Kang, Yingchao Song, Jinyu Xia, Qianfang Guo, Tie Song, Jianfeng He, Hui-Ling Yen, Malik Peiris, and Jie Wu. 2020. “SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients.” *New England Journal of Medicine* 382(12): 1177–79. [18]

Zuin, Marco, Valentina Gentili, Carlo Cervellati, Roberta Rizzo, and Giovanni Zuliani. 2021. “Viral Load Difference between Symptomatic and Asymptomatic COVID-19 Patients: Systematic Review and Meta-Analysis.” *Infectious Disease Reports* 13(3): 645–53. [67]

Chapter 2

Structural Models for Policy-Making: Coping with Parametric Uncertainty

Joint with Philipp Eisenhauer and Lena Janys

2.1 Introduction

Structural microeconometricians use highly parameterized computational models to investigate economic mechanisms, predict the impact of proposed policies, and inform optimal policy-making (Wolpin, 2013). These models represent deep structural relationships of theoretical economic models invariant to policy changes (Hood and Koopmans, 1953). The sources of uncertainty in such an analysis are ubiquitous (Saltelli, Bammer, Bruno, Charters, Di Fiore, et al., 2020). For example, models are often misspecified, there are numerical approximation errors in their implementation, and model parameters are uncertain. Therefore, most disciplines require a proper account of uncertainty before using computational models to

* Philipp Eisenhauer and Lena Janys are both funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2126/1- 390838866 and the TRA Modelling (University of Bonn) as part of the Excellence Strategy of the federal and state governments. Janos Gabler is grateful for financial support by the German Research Foundation (DFG) through CRC-TR 224 (Project C01) and funding by IZA Institute of Labor Economics. Lena Janys is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2047/1 - 390685813. Philipp Eisenhauer was funded by a post-doctoral fellowship by the AXA Research Fund. We thank Tim Mensinger for his help in the early stages of the project. We thank Max Blesch, Joachim Freyberger, Annica Gehlen, Daniel Harenberg, Ken Judd, Gregor Reich, Jörg Stoye, and Rafael Suchy for numerous helpful discussions. We thank Michael Keane and Kenneth Wolpin for providing the dataset used in our analysis. We thank Annica Gehlen and Emily Schwab for their outstanding research assistance. We are grateful to the Social Sciences Computing Service (SSCS) at the University of Chicago for the permission to use their computational resources

inform decision-making (National Research Council, 2012; Renn, Baghrarian, and Capaccioli, 2019).

The following study focuses on parametric uncertainty in structural microeconomic models that are estimated on observed data. Researchers often do not account for parametric uncertainty and conduct an as-if analysis in which the point estimates serve as a stand-in for the true model parameters. They then continue to study the implications of their models at the point estimates (Eisenhauer, Heckman, and Mosso, 2015; Blundell, Costa Dias, Meghir, and Shaw, 2016; Adda, Dustmann, and Stevens, 2017; Eckstein, Keane, and Lifshitz, 2019) and rank competing policy proposals based on the point predictions alone (Todd and Wolpin, 2006; Cunha, Heckman, and Schennach, 2010; Blundell and Shephard, 2012; Gayle and Shephard, 2019). In fact, Keane, Todd, and Wolpin (2011) states in their handbook article that they are unaware of any applied work that reports the distribution of policy predictions under parametric uncertainty. To the best of our knowledge, this statement remains true more than a decade later. Consequently, economists risk accepting fragile findings as facts, ignoring the trade-off between model complexity and prediction uncertainty, and neglecting to frame policy advice as a decision problem under uncertainty.

To mitigate these shortcomings, we develop an approach that copes with parametric uncertainty in structural microeconomic models and embeds model-informed policy-making in a decision-theoretic framework. Ideally, policy-makers fix the parameter space *ex-ante* and then evaluate the policy options according to decision rules. However, this approach is often computationally intractable. We, therefore, follow Manski (2021)'s suggestion and, instead of using the parameter estimates as-if they were true, incorporate uncertainty in the analysis by treating the estimated confidence set as-if it is correct. We use the confidence set to construct an uncertainty set that is anchored in empirical estimates, statistically meaningful, and computationally tractable (Ben-Tal, Hertog, De Waegenare, Melenberg, and Rennen, 2013). Instead of just focusing on the point estimates, we evaluate counterfactual policies based on all parametrizations within the uncertainty set.

We draw on statistical decision theory (Manski, 2013) to deal with the uncertainty in counterfactual predictions. This approach promotes a well-reasoned and transparent policy process. Before a decision, it clarifies trade-offs between choices (Gilboa, Rouziou, and Sibony, 2018). Afterward, decision-theoretic principles allow constituents to scrutinize the coherence of choices (Gilboa and Samuelson, 2021), ease the *ex-post* justification (Berger, Berger, Bosetti, Gilboa, Hansen, et al., 2021), and facilitate the communication of uncertainty (Manski, 2019).

We tailor our approach to the class of Eckstein-Keane-Wolpin (EKW) models (Aguirregabiria and Mira, 2010). Labor economists often use EKW models to learn about human capital investment and consumption-saving decisions and predict the impact of proposed reforms to education policy and welfare programs (Keane, Todd, and Wolpin, 2011; Blundell, 2017; Low and Meghir, 2017). The analysis of these models poses serious computational challenges. During estimation, EKW models are solved thousands of times and even a single solution often takes several minutes. Thus, a decision-theoretic ex-ante analysis of alternative decision rules across the whole parameter space, as intended by (Wald, 1950), is infeasible. Instead we construct an uncertainty set, a subset of the whole parameter space, and deal with the ex-post uncertainty after estimating the model. This compromise allows us to garner the benefits of using statistical decision theory to shape policy-making under uncertainty while ensuring the computational tractability of our analysis.

As an example of our approach, we analyze the seminal human capital investment model by Keane and Wolpin (1997) as a well-known, empirically grounded, and computationally demanding test case. We follow the authors and estimate the model on the National Longitudinal Survey of Youth 1979 (NLSY79) (Bureau of Labor Statistics, 2019) using the original dataset and reproduce all core results. We revisit their predictions for the impact of a tuition subsidy on completed years of schooling. The economics of the model implies that the nonlinear mapping between the model parameters and predictions is truncated at zero, and we thus use the Confidence Set (CS) bootstrap (Woutersen and Ham, 2019) to estimate the confidence set for the counterfactuals. We document considerable uncertainty in the policy predictions and highlight the resulting policy recommendations from different formal rules on decision-making under uncertainty.

Our work extends existing research exploring the sensitivity of implications and predictions to parametric uncertainty in macroeconomics and climate economics. For example, Harenberg, Marelli, Sudret, and Winschel (2019) study uncertainty propagation and sensitivity analysis for a standard real business cycle model. Cai and Lontzek (2019) examine how uncertainties and risks in economic and climate systems affect the social cost of carbon. However, neither of them estimates their model on data. Instead, they rely on expert judgments to inform the degree of parametric uncertainty. They do not investigate the consequences of uncertainty for policy decisions in a decision-theoretic framework.

We complement a burgeoning literature on the sensitivity analysis of policy predictions in light of model or moment misspecification. For example, Andrews, Gentzkow, and Shapiro (2017) and Andrews, Gentzkow, and Shapiro (2020) treat the model specification as given and then analyze the sensitivity of the parameter estimates to the misspecification of the moments used for estimation. Christensen

and Connault (2019) study global sensitivity of the model predictions to misspecification of the distribution of unobservables. Jørgensen (2021) provides a local measure for the sensitivity of counterfactuals to model parameters that are fixed before the estimation of the model.¹ This literature does not embed the counterfactual predictions in a decision-theoretic setting. Recent work by Kalouptsidei, Kitamura, Lima, and Souza-Rodrigues (2021), Kalouptsidei, Scott, and Souza-Rodrigues (2021), and Norets and Tang (2014) studies (partial) identification and inference on counterfactuals. However, they all adopt the setup outlined in Rust (1987) and exploit the additive separability of the immediate utility function between observed and unobserved state variables, which does not apply to EKW models. In related work, Blesch and Eisenhauer (2021) conduct a decision-theoretic ex-ante analysis to determine optimal decision rules in Rust (1987)'s stochastic dynamic investment model where the decision-maker directly accounts for uncertainty in the model's transition dynamics. They only consider uncertainty in a subset of the model's parameters which are estimated outside the model and remain fixed to their point estimates during the analysis.

In Section 2.2, we describe the decision-theoretic framework for making model-informed decisions under parametric uncertainty using an illustrative example. After summarizing the empirical setting of Keane and Wolpin (1997) in Section 2.3, we present our results in Section 2.4. We complete our analysis in Section 2.5 with a brief conclusion and outlook.

2.2 Structural Models for Policy-Making

In the following section, we discuss uncertainty propagation and the common practice of using estimated parameters as a plug-in replacement for the true model parameters. We then explore the limitations of this strategy and introduce our alternative approach, in which we implement estimated confidence sets to construct uncertainty sets. In so doing, we are able to cope with uncertain policy predictions in a proper decision-theoretic framework.

At a high level, a structural microeconomic model provides a mapping $\mathcal{M}(\theta)$ between the l model parameters $\theta \in \Theta$ and a quantity y that is of interest to policy-makers.

$$\mathbb{R}^l \supset \Theta \ni \theta \mapsto \mathcal{M}(\theta) = y$$

1. For other examples, see Armstrong and Kolesár (2021), Bonhomme and Weidner (2021), Bugni and Ura (2019), and Mukhin (2018).

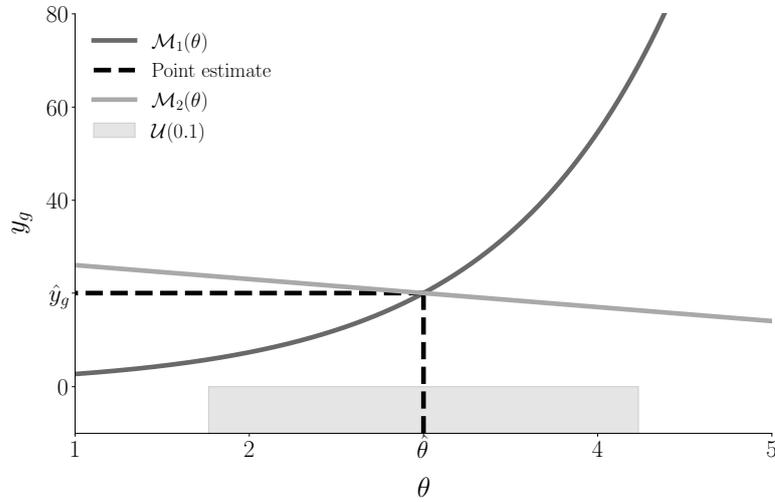
A policy $g \in \mathcal{G}$ changes the mapping to $\mathcal{M}_g(\theta)$ and produces a counterfactual y_g .

Estimation of a baseline model $\mathcal{M}(\theta)$ describing the status-quo on observed data allows researchers to learn about the true parameters. Frequentist estimation procedures such as maximum likelihood estimation and the method of simulated moments produce a point estimate $\hat{\theta}$. However, uncertainty about the true parameters remains.

Previewing our empirical analysis of Keane and Wolpin (1997), our \mathcal{M} is provided by a dynamic model of human capital accumulation, which we estimate on observed schooling and labor market decisions using simulated maximum likelihood estimation. The policy g is the implementation of a college tuition subsidy, and the counterfactual is the level of completed schooling in the population. Example parameters that drive the economics of the model are time preferences of individuals, the return to schooling, and the transferability of work experience across occupations.

The following illustrative example highlights our key points. We consider two policies $g \in \{1, 2\}$ that result in two different mappings ($\mathcal{M}_1, \mathcal{M}_2$) of the same scalar θ to a counterfactual y_g . Higher values of y_g are more desirable for a policy-maker. The point estimate $\hat{\theta}$ is determined by estimating a baseline model on an observed dataset. We denote the probability density function of its sampling distribution by $f_{\hat{\theta}}$.

Under the first policy, the counterfactual is an increasing nonlinear function of θ . In the case of the second policy, the relationship is decreasing and linear.



Notes: We parameterize the two models as $y_1 = \exp \theta$ and $y_2 = 29.08 - 3 \theta$.

Figure 2.2.1. Model comparison

Figure 2.2.1 traces the counterfactual from both models over a range of the parameter. At the point estimate, both models yield the same value for the counterfactual. Once we account for uncertainty in our estimates of the true parameter, deciding which policy to adopt becomes less straightforward: for higher values of θ , the first policy is preferred, while the opposite is true for lower values.

2.2.1 Uncertainty Sets

Manski (2021) suggests acknowledging parametric uncertainty by working with estimated confidence sets instead of point estimates. A confidence set $\Theta(\alpha) \subset \Theta$ covers the true parameters, from an ex-ante point of view, with a predetermined coverage probability of $(1 - \alpha)$. Proceeding with our analysis, we refine the status quo procedure, in which estimated parameter values serve as a stand-in for the model's true parametrization. Instead, we assume the estimated confidence set for the parameters $\hat{\Theta}(\alpha)$ and the counterfactual $\hat{\Theta}_{y_g}(\alpha)$ are correct and analyze policy decisions accordingly.

Based on the estimated confidence sets, we construct so-called uncertainty sets for the parameters $\mathcal{U}(\alpha)$ and the prediction $\mathcal{U}_{y_g}(\alpha)$ by only considering parameterizations that we cannot reject based on a hypothesis test with confidence level $1 - \alpha$. This approach ensures the tractability of our decision-theoretic analysis, as the uncertainty set of the parameters is much smaller than the whole parameter space of

the model. We adopt this procedure from the literature on data-driven robust optimization in operations research (Ben-Tal et al., 2013; Bertsimas, Gupta, and Kallus, 2018).

2.2.2 Statistical Decision Theory

In our setting, a policy-maker relies on a structural model with an uncertain parametrization to map alternative policies to counterfactual predictions. In most cases, the preferred policy depends on the model's uncertain true parameters. We, therefore, draw on statistical decision theory to organize the decision-making process (Gilboa, 2009; Marinacci, 2015).

Returning to our example, we rank the two policies according to alternative statistical decision rules using an uncertainty set derived from a confidence set with a 90% coverage probability. In what follows, we postulate a simple linear utility function $U(y_g)$ to describe the policy-maker's preferences.²

Figure 2.2.2 shows the implied sampling distribution of the predictions for the two alternative policies and the corresponding uncertainty sets $\mathcal{U}_{y_g}(0.1)$. The mapping \mathcal{M}_1 is highly nonlinear, while the mapping \mathcal{M}_2 is linear. When evaluated at the point estimate, the counterfactual is the same under both policies, so a policy-maker is indifferent. However, the spread of the uncertainty set differs considerably.

2. We assume that the sampling distribution of the point estimate is normal with a mean of three and a standard deviation of three-fourths. We can derive the uncertainty sets directly and simply consider realizations of $\theta \in [1.76, 4.23]$.

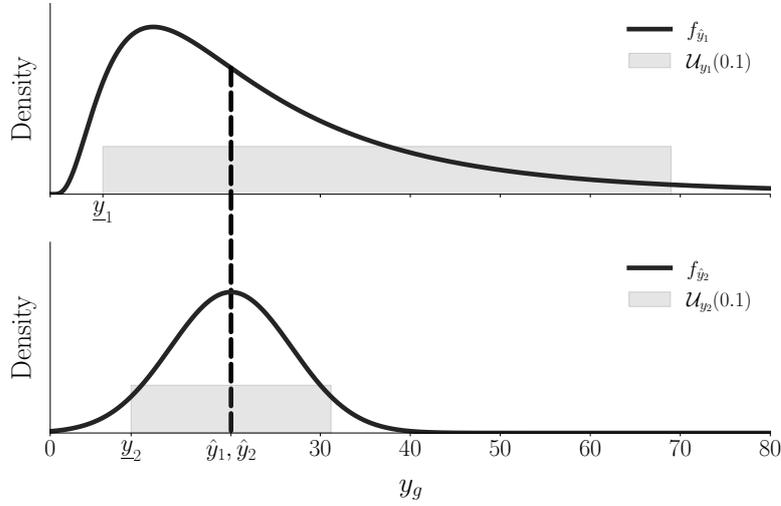


Figure 2.2.2. Comparing policy predictions

Decision theory proposes a variety of different rules for reasonable decisions in this setting. We explore the following four: (1) as-if optimization, (2) maximin criterion, (3) minimax regret rule, and (4) subjective Bayes.

As-if optimization describes the predominant practice. The estimation of the model produces point estimates that serve as a plug-in for the true parameters. The decision maximizes the utility at the point estimate. More formally,

$$g^* = \arg \max_{g \in \mathcal{G}} U(M_g(\hat{\theta})).$$

Given our example, an as-if policy-maker is indifferent between the two policies, since both policies result in the same counterfactual at the point estimates as indicated by the dashed line in Figure 2.2.2.

The maximin criterion and minimax regret rule are two common alternatives that favor actions that work uniformly well over all possible parameters in the uncertainty set. This approach departs from as-if optimization, which only considers a policy’s performance at a single point in the uncertainty set. The maximin decision (Wald, 1950; Gilboa and Schmeidler, 1989) is determined by computing the minimum utility for each policy within the uncertainty set and choosing the one with the highest worst-case outcome. Stated concisely,

$$g^* = \arg \max_{g \in \mathcal{G}} \min_{\theta \in \mathcal{U}(\alpha)} U(M_g(\theta)).$$

Returning to Figure 2.2.2, a maximin policy-maker prefers g_2 as the worst-case outcome. Within the uncertainty set, y_2 is better than under the alternative policy,

g_1 .

The minimax regret rule (Niehans, 1948; Manski, 2004) computes the maximum regret for each policy over the whole uncertainty set and chooses the policy that minimizes the maximum regret. The regret of choosing a policy g for a given parameterization of the model is the difference between the maximum possible utility achieved from adopting $\tilde{g} \in \mathcal{G}$ and the actual utility obtained. The decision maximizes:

$$g^* = \arg \min_{g \in \mathcal{G}} \max_{\theta \in \mathcal{U}(\alpha)} \underbrace{\left[\max_{\tilde{g} \in \mathcal{G}} U(M_{\tilde{g}}(\theta)) - U(M_g(\theta)) \right]}_{\text{regret}}.$$

Figure 2.2.3 compares our two policy examples over the uncertainty sets. A policy-maker adopting policy g_1 regrets his choice for small values of the model parameter, while the opposite is true for larger values. The regret of each policy is maximized at the boundaries of the uncertainty set. Maximum regret is minimized when a policy-maker chooses g_1 . It corresponds to the difference in the counterfactual at the lower boundary of the uncertainty set instead of the larger difference at its upper bound. This outcome contradicts the maximin decision in which policy g_2 is preferred.

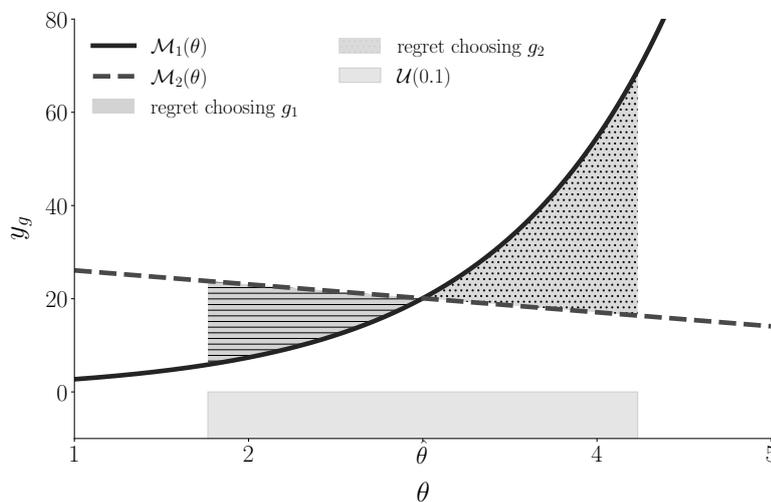


Figure 2.2.3. Comparing policy regret

Each decision rule presented so far focuses on a single point in the uncertainty set as the policy’s relevant performance measure. Bayesian approaches aggregate a policy’s performance over the complete uncertainty set.

Maximization of the subjective expected utility (Savage, 1954) requires the policy-maker to place a subjective probability distribution f_{θ} over the parameters in the uncertainty set. A policy-maker then selects the alternative with the highest expected subjective utility. Formally,

$$g^* = \arg \max_{g \in \mathcal{G}} \int_{\mathcal{U}(\alpha)} U(M_g(\theta)) df_{\theta}.$$

Applying a uniform distribution to our example, a policy-maker chooses g_1 , which performs well for high values of θ and still reasonably well for low values.

2.3 Eckstein-Keane-Wolpin Models

We now present the general structure of Eckstein-Keane-Wolpin (EKW) models (Aguirregabiria and Mira, 2010) and their solution approach. We then turn to the customized version used by Keane and Wolpin (1997) to study the career decisions of young men and investigate the consequences of parametric uncertainty in this empirically-grounded and computationally demanding setting. We outline their model's basic setup, provide some descriptive statistics of the empirical data used in our estimation, and then discuss the core findings.

2.3.1 General Structure

EKW models describe sequential decision-making under uncertainty (Gilboa, 2009; Machina and Viscusi, 2014). At time $t = 1, \dots, T$ each individual observes the state of their choice environment $s_t \in S$ and chooses an action a_t from the set of admissible actions \mathcal{A} . The decision has two consequences: an individual receives an immediate utility $u_t(s_t, a_t)$ and their environment evolves to a new state s_{t+1} . The transition from s_t to s_{t+1} is affected by the action but remains uncertain. Since individuals are forward-looking, they do not simply choose the alternative with the highest immediate utility. Instead, they take the future consequences of their actions into account.

A policy $\pi = (d_1^{\pi}, \dots, d_T^{\pi})$ provides the individual with instructions for choosing an action in any possible future state. It is a sequence of decision rules d_t^{π} that specify the action $d_t^{\pi}(s_t) \in \mathcal{A}$ at a particular time t for any possible state s_t under π . The implementation of a policy generates a sequence of utilities that depends on the objective transition probability distribution $p_t(s_t, a_t)$ for the evolution from state s_t to s_{t+1} induced by the model.

Figure 2.3.1 depicts the timing of events for two generic periods. At the beginning of period t , an individual fully learns about each action's immediate utility, selects one of the alternatives, and receives its immediate utility. Then, the state evolves

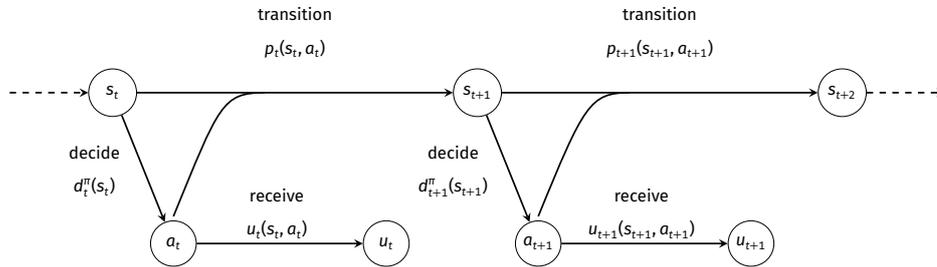


Figure 2.3.1. Timing of events

from s_t to s_{t+1} , and the process repeats itself in $t + 1$.

Individuals make their decisions facing uncertainty about the future and seek to maximize their expected total discounted utilities over all decision periods given all available information. They have rational expectations (Muth, 1961), so their subjective beliefs about the future agree with the objective probabilities for all possible future events provided by the model. Immediate utilities are separable between periods (Kahneman, Wakker, and Sarin, 1997), and a discount factor δ parameterizes a preference for immediate over future utilities (Samuelson, 1937).

Equation (2.3.1) formally describes the individual's objective. Given an initial state s_1 , they implement a policy π that maximizes the expected total discounted utilities over all decision periods given the information available at the time.

$$\max_{\pi \in \Pi} \mathbb{E}_{s_1}^{\pi} \left[\sum_{t=1}^T \delta^{t-1} u_t(s_t, d_t^{\pi}(s_t)) \right] \quad (2.3.1)$$

EKW models are set up as a standard Markov decision process (MDP) (White, 1993; Puterman, 1994; Rust, 1994) that can be solved by a simple backward induction procedure. In the final period T , there is no future to consider, and the optimal action is choosing the alternative with the highest immediate utility in each state. With the decision rule for the final period, we can determine all other optimal decisions recursively. We use our group's open-source research code `respy` (Gabler and Raabe, 2020), which allows for the flexible specification, simulation, and estimation of EKW models. Detailed documentation of the software and its numerical components is available at <http://respy.readthedocs.io>.

2.3.2 The Career Decisions of Young Men

Keane and Wolpin (1997) specialize the model above to explore the career decisions of young men regarding their schooling, work, and occupational choices using the National Longitudinal Survey of Youth 1979 (NLSY79) (Bureau of Labor Statistics,

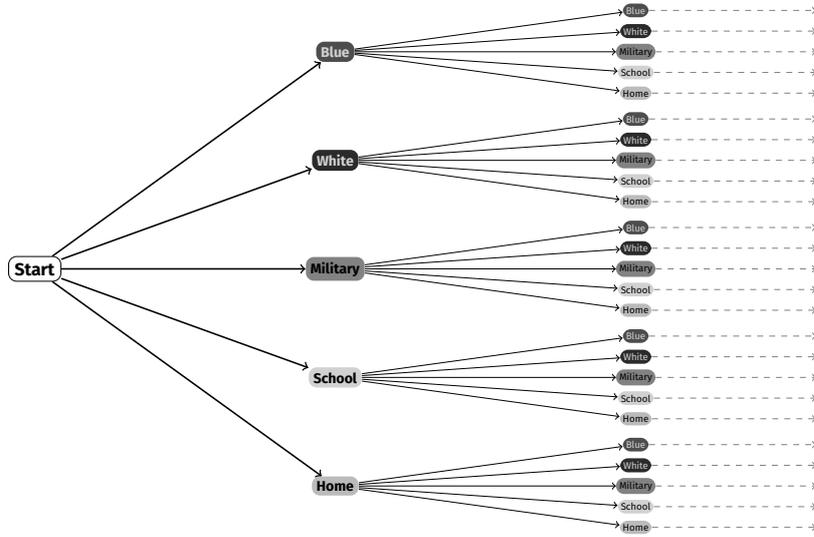


Figure 2.3.2. Decision tree

2019) for the estimation of the model. We restrict ourselves to a basic summary of their setup. Further documentation of the model specification and the observed dataset is available in the Appendix.

Keane and Wolpin (1997) follows individuals over their working life from young adulthood at age 16 to retirement at age 65. Each decision period $t = 16, \dots, 65$ represents a school year. Figure 2.3.2 illustrates the initial decision problem as individuals select one of five alternatives from the set of admissible actions $a \in \mathcal{A}$. They can decide to either work in a blue-collar or a white-collar occupation ($a = 1, 2$), serve in the military ($a = 3$), attend school ($a = 4$), or stay at home ($a = 5$).

Individuals are already heterogeneous when entering the model. They differ with respect to their level of initial schooling h_{16} , and have one of four different $\mathcal{J} = \{1, \dots, 4\}$ alternative-specific skill endowment types $e = (e_{j,a})_{\mathcal{J} \times \mathcal{A}}$.

The immediate utility $u_a(\cdot)$ of each alternative consists of a non-pecuniary utility $\zeta_a(\cdot)$ and, at least for the working alternatives, an additional wage component $w_a(\cdot)$. Both depend on the level of human capital as measured by their alternative-specific skill endowment e , their years of completed schooling h_t , and their occupation-specific work experience $\mathbf{k}_t = (k_{a,t})_{a \in \{1,2,3\}}$. The immediate utilities are influenced by last-period choices a_{t-1} and alternative-specific productivity shocks $\epsilon_t = (\epsilon_{a,t})_{a \in \mathcal{A}}$ as well. Their general form is given by:

$$u_a(\cdot) = \begin{cases} \zeta_a(\mathbf{k}_t, h_t, t, a_{t-1}) + w_a(\mathbf{k}_t, h_t, t, a_{t-1}, e_{j,a}, \epsilon_{a,t}) & \text{if } a \in \{1, 2, 3\} \\ \zeta_a(\mathbf{k}_t, h_t, t, a_{t-1}, e_{j,a}, \epsilon_{a,t}) & \text{if } a \in \{4, 5\}. \end{cases}$$

Work experience \mathbf{k}_t and years of completed schooling h_t evolve deterministically. There is no uncertainty about grade completion (Altonji, 1993) and no part-time enrollment. Schooling is defined by time spent in school, not by formal credentials acquired. Once individuals reach a certain amount of schooling, they acquire a degree.

$$\begin{aligned} k_{a,t+1} &= k_{a,t} + \mathbf{I}[a_t = a] & \text{if } a \in \{1, 2, 3\} \\ h_{t+1} &= h_t + \mathbf{I}[a_t = 4] \end{aligned}$$

The productivity shocks ϵ_t are uncorrelated across time and follow a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix Σ . Given the structure of the utility functions and the distribution of the shocks, the state at time t is $s_t = \{\mathbf{k}_t, h_t, t, a_{t-1}, \mathbf{e}, \epsilon_t\}$.

Skill endowments \mathbf{e} and initial schooling h_{16} are the only sources of persistent heterogeneity in the model. All remaining differences in life-cycle decisions result from different transitory shocks ϵ_t that occur over time.

Theoretical and empirical research from specialized disciplines within economics informs the specification of each $u_a(\cdot)$. As an example, we provide the exact functional form of the non-pecuniary utility from schooling in Equation (2.3.2). Further details on the specification of the utility functions are available in the Appendix.

$$\begin{aligned} \zeta_4(s_t) &= \underbrace{e_{j,4}}_{\text{type}} + \underbrace{\beta_{tc_1} \cdot \mathbf{I}[h_t \geq 12] + \beta_{tc_2} \cdot \mathbf{I}[h_t \geq 16]}_{\text{tuition costs}} + \underbrace{\gamma_{4,4} \cdot t + \gamma_{4,5} \cdot \mathbf{I}[t < 18]}_{\text{time trend}} \\ &+ \underbrace{\beta_{rc_1} \cdot \mathbf{I}[a_{t-1} \neq 4, h_t < 12] + \beta_{rc_2} \cdot \mathbf{I}[a_{t-1} \neq 4, h_t \geq 12]}_{\text{re-enrollment cost}} + \dots + \epsilon_{4,t} \end{aligned} \quad (2.3.2)$$

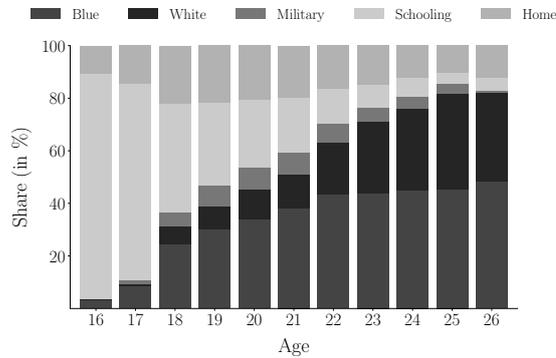
There is a direct cost in the form of tuition for continuing education after high school β_{tc_1} and college β_{tc_2} . The decision to leave school is reversible, but entails re-enrollment costs that differ by schooling category ($\beta_{rc_1}, \beta_{rc_2}$).

We analyze the original dataset used by Keane and Wolpin (1997). We only provide a brief description and relegate further details to the Appendix. The authors construct their sample based on the NLSY79, a nationally representative sample of young men and women living in the United States in 1979 and born between

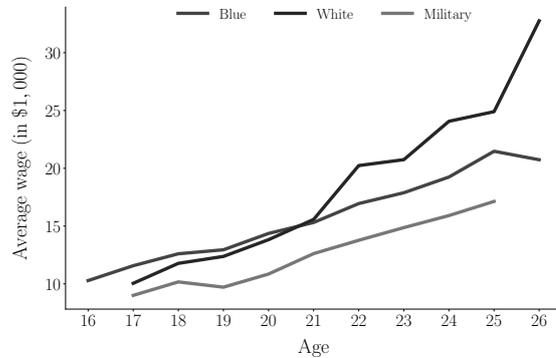
1957 and 1964. Individuals were followed from 1979 onwards and repeatedly interviewed about their schooling decisions and labor market experiences. Based on this information, individuals are assigned to either working in one of the three occupations, attending school, or simply staying at home.

Keane and Wolpin (1997) restrict attention to white men, who turned 16 between 1977 and 1981, and exploit information collected between 1979 and 1987. Thus, individuals in the sample range in age between 16 and 26 years old. While the sample initially consists of 1,373 individuals at age 16, this number drops to 256 at the age of 26 due to sample attrition and missing data. Overall, the final sample consists of 12,359 person-period observations.

Figure 2.3.3 summarizes the evolution of choices and wages over the sample period. Roughly 86% of individuals initially enroll in school, but this share steadily declines with age. Nevertheless, about 39% pursue some form of higher education and obtain more than a high school degree. As individuals leave school, most of them initially pursue a blue-collar occupation. However, the relative share of white-collar workers increases as individuals entering the labor market later gain access to higher levels of schooling. At age 26, about 48% work in a blue-collar occupation and 34% in a white-collar occupation. The share of individuals in the military peaks around age 20 at 8%. At its maximum around age 18, approximately 20% of individuals stay at home.



(a) Choices



(b) Average wage

Notes: The wage is a full-time equivalent deflated by the gross national product deflator, with 1987 as the base year. We do not report the wage if less than ten observations are available.

Figure 2.3.3. Data overview

For an individual, the average wage starts at about \$10,000 at age 16 and increases considerably up to about \$25,000 by the age of 26. While starting wages for blue-collar workers are about \$10,286, wages in white-collar occupations and the military start around \$9,000. However, wages for white-collar occupations increase sharply over time, overtaking blue-collar wages around age 21. By the end of the observation period, wages for white-collar occupations are about 50% higher than blue-collar wages at \$32,756 compared to only \$20,739. Military wages remain lowest throughout.

We consider observations for $i = 1, \dots, N$ individuals in each time period $t = 1, \dots, T_i$. For every observation (i, t) in the data, we observe the action a_{it} , some components \bar{u}_{it} of the utility, and a subset \bar{s}_{it} of the state s_{it} . Therefore, from an economist's point of view, we must distinguish between two types of state variables $s_{it} = \{\bar{s}_{it}, e, \epsilon_t\}$. At time t , the economist and individual both observe \bar{s}_{it} , while $\{e, \epsilon_t\}$ is only observed

by the individual.

We use simulated maximum likelihood (Fisher, 1922; Manski and Lerman, 1977) estimation and determine the 88 model parameters $\hat{\theta}$ that maximize the likelihood function $\mathcal{L}(\theta | \mathcal{D})$. As we only observe a subset $\bar{s}_t = \{\mathbf{k}_t, h_t, t, a_{t-1}\}$ of the state, we can determine the probability $p_{it}(a_{it}, \bar{u}_{it} | \bar{s}_{it}, \theta)$ of individual i at time t in \bar{s}_{it} choosing a_{it} and receiving \bar{u}_{it} given parametric assumptions about the distribution of ϵ_t . The objective function takes the following form:

$$\hat{\theta} \equiv \arg \max_{\theta \in \Theta} \underbrace{\prod_{i=1}^N \prod_{t=1}^{T_i} p_{it}(a_{it}, \bar{u}_{it} | \bar{s}_{it}, \theta)}_{\mathcal{L}(\theta | \mathcal{D})}.$$

Overall, our parameter estimates are in broad agreement with the results reported in the original paper and the related literature. For example, individuals discount future utilities by 6% per year. The returns to schooling vary according to occupation. While wages for white-collar occupations increase by about 6% with each additional year of schooling, they only increase by 2% for those working blue collar jobs. Skills are transferable across occupations as work experience increases wages in both blue and white-collar occupations.

Figure 2.3.4 shows the overall agreement between the empirical data and a dataset simulated using the estimated model parameters. We show average wages and the share of individuals choosing a blue-collar occupation over time. The results are based on a simulated sample of 10,000 individuals. Additional model fit statistics are available in the Appendix.

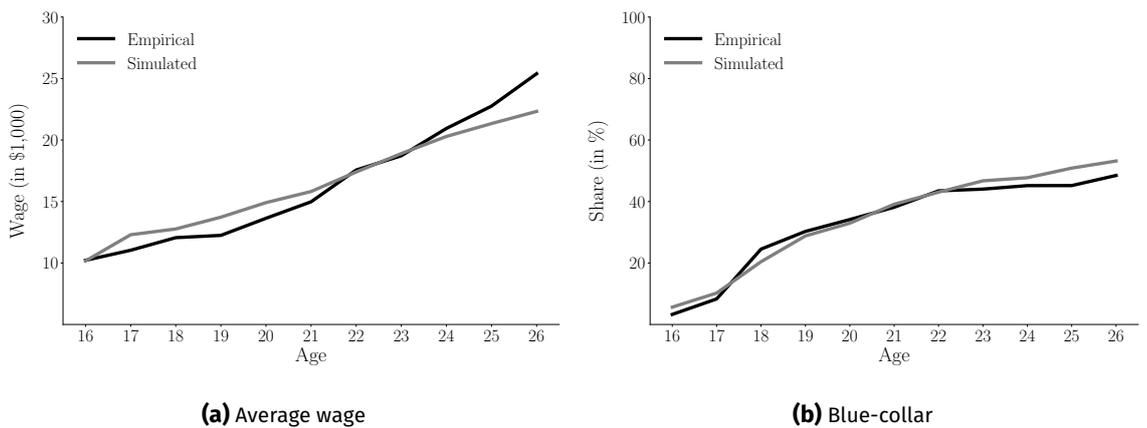


Figure 2.3.4. Model fit

We adhere to the procedure outlined by the authors of the original paper and use the estimated model to conduct the ex-ante evaluation of a \$2,000 tuition subsidy

on educational attainment. We simulate a sample of 10,000 individuals using the point estimates and compare completed schooling to a sample of the same size, but with a reduction of $\hat{\beta}_{tc_1}$ by \$2,000. The subsidy increases average final schooling by 0.65 years. College graduation increases by 13 percentage points, and high school graduation rates improve by 4 percentage points.

2.3.3 Confidence Set Bootstrap

The construction of confidence sets for counterfactuals in many structural models poses two distinct challenges. First, the computational burden of even a single estimation of the model is considerable. This makes the application of a standard bootstrap approach (Efron, 1979) infeasible. Second, the nonlinear mapping from the parameters of the model to the counterfactual predictions often has kinks or is truncated. For example, in our case, the predicted impact of a tuition subsidy is bounded from below by zero. This violates the smoothness requirements of the delta method.

We use the Confidence Set (CS) bootstrap to construct the confidence set of the counterfactual. Although the CS bootstrap was originally proposed in Rao (1973), it has only recently been formalized by Woutersen and Ham (2019). Its application does not require repeated estimations of the model, as it uses the asymptotic normal distribution of the estimator for $\hat{\theta}$. Furthermore, its validity does not depend on the differentiability of the prediction function.³

Algorithm 1 provides a concise description of the steps involved, where $\chi_l^2(1-\alpha)$ is the quantile function for probability $1-\alpha$ of the chi-square distribution with l degrees of freedom.

3. See Reich and Judd (2020) for a critical assessment of confidence sets based on asymptotic arguments. They advocate the use of likelihood-ratio confidence intervals instead and set up their computation as a constraint optimization problem.

Algorithm 1. Confidence Set bootstrap

```

for  $m = 1, \dots, M$  do
  Draw  $\hat{\theta}_m \sim \mathcal{N}(\hat{\theta}, \hat{\Sigma})$ 
  if  $(\hat{\theta}_m - \hat{\theta})' \hat{\Sigma}^{-1} (\hat{\theta}_m - \hat{\theta}) \leq \chi_l^2(1 - \alpha)$  then
    Compute  $\hat{y}_{g,m} = \mathcal{M}_g(\hat{\theta}_m)$ 
    Add  $\hat{y}_{g,m}$  to sample  $Y = \{\hat{y}_{g,1}, \dots, \hat{y}_{g,m-1}\}$ 
  end if
end for
Set  $\theta_{y_g}(\alpha) = [\min(Y), \max(Y)]$ 

```

To summarize, we draw a large sample of M parameters from the estimated asymptotic normal distribution of our estimator with mean $\hat{\theta}$ and covariance matrix $\hat{\Sigma}$, accepting only those draws that are elements of the confidence set of the model parameters. We then compute the counterfactual for all remaining draws and calculate the confidence set for the counterfactual based on its lowest and highest value.

The CS bootstrap poses a considerable computational challenge. In many applications, including our own, a single prediction of a counterfactual takes several minutes. At the same time, the number of parameter samples must be large to ensure that the minimum and maximum values for the counterfactual prediction are reliable. However, the algorithm is amenable to parallelization using modern high-performance computational resources by processing each of the M parameter draws independently.

Our uncertainty sets then take the following form:

$$\begin{aligned} \mathcal{U}(\alpha) &\equiv \{\theta \in \Theta : (\theta - \hat{\theta})' \hat{\Sigma}^{-1} (\theta - \hat{\theta}) \leq \chi_l^2(1 - \alpha)\} \\ \mathcal{U}_{y_g}(\alpha) &\equiv \{M_g(\theta) : (\theta - \hat{\theta})' \hat{\Sigma}^{-1} (\theta - \hat{\theta}) \leq \chi_l^2(1 - \alpha), \theta \in \Theta\}. \end{aligned}$$

2.4 Results

Turning to the presentation of our results, we focus on the impact of a \$2,000 tuition subsidy on completed schooling and use the 90% uncertainty set to measure the degree of uncertainty. All our results potentially depend on the size of the uncertainty set. In practice, policy-makers choose the uncertainty set's size in line with their underlying preferences - the more desirable protection against

unfavorable outcomes is, the larger the uncertainty set will be.⁴

All results are based on 30,000 draws from the asymptotic normal distribution of our parameter estimates. We follow Keane and Wolpin (1997) and start by analyzing the prediction for a general subsidy. Then we turn to the situation where we use endowment types for policy targeting. Throughout our analysis, we postulate a linear utility function for the policy-maker.

2.4.1 General Subsidy

Figure 2.4.1 explores the impact prediction for a general tuition subsidy. We show the point prediction, its sampling distribution, and the uncertainty set. At the point estimate, average schooling increases by 0.65 years. However, there is considerable uncertainty about the prediction, as the uncertainty set ranges from 0.15 to 1.10 years.

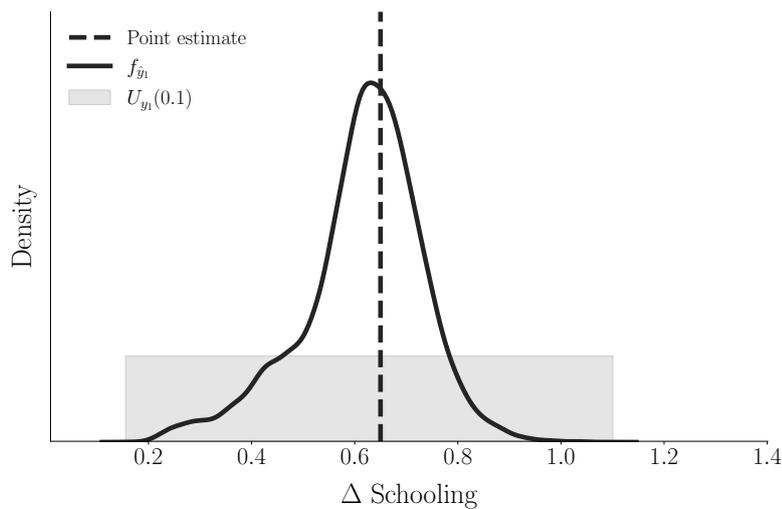


Figure 2.4.1. General subsidy

In Figure 2.4.2, we trace the effect of the discount rate δ on the subsidy's impact over the uncertainty set, while keeping all other parameters at their point estimate. Initially, as δ increases, so does the policy's impact as individuals value the long-term benefits from increasing their level of schooling more and more. However, for high levels of the discount factor, the policy's impact starts to decrease as most individuals already complete a high school or college degree even without the subsidy.

4. In a different setting, Blesch and Eisenhauer (2021) conduct an ex-ante performance evaluation of the statistical decision functions over the whole parameter space (Wald, 1950; Manski, 2021).

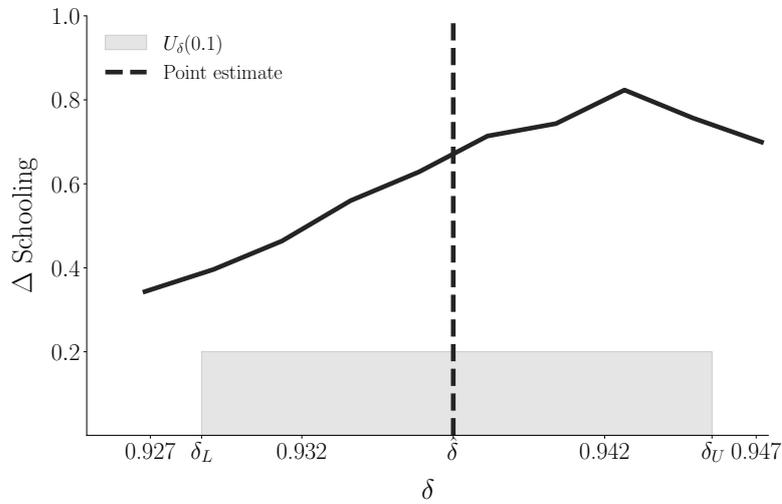


Figure 2.4.2. Time preference

2.4.2 Targeted Subsidy

So far, we restricted the analysis to a general subsidy available to the whole population and the average predicted impact. We now examine the setting in which a policy-maker can target individuals based on the type of their initial endowment. The importance of early endowment heterogeneity in shaping economic outcomes over the life-cycle is the most important finding from Keane and Wolpin (1997). It served as motivation for a host of subsequent research on the determinants of skill heterogeneity among adolescents (Todd and Wolpin, 2007; Erosa, Koreshkova, and Restuccia, 2010; Caucutt and Lochner, 2020).

To ease the exposition, we initially focus our discussion of results on Type 1 and Type 3 individuals. We later rank policies targeting either of the four types based on the different decision-theoretic criteria. Additional results are available in our Appendix.

Figure 2.4.3 confirms that life-cycle choices differ considerably by initial endowment type. On the left, we show the number of periods the two types spend on average in each of the five alternatives. Those characterized as Type 1 individuals spend more than six years on their education even after entering the model. Type 3 individuals, on the other hand, extend their academic pursuits for only an additional two years. This difference translates into very different labor market experiences. While Type 1 individuals work for about 35 years in a white-collar occupation, Type 3 workers switch more frequently between white and blue-collar occupations and spend a comparable amount of time working in either occupation – approximately 44 years

split equally among white and blue-collar occupations. Both types only spend a short time at home.

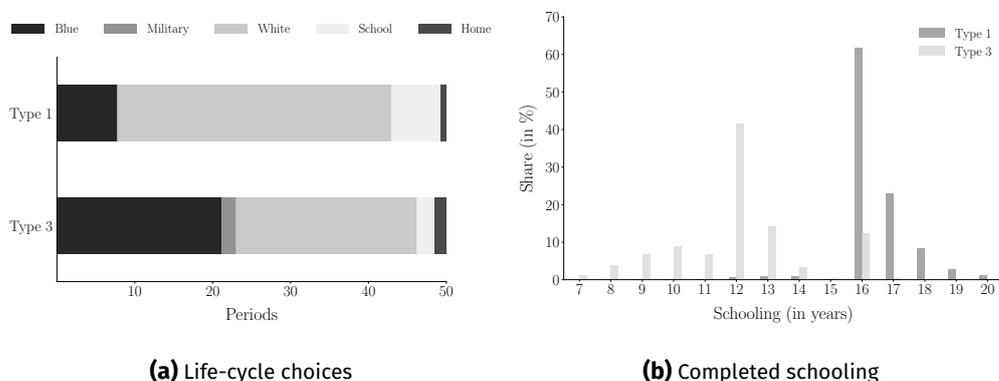


Figure 2.4.3. Type heterogeneity

On the right, we show the distribution of final schooling for both types. Years of schooling are considerably higher for Type 1 individuals with an average of more than 16 years compared to only 12 years for those identified as Type 3 individuals. Nearly all Type 1 individuals enroll in college and most graduate with a degree.

Figure 2.4.4 provides a visualization of our core results for a targeted subsidy. At the point estimates, the predicted impact is considerably lower for Type 1 than Type 3. However, the prediction uncertainty is much larger for Type 3 compared to Type 1. The uncertainty set for Type 3 ranges all the way from 0 to 1.2 years, while the prediction for Type 1 is between 0.18 and 0.75.

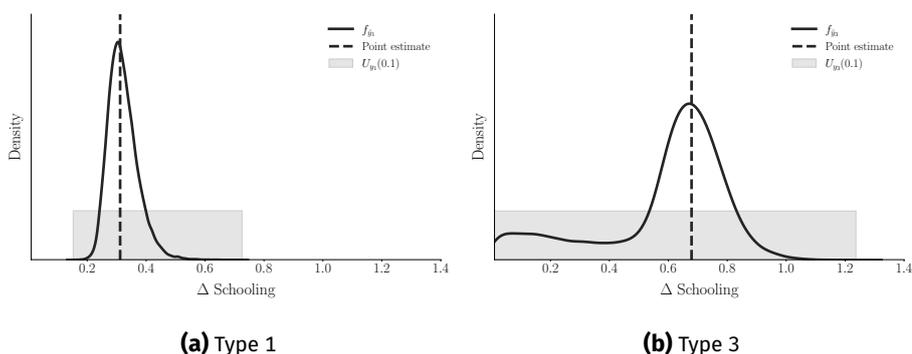


Figure 2.4.4. Targeted subsidy

This heterogeneity in impact and prediction uncertainty follows directly from the underlying economics of the model. Type 1 individuals are already more likely to have a college degree before the subsidy, and thus, the predicted impact is

smaller. Alternatively, Type 1 individuals affected by the subsidy are in the middle of pursuing a college education and thus directly benefit from it. Since Type 3 individuals are at the lower end of the schooling distribution, a tuition subsidy can considerably increase their level of schooling. Whether the subsidy succeeds in doing so, however, remains uncertain.

We now consider the policy option to target Type 2 and Type 4 as well. Their point predictions are actually highest with an additional 0.81 years on average for Type 2 and 0.75 years for Type 4. However, both predictions are fraught with uncertainty. For Type 2 the uncertainty set ranges from 0.17 to 1.3, while for Type 4 it starts at zero and spans all the way to 1.18.

Figure 2.4.5 shows the policy alternative’s ranking by the decision-theoretic criteria we discussed in Section 2.2.2. Ranking alternatives using as-if optimization is straightforward. A policy targeting Type 2 is the most preferred alternative, while a focus on Type 1 is the least attractive. However, once we account for the presence of uncertainty in the predictions, a more nuanced picture emerges. Moving from as-if optimization to a subjective Bayes criterion using a uniform distribution over the uncertainty set does not change the ordering. However, once a decision-maker is concerned with performance across the whole range of values in the uncertainty set – we move to the minimax regret or maximin criterion – a policy targeting Type 1 becomes more and more attractive despite its low point prediction because its worst-case utility is highest.

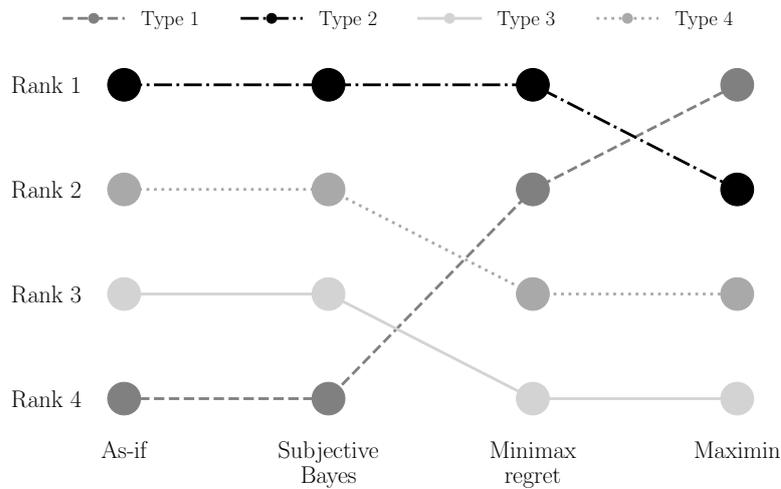


Figure 2.4.5. Policy ranking

In general, framing policy advice as a decision problem under uncertainty shows that there are many different ways of making reasonable decisions. The ranking of policies varies depending on the decision criteria. Not only that, but due to the necessary ex-post nature of our implementation, the ranking for a given criteria also depends on the choice of α . The selection of α is part of the decision problem: the more a policy-maker is concerned about worst-case scenarios, the smaller the appropriate value for α will be. After deciding on a preferred decision rule, we suggest performing a sensitivity analysis around the selected α value by checking how much the policy ranking varies within a neighborhood.

2.5 Conclusion

We develop a generic approach that addresses parametric uncertainty when using models to inform policy-making. We propose a decision-theoretic analysis of computationally demanding structural models based on uncertainty sets. We construct the uncertainty sets from empirical estimates and ensure their computational tractability by using the confidence set bootstrap. We revisit the seminal work by Keane and Wolpin (1997) to document the empirical relevance of prediction uncertainty and showcase our analysis. Focusing on their ex-ante evaluation of a tuition subsidy, we report considerable uncertainty in the policy's impact on completed schooling. We show how a policy-maker's preferred policy depends on the choice of alternative formal rules for decision-making under uncertainty.

In our ongoing research, we pursue three avenues for further improvements. First, we link our work with the literature on inference under (local) model misspecification to refine the construction of our uncertainty sets. For example, Armstrong and Kolesár (2021) and Bonhomme and Weidner (2021) propose different methods for taking misspecification into account when constructing confidence sets. Second, we incorporate ideas from the literature on global sensitivity analysis (Razavi, Jake-man, Saltelli, Prieur, Iooss, et al., 2021) to identify the parameters most responsible for uncertainty in predictions. The attribution of importance based on Shapely values, familiar to economists from game theory, appears promising (Shapley, 1953; Owen, 2014) as well. Third, we address our analysis's computational burden using surrogate modeling (Forrester, Sobester, and Keane, 2008), which emulates the full model's behavior at a negligible cost per run and allows us to determine prediction uncertainty using a nonparametric bootstrap procedure.

Appendix 2.A Supplementary Material

The Appendix contains details on our computational implementation, the estimation dataset, and additional results.

2.A.1 Computation

Using the same computational implementation as Keane and Wolpin (1997), we outline the immediate utility functions for each of the five alternatives. We first focus on their common structure and then present their parameterization. We also provide the economic motivation for their specification.

We follow individuals over their working life from age 16 until retirement at age 65. Each decision period $t = 16, \dots, 65$ represents a school year. Individuals can select one of five alternatives from the set of admissible actions $a \in \mathcal{A}$. They can decide to either work in a blue-collar or white-collar occupation ($a = 1, 2$), serve in the military ($a = 3$), attend school ($a = 4$), or stay at home ($a = 5$).

Individuals differ with respect to their initial level of completed schooling h_{16} , and they possess one of four $\mathcal{J} = \{1, \dots, 4\}$ alternative-specific skill endowments $e = (e_{j,a})_{\mathcal{J} \times \mathcal{A}}$.

The immediate utility $u_a(\cdot)$ of each alternative consists of a non-pecuniary utility $\zeta_a(\cdot)$ and, at least for the working alternatives, an additional wage component $w_a(\cdot)$, both of which depend on the level of human capital as measured by their occupation-specific work experience $\mathbf{k}_t = (k_{a,t})_{a \in \{1,2,3\}}$, years of completed schooling h_t , and alternative-specific skill endowment e . The immediate utility functions are influenced by last-period choices a_{t-1} and alternative-specific productivity shocks $\epsilon_t = (\epsilon_{a,t})_{a \in \mathcal{A}}$ as well. Their general form is given by:

$$u_a(\cdot) = \begin{cases} \zeta_a(\mathbf{k}_t, h_t, t, a_{t-1}) + w_a(\mathbf{k}_t, h_t, t, a_{t-1}, e_{j,a}, \epsilon_{a,t}) & \text{if } a \in \{1, 2, 3\} \\ \zeta_a(\mathbf{k}_t, h_t, t, a_{t-1}, e_{j,a}, \epsilon_{a,t}) & \text{if } a \in \{4, 5\}. \end{cases}$$

Work experience \mathbf{k}_t and years of completed schooling h_t evolve deterministically:

$$\begin{aligned} k_{a,t+1} &= k_{a,t} + \mathbf{I}[a_t = a] & \text{if } a \in \{1, 2, 3\} \\ h_{t+1} &= h_t + \mathbf{I}[a_t = 4]. \end{aligned}$$

The productivity shocks are uncorrelated across time and follow a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix Σ . Given the structure of the utility functions and the distribution of the shocks, the state at time t is

$$s_t = \{\mathbf{k}_t, h_t, t, a_{t-1}, \mathbf{e}, \boldsymbol{\epsilon}_t\}.$$

Empirical and theoretical research from specialized disciplines within economics informs the exact specification of $u_a(\cdot)$. We now discuss each of its components in detail.

Non-Pecuniary Utility

We begin by presenting the parameterization of the non-pecuniary utility for all five alternatives.

Blue-collar. Equation (2.A.1) shows the parameterization of the non-pecuniary utility from working in a blue-collar occupation:

$$\begin{aligned} \zeta_1(\mathbf{k}_t, h_t, a_{t-1}) = & \alpha_1 + c_{1,1} \cdot \mathbf{I}[a_{t-1} \neq 1] + c_{1,2} \cdot \mathbf{I}[k_{1,t} = 0] \\ & + \vartheta_1 \cdot \mathbf{I}[h_t \geq 12] + \vartheta_2 \cdot \mathbf{I}[h_t \geq 16] + \vartheta_3 \cdot \mathbf{I}[k_{3,t} = 1]. \end{aligned} \quad (2.A.1)$$

A constant α_1 captures the net monetary equivalent of on-the-job amenities. Non-pecuniary utility includes mobility and search costs $c_{1,1}$, which are higher for individuals who had previously never worked in a blue-collar occupation, $c_{1,2}$, and captures returns from high school, ϑ_1 , and college degrees, ϑ_2 . Additionally, there is a detrimental effect of prematurely leaving the military after one year, ϑ_3 .

White-collar. The non-pecuniary utility from working in a white-collar occupation is specified analogously. Equation (2.A.2) shows its parameterization:

$$\begin{aligned} \zeta_2(\mathbf{k}_t, h_t, a_{t-1}) = & \alpha_2 + c_{2,1} \cdot \mathbf{I}[a_{t-1} \neq 2] + c_{2,2} \cdot \mathbf{I}[k_{2,t} = 0] \\ & + \vartheta_1 \cdot \mathbf{I}[h_t \geq 12] + \vartheta_2 \cdot \mathbf{I}[h_t \geq 16] + \vartheta_3 \cdot \mathbf{I}[k_{3,t} = 1]. \end{aligned} \quad (2.A.2)$$

Military. Equation (2.A.3) shows the parameterization of the non-pecuniary utility from working in the military:

$$\zeta_3(k_{3,t}, h_t) = c_{3,2} \cdot \mathbf{I}[k_{3,t} = 0] + \vartheta_1 \cdot \mathbf{I}[h_t \geq 12] + \vartheta_2 \cdot \mathbf{I}[h_t \geq 16]. \quad (2.A.3)$$

Although search costs $c_{3,1} = 0$ are absent, there is a mobility cost if an individual has never previously served in the military, $c_{3,2}$. Individuals still experience a non-pecuniary utility from completing high school, ϑ_1 , and college, ϑ_2 .

School. Equation (2.A.4) shows the parameterization of the non-pecuniary utility from schooling:

$$\begin{aligned} \zeta_4(k_{3,t}, h_t, t, a_{t-1}, e_{j,4}, \epsilon_{4,t}) = & e_{j,4} + \beta_{tc_1} \cdot \mathbf{I}[h_t \geq 12] + \beta_{tc_2} \cdot \mathbf{I}[h_t \geq 16] \quad (2.A.4) \\ & + \beta_{rc_1} \cdot \mathbf{I}[a_{t-1} \neq 4, h_t < 12] \\ & + \beta_{rc_2} \cdot \mathbf{I}[a_{t-1} \neq 4, h_t \geq 12] + \gamma_{4,4} \cdot t \\ & + \gamma_{4,5} \cdot \mathbf{I}[t < 18] + \vartheta_1 \cdot \mathbf{I}[h_t \geq 12] \\ & + \vartheta_2 \cdot \mathbf{I}[h_t \geq 16] + \vartheta_3 \cdot \mathbf{I}[k_{3,t} = 1] + \epsilon_{4,t}. \end{aligned}$$

There are direct costs for pursuing higher education, which primarily take the form of college, β_{tc_1} , and graduate school tuition fees, β_{tc_2} . The decision to leave school is reversible, but entails adjustment costs that differ by schooling category ($\beta_{rc_1}, \beta_{rc_2}$). Education is defined by time spent in school, not by formal credentials acquired. Once individuals reach a certain amount of schooling, they acquire a degree. There is no uncertainty about grade completion (Altonji, 1993) and no part-time enrollment. Individuals value the completion of high school and college (ϑ_1, ϑ_2).

Home. Equation (2.A.5) shows the parameterization of the non-pecuniary utility from staying at home:

$$\begin{aligned} \zeta_5(k_{3,t}, h_t, t, e_{j,5}, \epsilon_{5,1}) = & e_{j,5} + \gamma_{5,4} \cdot \mathbf{I}[18 \leq t \leq 20] + \gamma_{5,5} \cdot \mathbf{I}[t \geq 21] \quad (2.A.5) \\ & + \vartheta_1 \cdot \mathbf{I}[h_t \geq 12] + \vartheta_2 \cdot \mathbf{I}[h_t \geq 16] \\ & + \vartheta_3 \cdot \mathbf{I}[k_{3,t} = 1] + \epsilon_{5,t}. \end{aligned}$$

Staying at home as a young adult, $\gamma_{5,4}$, is less stigmatized than doing so as an older individual, $\gamma_{5,5}$. Possessing a degree (ϑ_1, ϑ_2) or leaving the military prematurely, ϑ_3 , influences the immediate utility as well.

Wage Component

The wage component $w_a(\cdot)$ for the working alternatives is given by the product of the market-equilibrium rental price r_a and an occupation-specific skill level $x_a(\cdot)$. The latter is determined by the overall level of human capital:

$$w_a(\cdot) = r_a x_a(\cdot).$$

This specification leads to a standard logarithmic wage equation in which the constant term is the skill rental price $\ln(r_a)$ and wages follow a log-normal distribution.

The occupation-specific skill level $x_a(\cdot)$ is determined by a skill production function, which includes a deterministic component $\Gamma_a(\cdot)$ and a multiplicative stochastic productivity shock $\epsilon_{a,t}$:

$$x_a(\mathbf{k}_t, h_t, t, a_{t-1}, e_{j,a}, \epsilon_{a,t}) = \exp(\Gamma_a(\mathbf{k}_t, h_t, t, a_{t-1}, e_{j,a}) \cdot \epsilon_{a,t}).$$

Blue-collar. Equation (2.A.6) shows the parameterization of the deterministic component of the skill production function:

$$\begin{aligned} \Gamma_1(\mathbf{k}_t, h_t, t, a_{t-1}, e_{j,1}) = & e_{j,1} + \beta_{1,1} \cdot h_t + \beta_{1,2} \cdot \mathbf{I}[h_t \geq 12] & (2.A.6) \\ & + \beta_{1,3} \cdot \mathbf{I}[h_t \geq 16] + \gamma_{1,1} \cdot k_{1,t} + \gamma_{1,2} \cdot (k_{1,t})^2 \\ & + \gamma_{1,3} \cdot \mathbf{I}[k_{1,t} > 0] + \gamma_{1,4} \cdot t + \gamma_{1,5} \cdot \mathbf{I}[t < 18] \\ & + \gamma_{1,6} \cdot \mathbf{I}[a_{t-1} = 1] + \gamma_{1,7} \cdot k_{2,t} + \gamma_{1,8} \cdot k_{3,t}. \end{aligned}$$

There are several notable features. The first part of the skill production function is motivated by Mincer (1974) and, hence, linear in years of completed schooling, $\beta_{1,1}$, quadratic in experience ($\gamma_{1,1}, \gamma_{1,2}$), and separable between the two of them. There are so-called sheep-skin effects (Hungerford and Solon, 1987; Jaeger and Page, 1996) associated with completing a high school, $\beta_{1,2}$, or graduate education, $\beta_{1,3}$, which capture the impact of completing a degree beyond the associated years of schooling. There is also a first-year blue-collar experience effect $\gamma_{1,3}$. Additionally, job skills depreciate for blue-collar workers, who were unemployed in the previous period, $\gamma_{1,6}$. All forms of work experience ($\gamma_{1,7}, \gamma_{1,8}$) are transferable.

White-collar. The wage component from working in a white-collar occupation is specified analogously. Equation (2.A.7) shows the parameterization of the deterministic component of the skill production function:

$$\begin{aligned} \Gamma_2(\mathbf{k}_t, h_t, t, a_{t-1}, e_{j,2}) = & e_{j,2} + \beta_{2,1} \cdot h_t + \beta_{2,2} \cdot \mathbf{I}[h_t \geq 12] & (2.A.7) \\ & + \beta_{2,3} \cdot \mathbf{I}[h_t \geq 16] + \gamma_{2,1} \cdot k_{2,t} + \gamma_{2,2} \cdot (k_{2,t})^2 \\ & + \gamma_{2,3} \cdot \mathbf{I}[k_{2,t} > 0] + \gamma_{2,4} \cdot t + \gamma_{2,5} \cdot \mathbf{I}[t < 18] \\ & + \gamma_{2,6} \cdot \mathbf{I}[a_{t-1} = 2] + \gamma_{2,7} \cdot k_{1,t} + \gamma_{2,8} \cdot k_{3,t}. \end{aligned}$$

Military. Equation (2.A.8) shows the parameterization of the deterministic component of the skill production function:

$$\begin{aligned} \Gamma_3(k_{3,t}, h_t, t, e_{j,3}) = & e_{j,3} + \beta_{3,1} \cdot h_t & (2.A.8) \\ & + \gamma_{3,1} \cdot k_{3,t} + \gamma_{3,2} \cdot (k_{3,t})^2 + \gamma_{3,3} \cdot \mathbf{I}[k_{3,t} > 0] \\ & + \gamma_{3,4} \cdot t + \gamma_{3,5} \cdot \mathbf{I}[t < 18]. \end{aligned}$$

Unlike the civilian sector, there are no sheep-skin effects from completing military training, ($\beta_{3,2} = \beta_{3,3} = 0$). Furthermore, the previous occupational choice has no influence ($\gamma_{3,6} = 0$), and any experience other than military is non-transferable ($\gamma_{3,7} = \gamma_{3,8} = 0$).

Remark 1. Our parameterization for the immediate utility of serving in the military differs from Keane and Wolpin (1997), as we remain unsure about their exact specification. The authors state in Footnote 31 (p. 498) that the constant for the non-pecuniary utility $\alpha_{3,t}$ depends on age. However, we are unable to determine the precise nature of the relationship. Equation (C3) (p. 521) also indicates no productivity shock $\epsilon_{a,t}$ in the wage component. Table 7 (p. 500) reports such estimates.

Table 2.A.1 presents an overview of the model's parameters.

Table 2.A.1. Overview of parameters in the Keane and Wolpin (1997) extended model.

Parameter	Description
Preference and type-specific parameters	
δ	discount factor
$e_{j,a}$	initial endowment of type j in alternative a specific skills
Common parameters immediate utility	
α_a	return on non-wage working conditions
ϑ_1	non-pecuniary premium for finishing high school
ϑ_2	non-pecuniary premium for finishing college
ϑ_3	non-pecuniary premium for leaving the military early
Schooling-related parameters	
$\beta_{a,1}$	return on each additional year of completed schooling
$\beta_{a,2}$	skill premium for high school graduates

continued on next page

 continued from previous page

$\beta_{a,3}$	skill premium for college graduates
β_{tc_1}	tuition costs for high school
β_{tc_2}	tuition costs for college
β_{rc_1}	re-entry costs for high school
β_{rc_2}	re-entry costs for college
$\beta_{5,2}$	skill premium for high school graduates
$\beta_{5,3}$	skill premium for college graduates

Experience-related parameters

$\gamma_{a,1}$	return on same-sector experience
$\gamma_{a,2}$	return squared on same-sector experience
$\gamma_{a,3}$	premium for having previously worked in sector
$\gamma_{a,4}$	return on age effect
$\gamma_{a,5}$	return on age effect for minors
$\gamma_{a,6}$	premium for remaining in same sector
$\gamma_{a,7}$	return on civilian cross-sector experience
$\gamma_{a,8}$	return on non-civilian sector experience
$\gamma_{3,1}$	return on same-sector experience
$\gamma_{3,2}$	return squared on same-sector experience
$\gamma_{3,3}$	premium for having previously worked in sector
$\gamma_{3,4}$	return on age effect
$\gamma_{3,5}$	return on age effect for minors
$\gamma_{4,4}$	return on age effect
$\gamma_{4,5}$	return on age effect for minors
$\gamma_{5,4}$	return on age, between 17 and 21
$\gamma_{5,5}$	return on age, older than 21

Mobility and search parameters

$c_{a,1}$	premium for switching to occupation a
$c_{a,2}$	premium for working in occupation a for the first time
$c_{3,2}$	premium for serving in the military for the first time

Error correlation

 continued on next page

 continued from previous page

$\sigma_{a,a}$	standard deviation of shock in alternative a
$\sigma_{i,j}$	correlation between shocks in alternative $a = i$ and $a = j$ with $i \neq j$

Note: The above list is an overview of the model parameters. The immediate utilities for the alternatives do not necessarily include all of them.

2.A.2 Data

We use the same data as Keane and Wolpin (1997), who derive their sample from the National Longitudinal Survey of Youth 1979 (NLSY79) (Bureau of Labor Statistics, 2019). The NLSY79 is a nationally representative sample of young men and women living in the United States in 1979 and born between 1957 and 1964. Individuals were followed from 1979 onwards and repeatedly interviewed about their educational decisions and labor market experiences. Based on this information, individuals are assigned to either working in one of three occupations, attending school, or simply staying at home. The decision period is represented by the school year. The sample is restricted to white men, who turned 16 between 1977 and 1981, and it uses information collected between 1979 and 1987. Thus, the individuals in the sample range in age between 16 and 26 years old.

Figure 2.A.1 shows the sample size by age. While the sample initially consists of 1,373 16-year-olds, this value drops to 256, once the sampled individuals reach the age of 26 due to sample attrition, missing data, and the short observation period. Overall, the final sample consists of 12,359 person-period observations.

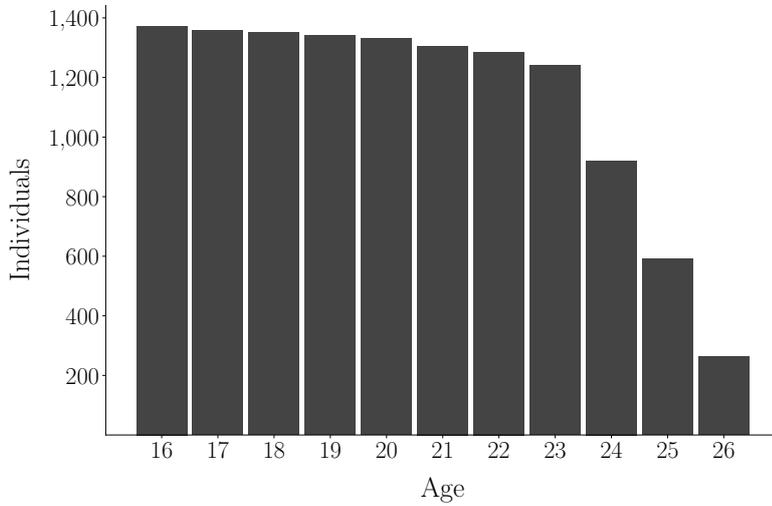


Figure 2.A.1. Sample size

Figure 2.A.2 shows the distribution of initial schooling among individuals at the time they enter the model. The majority of individuals enter the model with ten years of schooling, while about a quarter of the sample has less than ten years of schooling. About 7.5% of individuals already attended school for 11 years.

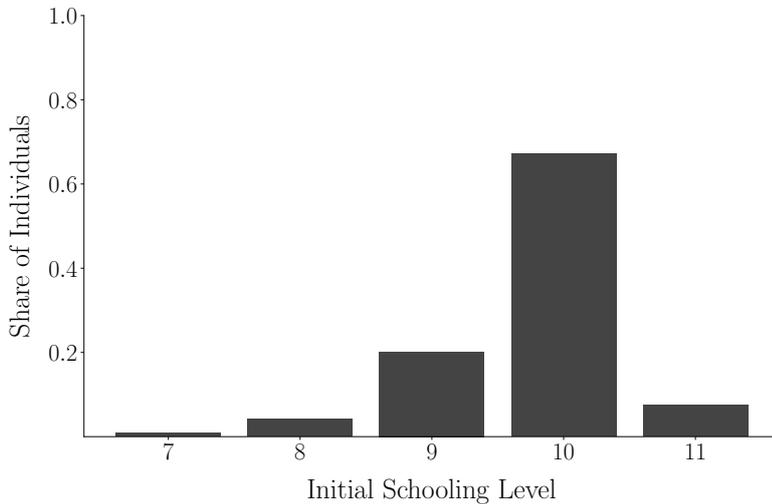


Figure 2.A.2. Initial schooling

Figure 2.A.3 shows heterogeneity of choices by the level of initial schooling. Individuals who enter the model with only seven years of schooling spend an additional 0.65 years in school after age 16. Consequently, they spend around

four years at home. In the event that they are working, it is likely in a blue-collar occupation. When starting with ten years of schooling, then individuals add roughly another three years while in the model. This increase is about half a year more than individuals that start with eleven years.

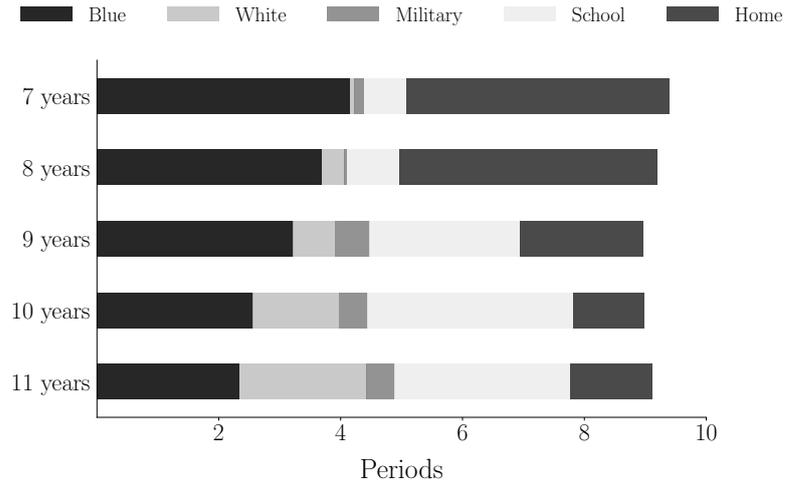


Figure 2.A.3. Average choices by initial schooling

Figure 2.A.4 documents strong persistence in choices over time. For example, among those with a white-collar occupation in t , 67% work in the same occupation in $t + 1$, while 20% switch to a blue-collar job.

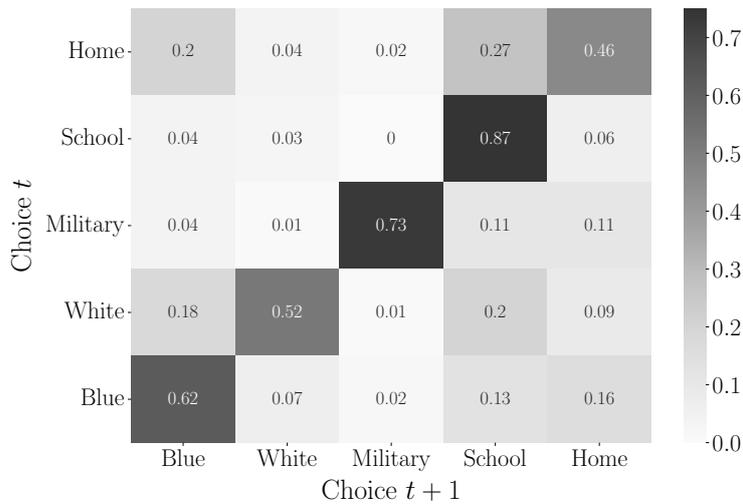


Figure 2.A.4. Transition matrix

2.A.3 Results

Figure 2.A.5 shows further comparisons between the simulated and empirical data. All results from the estimated model are based on 10,000 individuals.

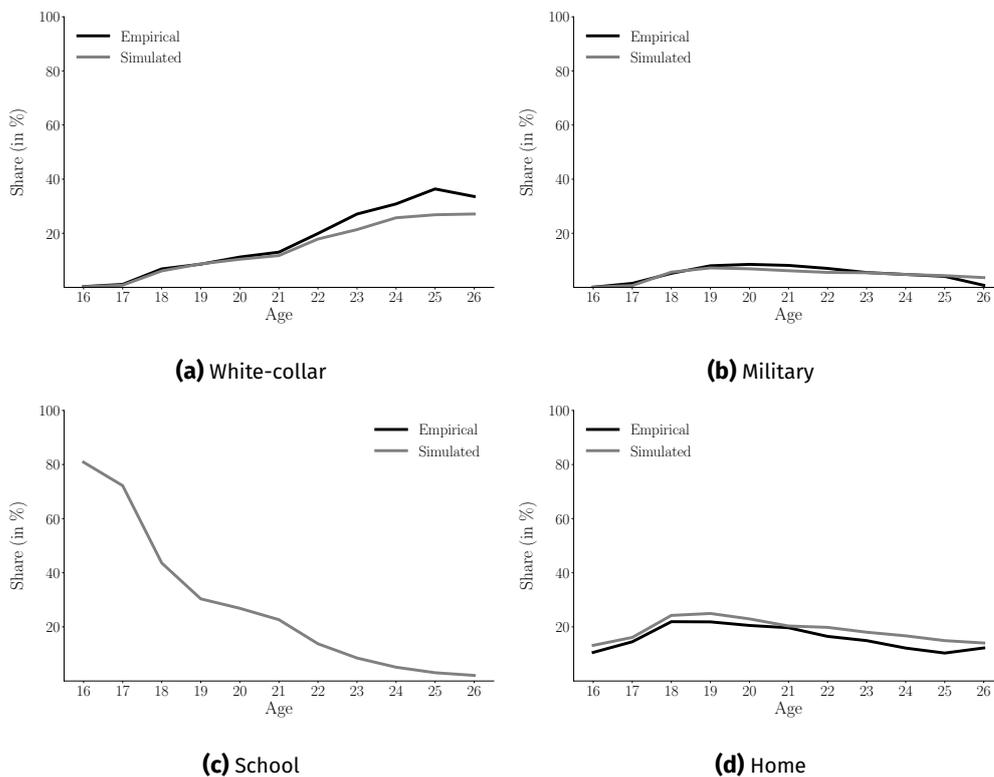


Figure 2.A.5. Model fit

Figure 2.A.6 provides the point prediction, its sampling distribution, and the estimated confidence set for the impact of the tuition subsidy on all types. All results are based on 30,000 draws from the asymptotic normal distribution of our parameter estimates.

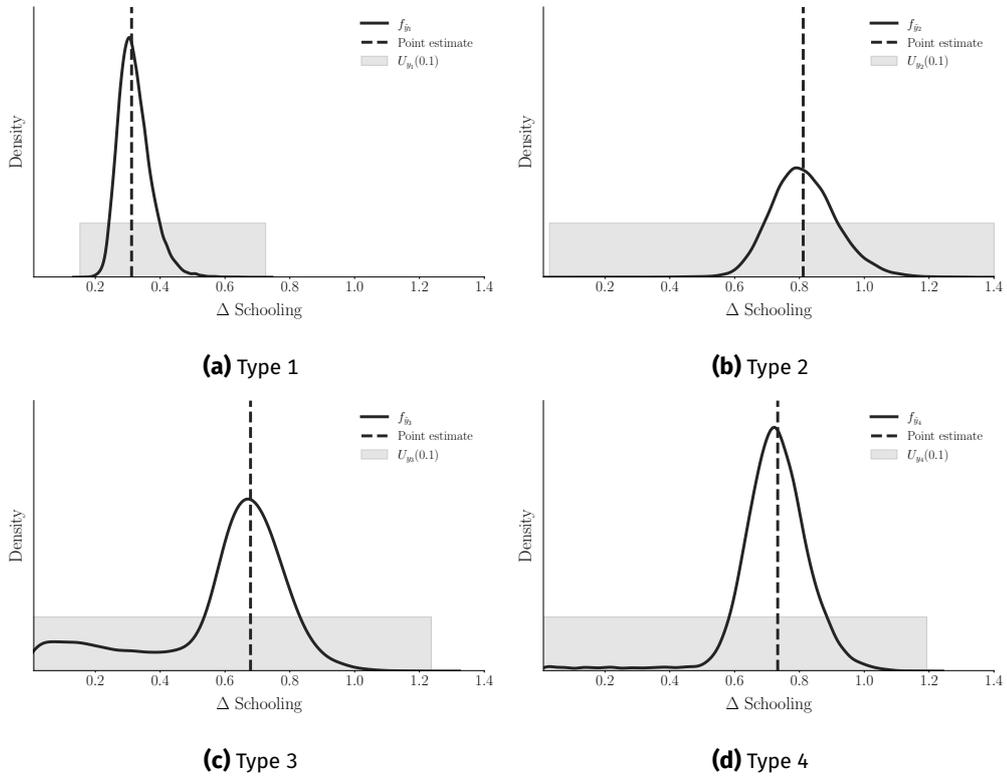


Figure 2.A.6. Targeted subsidy for all types

Figure 2.A.7 shows the impact of the tuition subsidy at the upper δ_H and lower δ_L bound of the estimated confidence set for δ . The results for both scenarios are based on simulated samples of 10,000 individuals.

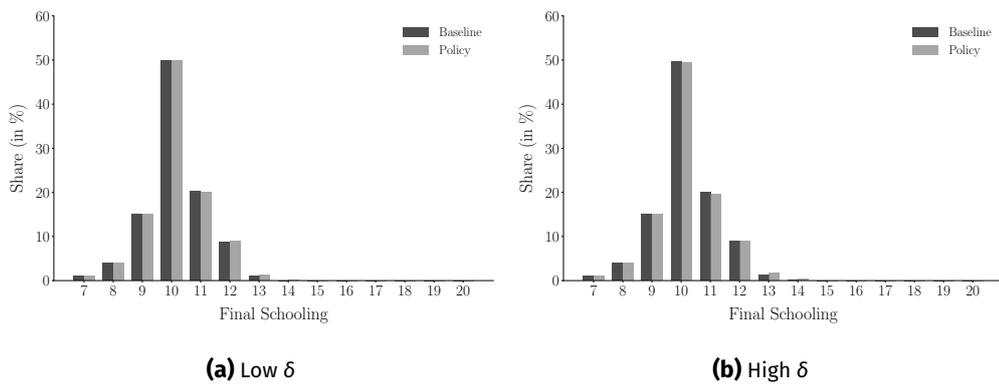


Figure 2.A.7. Policy impact and time preference

References

- Adda, Jérôme, Christian Dustmann, and Katrien Stevens.** 2017. "The career costs of children." *Journal of Political Economy* 125 (2): 293–337. [90]
- Aguirregabiria, Victor, and Pedro Mira.** 2010. "Dynamic Discrete Choice Structural Models: A survey." *Journal of Econometrics* 156 (1): 38–67. [91, 98]
- Altonji, Joe.** 1993. "The Demand for and Return to Education when education outcomes are uncertain." *Journal of Labor Economics* 11 (1): 48–83. [101, 114]
- Andrews, Isaiah, Matthew Gentzkow, and Jesse M. Shapiro.** 2017. "Measuring the sensitivity of parameter estimates to estimation moments." *Quarterly Journal of Economics* 132 (4): 1553–92. [91]
- Andrews, Isaiah, Matthew Gentzkow, and Jesse M. Shapiro.** 2020. "On the informativeness of descriptive statistics for structural estimates." *Econometrica* 88 (6): 2231–58. [91]
- Armstrong, Timothy B, and Michal Kolesár.** 2021. "Sensitivity analysis using approximate moment condition models." *Quantitative Economics* 12 (1): 77–108. [92, 111]
- Ben-Tal, Aharon, Dick den Hertog, Anja De Waegenaere, Bertrand Melenberg, and Gijs Rennen.** 2013. "Robust Solutions of Optimization Problems Affected by Uncertain Probabilities." *Management Science* 59 (2): 341–57. [90, 95]
- Berger, Loïc, Nicolas Berger, Valentina Bosetti, Itzhak Gilboa, Lars Peter Hansen, Christopher Jarvis, Massimo Marinacci, and Richard D Smith.** 2021. "Rational policymaking during a pandemic." *Proceedings of the National Academy of Sciences* 118 (4): [90]
- Bertsimas, Dimitris, Vishal Gupta, and Nathan Kallus.** 2018. "Data-driven robust optimization." *Mathematical Programming* 167 (2): 235–92. [95]
- Blesch, Maximilian, and Philipp Eisenhauer.** 2021. "Robust decision-making under risk and ambiguity." *arXiv preprint arXiv:2104.12573* submitted: [92, 107]
- Blundell, Richard.** 2017. "What have we learned from structural models?" *American Economic Review* 107 (5): 287–92. [91]
- Blundell, Richard, Monica Costa Dias, Costas Meghir, and Jonathan Shaw.** 2016. "Female labor supply, human capital, and welfare reform." *Econometrica* 84 (5): 1705–53. [90]
- Blundell, Richard, and Andrew Shephard.** 2012. "Employment, hours of work and the optimal taxation of low-income families." *Review of Economic Studies* 79 (2): 481–510. [90]
- Bonhomme, Stéphane, and Martin Weidner.** 2021. "Minimizing sensitivity to model misspecification." *Quantitative Economics, forthcoming*, [92, 111]
- Bugni, Federico, and Takuya Ura.** 2019. "Inference in dynamic discrete choice problems under local misspecification." *Quantitative Economics* 10 (1): 67–103. [92]
- Bureau of Labor Statistics.** 2019. *National Longitudinal Survey of Youth 1979 cohort, 1979-2016 (rounds 1-27)*. Columbus, OH: Center for Human Resource Research, Ohio State University. [91, 99, 118]
- Cai, Yongyang, and Thomas S. Lontzek.** 2019. "The social cost of carbon with economic and climate risks." *Journal of Political Economy* 127 (6): 2684–734. [91]

- Caucutt, Elizabeth M., and Lance Lochner.** 2020. "Early and late human capital investments, borrowing constraints, and the family." *Journal of Political Economy* 128 (3): 1065–147. [108]
- Christensen, Timothy, and Benjamin Connault.** 2019. "Counterfactual sensitivity and robustness." *arXiv preprint arXiv:1904.00989*, [91]
- Cunha, Flavio, James J. Heckman, and Susanne Schennach.** 2010. "Estimating the technology of cognitive and noncognitive skill formation." *Econometrica* 78 (3): 883–931. [90]
- Eckstein, Zvi, Michael Keane, and Osnat Lifshitz.** 2019. "Career and Family Decisions: Cohorts Born 1935-1975." *Econometrica* 87 (1): 217–53. [90]
- Efron, Bradley.** 1979. "Bootstrap methods: Another look at the jackknife." *Annals of Statistics* 7 (1): 1–26. [105]
- Eisenhauer, Philipp, James J. Heckman, and Stefano Mosso.** 2015. "Estimation of dynamic discrete choice models by maximum likelihood and the simulated method of moments." *International Economic Review* 56 (2): 331–57. [90]
- Erosa, Andres, Tatyana Koreshkova, and Diego Restuccia.** 2010. "How important is human capital? A quantitative theory assessment of world income inequality." *Review of Economic Studies* 77 (4): 1421–49. [108]
- Fisher, R A.** 1922. "On the Mathematical Foundations of Theoretical Statistics." *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 222 (594-604): 309–68. [104]
- Forrester, Dr Alexander I. J., Dr Andras Sobester, and Professor Andy J. Keane.** 2008. *Engineering design via surrogate modelling: A practical guide*. Chichester, England: John Wiley & Sons. [111]
- Gabler, Janoś, and Tobias Raabe.** 2020. "respy - A framework for the simulation and estimation of Eckstein-Keane-Wolpin models." [99]
- Gayle, George-Levi, and Andrew Shephard.** 2019. "Optimal taxation, marriage, home production, and family labor supply." *Econometrica* 87 (1): 291–326. [90]
- Gilboa, Itzhak.** 2009. *Theory of decision under uncertainty*. New York City, NY: Cambridge University Press. [95, 98]
- Gilboa, Itzhak, Maria Rouziou, and Olivier Sibony.** 2018. "Decision theory made relevant: Between the software and the shrink." *Research in Economics* 72 (2): 240–50. [90]
- Gilboa, Itzhak, and Larry Samuelson.** 2021. "What were you thinking? Revealed preference theory as coherence test." *Theoretical Economics, forthcoming*, [90]
- Gilboa, Itzhak, and David Schmeidler.** 1989. "MaxMin Expected Utility With Non-Unique Prior." *Journal of Mathematical Economics* 18 (2): 141–53. [96]
- Harenberg, Daniel, Stefano Marelli, Bruno Sudret, and Viktor Winschel.** 2019. "Uncertainty quantification and global sensitivity analysis for economic models." *Quantitative Economics* 10 (1): 1–41. [91]
- Hood, William, and Tjalling Koopmans.** 1953. *Studies in econometric method*. New York City, NY: John Wiley & Sons. [89]

- Hungerford, Thomas, and Gary Solon.** 1987. "Sheepskin Effects in the Returns to Education." *Review of Economics and Statistics* 69 (1): 175–77. [115]
- Jaeger, David A, and Marianne E Page.** 1996. "Degrees matter: New evidence on sheepskin effects in the returns to education." *Review of Economics and Statistics* 78 (4): 733–40. [115]
- Jørgensen, Thomas H.** 2021. "Sensitivity to calibrated parameters." *Review of Economics and Statistics*, forthcoming, [92]
- Kahneman, Daniel, Peter P. Wakker, and Rakesh Sarin.** 1997. "Back to Bentham? Explorations of Experienced Utility." *Quarterly Journal of Economics* 112 (2): 375–406. [99]
- Kalouptzidi, Myrto, Yuichi Kitamura, Lucas Lima, and Eduardo A. Souza-Rodrigues.** 2021. "Counterfactual analysis for structural dynamic discrete choice models." *NBER Working Paper Series* No. 26761: [92]
- Kalouptzidi, Myrto, Paul T. Scott, and Eduardo Souza-Rodrigues.** 2021. "Identification of counterfactuals in dynamic discrete choice models." *Quantitative Economics* 12 (2): 351–403. [92]
- Keane, Michael P., Petra E. Todd, and Kenneth I. Wolpin.** 2011. "The structural estimation of behavioral models: Discrete choice dynamic programming methods and applications." In *Handbook of Labor Economics*. Edited by Orley Ashenfelter and David Card. Amsterdam, Netherlands: Elsevier Science, 331–461. [90, 91]
- Keane, Michael P., and Kenneth I. Wolpin.** 1997. "The career decisions of young men." *Journal of Political Economy* 105 (3): 473–522. [91–93, 98–102, 107, 108, 111, 112, 116, 118]
- Low, Hamish, and Costas Meghir.** 2017. "The use of structural models in econometrics." *Journal of Economic Perspectives* 31 (2): 33–58. [91]
- Machina, Mark J, and Kip Viscusi.** 2014. *Handbook of the economics of risk and uncertainty*. Amsterdam, Netherlands: North-Holland Publishing Company. [98]
- Manski, Charles F.** 2004. "Statistical Treatment Rules for Heterogeneous Populations." *Econometrica* 72 (4): 1221–46. [97]
- Manski, Charles F.** 2019. "Communicating uncertainty in policy analysis." *Proceedings of the National Academy of Sciences* 116 (16): 7634–41. [90]
- Manski, Charles F, and Steven R Lerman.** 1977. "The Estimation of Choice Probabilities from Choice Based Samples." *Econometrica* 45 (8): 1977–88. [104]
- Manski, Charles F.** 2013. *Public policy in an uncertain world: Analysis and decisions*. Cambridge, MA: Harvard University Press. [90]
- Manski, Charles F.** 2021. "Econometrics for decision making: Building foundations sketched by Haavelmo and Wald." *Econometrica* 89 (6): 2827–53. [90, 94, 107]
- Marinacci, Massimo.** 2015. "Model uncertainty." *Journal of the European Economic Association* 13 (6): 1022–100. [95]
- Mincer, Jacob.** 1974. *Schooling, experience and earnings*. New York City, NY: National Bureau of Economic Research. [115]
- Mukhin, Yaroslav.** 2018. "Sensitivity of regular estimators." *arXiv Working Paper*, [92]
- Muth, John F.** 1961. "Rational Expectations and the Theory of Price Movements." *Econometrica* 29 (3): 315–35. [99]

- National Research Council.** 2012. *Assessing the reliability of complex models: Mathematical and statistical foundations of verification, validation, and uncertainty quantification*. Washington, DC: The National Academies Press. [90]
- Niehaus.** 1948. "Zur Preisbildung bei ungewissen Erwartungen." *Swiss Journal of Economics and Statistics* 84 (5): 433–56. [97]
- Norets, A., and X. Tang.** 2014. "Semiparametric inference in dynamic binary choice models." *Review of Economic Studies* 81 (3): 1229–62. [92]
- Owen, Art B.** 2014. "Sobol' Indices and Shapley Value." *Journal on Uncertainty Quantification* 2 (1): 245–51. [111]
- Puterman, Martin L.** 1994. *Markov decision processes: Discrete stochastic dynamic programming*. New York City, NY: John Wiley & Sons. [99]
- Rao, Calyampudi R.** 1973. *Linear statistical inference and its applications*. New York City, NY: John Wiley & Sons. [105]
- Razavi, Saman, Anthony Jakeman, Andrea Saltelli, Clémentine Prieur, Bertrand Iooss, Emanuele Borgonovo, Elmar Plischke, Samuele Lo Piano, Takuya Iwanaga, William Becker, et al.** 2021. "The Future of Sensitivity Analysis: An essential discipline for systems modeling and policy support." *Environmental Modeling & Software* 137: 104954. [111]
- Reich, Gregor, and Kenneth L. Judd.** 2020. "Efficient likelihood ratio confidence intervals using constrained optimization." *SSRN Working Paper*, [105]
- Renn, Ortwin, Maria Baghramian, and Massimo Capaccioli.** 2019. "Making sense of science for policy under conditions of complexity and uncertainty." Working paper. Science Advice for Policy by European Academies. [90]
- Rust, John.** 1987. "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher." *Econometrica* 55 (5): 999–1033. [92]
- Rust, John.** 1994. "Structural Estimation of Markov Decision Processes." In *Handbook of Econometrics*. Edited by Robert Engle and Daniel McFadden. Amsterdam, Netherlands: North-Holland Publishing Company, 3081–143. [99]
- Saltelli, Andrea, Gabriele Bammer, Isabelle Bruno, Erica Charters, Monica Di Fiore, Emmanuel Didier, Wendy Nelson Espeland, John Kay, Samuele Lo Piano, Deborah Mayo, et al.** 2020. "Five ways to ensure that models serve society: A manifesto." *Nature* 582: 482–84. [89]
- Samuelson, Paul A.** 1937. "A note on measurement of utility." *Review of Economic Studies* 4 (2): 155–61. [99]
- Savage, Leonard J.** 1954. *The foundations of statistics*. New York City, NY: John Wiley & Sons. [98]
- Shapley, L. S.** 1953. *A value for n -person games*. Princeton, NJ: Princeton University Press, 307–17. [111]
- Todd, Petra E, and Kenneth I Wolpin.** 2007. "The Production of Cognitive Achievement in Children: Home, School and Racial Test Score Gaps." *Journal of Human Capital* 1 (1): 91–136. [108]
- Todd, Petra E., and Kenneth I. Wolpin.** 2006. "Assessing the impact of a school subsidy program in Mexico: Using a social experiment to validate a dynamic behavioral model of child schooling and fertility." *American Economic Review* 96 (5): 1384–417. [90]

- Wald, Abraham.** 1950. *Statistical decision functions*. New York City, NY: John Wiley & Sons, 165–205. [91, 96, 107]
- White, D. J.** 1993. *Markov decision processes*. New York City, NY: John Wiley & Sons. [99]
- Wolpin, Kenneth I.** 2013. *The limits to inference without theory*. Cambridge, MA: MIT University Press. [89]
- Woutersen, Tiemen, and John Ham.** 2019. “Confidence sets for continuous and discontinuous functions of parameters.” *Journal of Economics*, forthcoming, [91, 105]

Chapter 3

Mens Sana in Corpore Sano?

Joint with Hans-Martin von Gaudecker, Jürgen Maurer and Mariam Petrosyan

3.1 Introduction

Development and maintenance of human capital throughout the life-course enables individuals to lead longer, more productive and more satisfactory lives. The notion of human capital generally comprises a broad range of useful abilities that shape individuals' capabilities, behaviors and wellbeing such as their knowledge, skills, and health among others (World Bank, 2018). While there is a large economic literature on early-life human capital development and its effects on adult outcomes (Heckman and Mosso, 2014), fewer studies in economics have analyzed the roles individual investments and corresponding technologies for the maintenance and depreciation of human capital during later life within an integrated framework to model later-life human capital dynamics (McFadden, 2008).

Physical and cognitive capacity represent two key forms of human capital during adulthood and are perhaps the most important forms of human capital at older ages, especially after retirement. Physical and cognitive capacity are key determinants of many important outcomes in health economics and beyond such as mortality, healthcare use and healthcare cost and spending, falls and disability, long-term care needs and nursing home use, economic and social participation and subjective wellbeing to name but a few. As a result, investments in the maintenance of physical and cognitive capacity are key to ensuring a healthier, longer, and happier old-age. Moreover, since many of these outcomes are highly uncertain, demand for various healthcare

*

We would like to thank Johannes Ewald for preparatory work in his M.Sc. thesis at the University of Bonn (March 2020). The authors are grateful for support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1– 390838866 – and through CRC-TR 224 (Project C01).

and long-term care related insurance products depends on the later-life dynamics of physical and cognitive capacity (Hosseini, Kopecky, and Zhao, 2022). Understanding the later-life dynamics of physical and cognitive capacity is, therefore, a key pre-requisite and input into models aimed at studying the role of later-life human capital on these important later-life outcomes and related investment and insurance decisions.

While physical and cognitive capacity tend to decline during later life (Niccoli and Partridge, 2012), there is considerable heterogeneity in the onset and speed of such aging-related declines across individuals, which is often related to individual differences in exposures and investments (Crimmins, 2020). What is more, several studies have in fact shown significant improvements in later life physical and cognitive capacity following targeted investments such as physical exercise programs or cognitive trainings, suggesting that both physical and cognitive function remain malleable even at very high ages (Fiatarone, O'Neill, Ryan, Clements, Solares, et al., 1994; Ball, Berch, Helmers, Jobe, Leveck, et al., 2002). This evidence suggests that aging-related changes in function are not fully pre-determined biologically but can be postponed, slowed down, compensated and in certain instances perhaps even (temporarily) reversed or overcompensated through appropriate later-life investments. These findings highlight the important role of health investments for physical and cognitive capacity throughout the entire life course, even if early-life health investments into health to build up "reserves" for later life may be more efficient due to a higher degree of malleability early in life, the longer time horizon available to capitalize on early investments and potentially important complementarities of health investments over time (Cunha, Heckman, and Schennach, 2010).

Besides documenting the continued malleability of physical and cognitive capacity during later life, the more recent literature in gerontological science has also found for evidence potentially important cross-effects of physical function on cognitive function and vice versa. These cross-effects may go beyond the responses of physical and cognitive function due to common risk factors such as physical inactivity or diseases affecting both physical and cognitive capacities such as Parkinson's disease, and represent more general connections between physical and cognitive capacity (Clouston, Brewster, Kuh, Richards, Cooper, et al., 2013). Evidence for such connections comes from both observational studies and RCTs, often but not always focused on the connection between cognitive and gait (dys-)function (Montero-Odasso, Verghese, Beauchet, and Hausdorff, 2012). In view of these findings, economic models of human capital maintenance and depreciation during later life should thus allow for flexible later-life dynamics of physical and cognitive capacities that can incorporate different forms of investment, and possible cross-effects between physical and cognitive capacities.

Varied existing conceptualizations of physical and cognitive capacity used in the literature and potentially widespread measurement error in physical and cognitive assessments in survey data and self-reported health investments further com-

plicate the already complex task of capturing the joint dynamics of later-life physical and cognitive capacity and related investments (Bound, Brown, and Mathiowetz, 2001; Baker, Stabile, and Deri, 2004; Kapteyn, Banks, Hamer, Smith, Steptoe, et al., 2018; Hosseini, Kopecky, and Zhao, 2022). Physical capacity, for example, is a multifaceted concept that is generally assessed through multiple self-reported and/or performance-based survey items presenting noisy measurements for underlying true physical capacity (Kasper, Chan, and Freedman, 2017). Similarly, cognition comprises a range of different cognitive functions such as perception, attention, intelligence, knowledge, memory and working memory, judgement, reasoning, computation, problem solving or comprehension, whose corresponding measurements have signal value for overall cognitive capacity (Salthouse, 2010; Salthouse, 2012). Perhaps more surprisingly, even commonly used survey items for health investments such as self-reported physical activity contain substantial measurement error relative to actual health investments and, therefore, need to be treated with caution (Kapteyn et al., 2018). Given the large potential for significant measurement error in survey-based assessments of physical and cognitive capacity and corresponding health investments documented in the literature, it seems prudent to employ an analytical framework that can readily accommodate such measurement errors when analyzing the joint dynamics of these outcomes.

The main objective of this paper is to estimate the technology for human capital maintenance and depreciation in later-life focusing on the dynamic interplay between later-life physical and cognitive capacity and corresponding investments among older adults in the US. To this end, we propose the use of a non-linear dynamic latent factor model as first proposed by Cunha, Heckman and Schennach (Cunha, Heckman, and Schennach, 2010) as a framework to model early-life human capital accumulation, to study later-life human capital depreciation processes using longitudinal data from the US Health and Retirement Study (HRS). Applying this framework to investigate the joint dynamics of later-life physical and cognitive capacity and related investments is very attractive as such a non-linear dynamic latent factor model can incorporate the main aforementioned stylized facts about human capital depreciation, i.e., (1) allowing for a joint modelling of physical and cognitive capacity and investments that can incorporate potentially important cross-domain effects; (2) integrating the continued malleability of both physical and cognitive capacity into the model to study dynamically optimal investment paths and (3) accounting for error in the measurement of physical and cognitive function and corresponding investments in a context where there are several measurements of each of these domains in many commonly used data sets, but each measurement is likely to provide only a noisy signal for the underlying construct at hand. In addition to accommodating key stylized facts about human capital maintenance and depreciation into a unified framework, our model also allows us to identify the distribution of latent factors from noisy measurements, simulate the effects of different investment patterns on physical and cognitive capacity, calculate optimal investment

patterns, notably the role of investments for human capital maintenance in younger old vs older old individuals, and anchor the results in interpretable metrics such as survival probabilities.

Our paper relates to two strands of research in economics, a methodological one on the use of non-linear dynamic latent factor models for estimating dynamic human capital production, which has to the best of our knowledge-so far only been applied to the case of human capital accumulation in early life but not to human capital maintenance and depreciation in later life, and a more substantive one on the measurement and modelling of health dynamics during adulthood and later life. From a methodological point of view, our paper transfers widely used methods for the study of early-life human capital accumulation to the study of later-life dynamics of physical and cognitive function and eventual mortality. As a technical contribution, we show how to incorporate mortality into the framework and improve the numerical stability of a well known maximum likelihood estimator. By applying non-linear dynamic latent factor models to questions of aging and later life health dynamics, we show the usefulness of these methods to study human development not just in early life but across the entire life-course, especially since many of the modelling and measurement issues mentioned above seem common to both ends of the life-course. As a result, we hope that our paper will inspire a larger group of life-course and aging researchers to consider such models in their research both in health economics and related fields.

Substantively, we contribute to the literature on how to measure and model later-life health dynamics in situations where we observe multiple potentially very noisy measurements for fewer latent concepts such as physical and cognitive capacity, which has long challenged empirical analyses in health economics and beyond. More specifically, one important issue in this literature is how to measure health in a comprehensive yet parsimonious way in view of the multifaceted nature of health on the one hand and the common need for dimensionality reduction in econometric models on the other. To address this trade-off, one set of commonly adopted approach to measuring health is to directly use (usually ordered measurements of) self-rated health as summary measure of health as outcome of interest (Contoyannis, Jones, and Rice, 2004; Heiss, 2011; Latham and Peek, 2012). This approach is generally motivated by a high predictive value of self-rated health for mortality (Idler and Benyamini, 1997). Alternatively to directly using self-reports to measure health, a commonly used approach is to "instrument" health via a larger and "more objective" set of individual health measurements, such as information on specific health conditions, functional limitations, performance test results or anthropometric measures. This approach endogenously derives weights for aggregating the more detailed set of individual health measurements into a single health index that can then be used in further analysis (Cutler and Richardson, 1997; Jürges, 2007). Relative to using self-rated health directly as outcome, the approach aims to improve measurement by using "more objective" measures of health to construct an underlying health in-

dex, whereby the weights attributed to each detailed and "more objective" health measure in the final health index is determined by the partial association of the respective detailed health measure with self-rated health. While this approach can address some known issues with self-rated health, such as potential age-, sex- or SES-dependent reporting heterogeneity (Lindeboom and Van Doorslaer, 2004; Dowd and Zajacova, 2007; Dowd and Zajacova, 2010), there is often still considerable measurement error in the "more objective" health measures that cannot be purged using this approach and may require further consideration (Baker, Stabile, and Deri, 2004; Maurer, Klein, and Vella, 2011). A second related approach side-steps the use of self-rated health entirely and instead uses principal component analysis of the more detailed health measurements to derive lower dimensional health indices (Jenicek, Cleroux, and Lamoureux, 1979; Poterba, Venti, and Wise, 2017; Nakazato, Sugiyama, Ohno, Shimoyama, Leung, et al., 2020). A third and increasingly popular approach simplifies the aggregation process for the more detailed health measurements even further by constructing a so-called "frailty index" or "deficit index", which simply consists of the total number of prevalent "health deficits" divided by the total number of potential "health deficits" (Rockwood and Mitnitski, 2007; Hosseini, Kopecky, and Zhao, 2022). A such constructed "frailty index"/"deficit index" is thus bounded to lie between zero and one and represents the percentage of potential "health deficits" already suffered by a given individual. A final set of studies refrains from performing some form of dimensionality reduction and uses the more detailed health measures directly in their analyses, either in isolation or simultaneously. As this is, for example,, the standard approach of disease-based analyses, most published papers on health adopt this latter approach.

While all of the aforementioned approaches have their respective advantages and disadvantages in measuring and modelling health in economic applications and have been employed with some success in the literature, they have mainly been used to describe the dynamic evolution of health during adulthood as inputs for structural models in health economics concerning retirement, housing or insurance decisions rather than studying the production technology of later life health maintenance or depreciation directly. Regarding the latter, the aforementioned approaches have some potential downsides that we aim to address in this paper. First, to the best of our knowledge, our paper is the first to explicitly study the dynamic interplay between physical capacity, cognitive capacity and related investments in the context of a structural non-linear dynamic latent factor model as first proposed by Cunha, Heckman and Schennach (Cunha, Heckman, and Schennach, 2010), which can generate new insights on the dynamic relationships between physical and cognitive capacity as well as investment into these important facets of human capital. Second, explicitly distinguishing between physical and cognitive capacity is thereby not only important due to increasing evidence for potentially important cross-effects between the two health domains cited above but also in view of likely differences in the consequences of depleted levels of physical vs cognitive capacity for functioning,

participation and other important later life outcomes (Crimmins, 2020; Amengual, Bueren, and Crego, 2021). In the economics literature, there is to date only limited evidence on the potential cross-effects between physical and cognitive capacity maintenance with Schiele and Schmitz (Schiele and Schmitz, 2021) being a notable exception studying the effects of adverse physical health shocks on cognitive capacity in later life using non-structural event study methods. Third, our approach can accommodate a situation where information about a few latent factors needs to be extracted from many measurements of the underlying construct which can potentially suffer from severe measurement error.

Our analysis complements the aforementioned approaches to modelling and analyzing later-life health by delivering new insights on the dynamics of later-life human capital and related investments among older adults in the US. Our approach, thereby, highlights the structural production function of older adults concerning the maintenance and depreciation of physical and cognitive capacity. Our key findings are as follows: 1) There is substantial noise in all observed variables. While most measurements have a high correlation with the latent factor they measure, no single measurement dominates to an extent where it would be justified to just use a single variable and ignore the measurement error in the econometric analysis. 2) Despite a strong decline in means of physical and cognitive capacity, the rank order of these latent factors is remarkably stable. 3) Physical and cognitive capacity can be influenced by investments until very high ages. Cognitive stimulation is a specific investment into cognitive capacity. Physical exercise has a larger effect on physical capacity and a small effect on cognitive capacity.

The remainder of the paper is organized as follows: Section 3.2 provides information on our main data source and gives detailed description of the factor measurements. Section 3.3 describes our empirical approach and the challenges associated with it. Section 3.4 presents and discusses our results, and section 3.5 concludes.

3.2 Data and Measurements

We base our empirical analysis on the 1992-2016 waves of the Health and Retirement Study (HRS) conducted by The University of Michigan. The HRS offers longitudinal panel data with representative sample of approximately 40,000 individuals living in the U.S. and aged 50 and above. The HRS core questionnaire offers rich set of measures of physical health, mental status, and behaviors. Measures of physical and cognitive capacities include self-reported diagnoses, subjective assessments, and objective biomedical markers. Additional off-wave surveys offer additional measures that are particularly relevant for our analysis. Specifically, we employ the Consumption and Activities Mail Survey (CAMS) (Health and Retirement Study, 2022b) to extract measurements for Exercise and Cognitive Stimulation.

Wherever possible, we include data prepared by the RAND corporation (Health and Retirement Study, 2022c), which provides a harmonized and easy-to-use version of the core HRS data. Out of the many variables we need, several are not included in the RAND HRS data, however, and we recur to the original core files (Health and Retirement Study, 2022b).

We start our analysis at age 68, when most people are retired and we start to see meaningful variation in the measures at our disposal for physical and cognitive capacity. The last age we consider is 93, after which the sample size becomes small. Since the HRS questionnaire is administered biannually, we work with two-year transitions and age groups. For conciseness, we refer to these age groups by the lower bound included – “age 68” thus includes ages 68 and 69, and at the other end of the spectrum “age 92” comprises ages 92 and 93. Because men and women show very different aging patterns, we present all statistics by gender. We will also estimate the model separately for each gender.

We standardize almost all measures to have mean zero and unit variance in the first age group included in our data. Any age trends are thus preserved. For example, until age 90, the mean of (residualized) grip strength declines by around 1.4 original standard deviations. At the same time, the dispersion of grip strength shrinks to around 80% of its original standard deviation. For categorical variables, all of which have numerical values with spacing 1, we add noise using uniform distributions on $(-0.5, 0.5)$. This preserves the original ordering and add to the numerical stability of the estimator below. Changing the seed of the random number generator did not affect any results; future work will pursue additional robustness exercises.

3.2.1 Physical Capacity

We employ six variables as measurements for physical capacity. Quite naturally, **vital status** is a dummy for being alive, which becomes zero in the first HRS wave after an individual has died. It is set to missing thereafter, so that the average of this variable can be interpreted as the probability of surviving until the next survey wave. The first row of Figure 3.2.1 shows the age trends in our measures of physical capacity.¹ Unsurprisingly, survival probabilities decreases in age both for women (Figure 3.2.1a) and men (Figure 3.2.1b). Note that the level of survival probabilities is depressed because the HRS is very good at tracking respondents’ dates of death even when they have not responded to previous waves. In this version of the data preparation, individuals who did not respond to a survey round would not enter the denominator of vital status.

The second measurement shown in Figures 3.2.1a and 3.2.1b is a version of the **frailty index** used, for example, in Hosseini, Kopecky, and Zhao (2022). The

1. Figure 3.A.1 in Appendix 3.A shows the same trends for the standard deviations of our measurements.

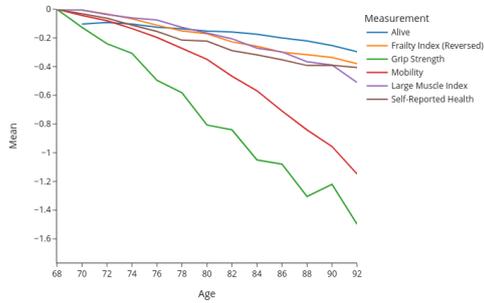
frailty index is the unweighted sum of all recorded medical conditions a doctor has diagnosed in an individual. These conditions comprise high blood pressure, diabetes, cancer, lung disease, heart disease, stroke, psychiatric problems, and arthritis. We reverse it so that higher values indicate better health. The reversed frailty index declines by 0.4 (women) and 0.3 (men) original standard deviations until the end of the age range we consider. Note that this trend and all those we will subsequently discuss are conditional on survival. Due to the high predictive power of the frailty index for mortality—as noted by Hosseini, Kopecky, and Zhao (2022) and others—the effect of mortality selection is particularly large here. For individuals still alive at age 80, average frailty at age 68 is 0.39 among women and 0.34 among men. By including vital status among the health measures, our model below will take care of this to some extent, but it is important to keep in mind for the descriptive statistics.

Grip Strength measurements were introduced to the HRS survey in 2006 and consist of in-home physical tests of the hand grip strength, conducted twice for each hand. To obtain our variable of use, we average the four measurements. Our measure of grip strength is then the residual of a regression of average grip strength on individuals' height. We partial height out because of the high correlation between height and grip strength (Steiber, 2016) and we do not expect differences in grip strength associated with differences in height to be indicative of physical capacity. Among all measures pertaining to physical health, grip strength shows the steepest decline.

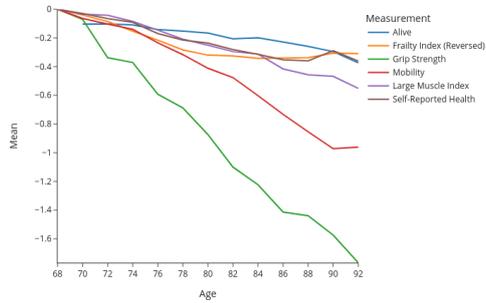
Mobility summarizes difficulties in performing the various activities of daily living: walking several blocks, walking one block, walking across the room, climbing several flights of stairs, and climbing one flight of stairs. As with the frailty index, we add up indicators for each measurement and reverse the scale so that higher values are associated with greater mobility. Mobility declines strongly in age. At the same time, its standard deviation rises as mobility impairments become more frequent over time.

Closely related, the **Large Muscle Index** summarizes difficulties in performing a number of activities associated with large muscles' strength. These activities are sitting for two hours, getting up from a chair, stooping or kneeling or crouching, and pushing or pulling a large object. Again, we revert the order of the values to have a positive association between the variable and physical capacity.

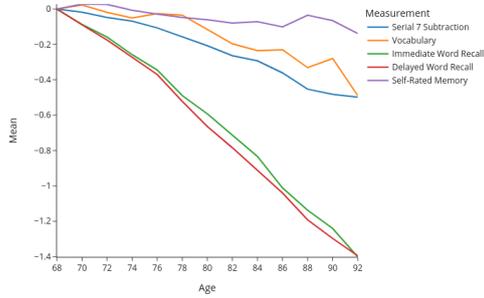
Finally, **Self-Reported Health** is a measure of health that is based on the respondent's self-assessed rating of their general health status. The values range from 1 (poor) to 5 (excellent). It probably is the most common health measure employed by economists as it provides an individuals' summary of her/his health in a single measure.



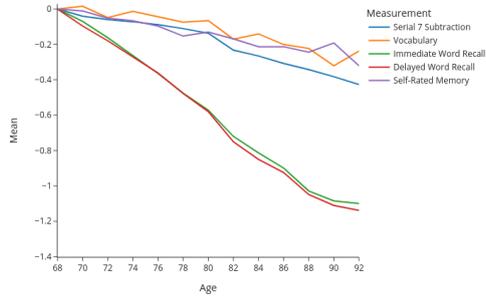
(a) Physical capacity, females



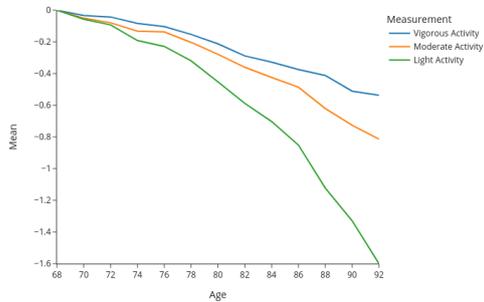
(b) Physical capacity, males



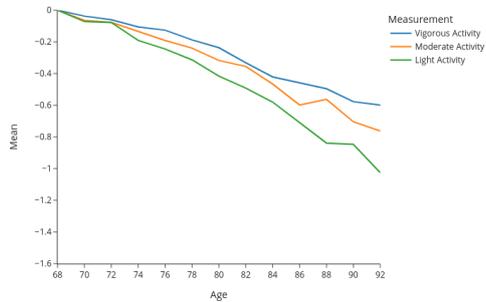
(c) Cognitive capacity, females



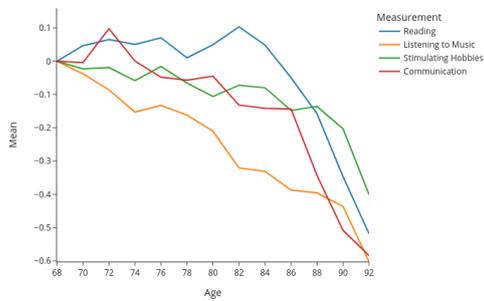
(d) Cognitive capacity, males



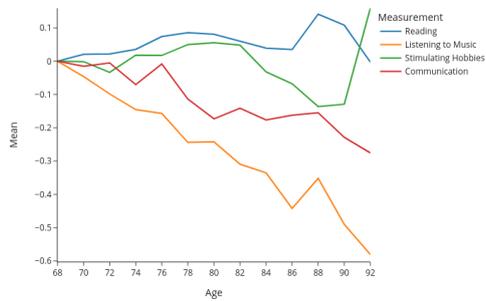
(e) Exercise, females



(f) Exercise, males



(g) Cognitive Stimulation, females



(h) Cognitive Stimulation, males

Figure 3.2.1. Average measurements by age

3.2.2 Cognitive Capacity

During interviews (in-person and via phone) HRS conducts a rich set of tests measuring respondents' cognitive capacity. For respondents that do not answer some of the cognitive test questions, HRS assumes non-random missing values and provides cross-wave imputation data in special data files (Health and Retirement Study, 2022a). Our measures of cognitive capacity are based on these cognitive tests and respondents' subjective ranking of their general memory status. In total, we employ five measures of cognitive capacity.

Serial 7 Subtraction is our first measure and is based on the test of serial sevens (SST) during which respondents are asked to subtract 7 from 100 and from continue subtracting 7 from each resulting number for a maximum of five times. The respondents are then assigned scores based on the total number of correct answers. In psycho-medical literature SST has widely been used to assess mental status of patients with dementia and been generally regarded as a measure of concentration (Karzmark, 2000). Figures 3.2.1c and 3.2.1d demonstrate a steady decline in concentration, as measured by the serial sevens, for both men and women, from our youngest age group to the oldest one being somewhat larger for women (0.5 units of original standard deviation) than for men (0.4 units of original standard deviation).

Our second measure of cognitive capability is **Vocabulary** which is a test summarizing respondents' ability to provide correct definitions of words from a list of five words. One of two sets of words is assigned randomly at the first interview, and alternating sets are given during subsequent interviews. The two alternating sets of words are 1) repair, fabric, domestic, remorse, plagiarize; and 2) conceal, enormous, perimeter, compassion, audacious. We can see in Figures 3.2.1c and 3.2.1d that Vocabulary test has an age trend similar to that of the Serial 7 Subtraction, both in terms of absolute slopes and relative differences between men and women.

Immediate Word Recall is the third variable in Figures 3.2.1c and 3.2.1d and results from a test that asks the respondents to recall words (in any order) from a list of ten (later waves) or twenty (earlier waves) words, directly after being read the list. Examples of words included in a list are lake, car, army, etc. In the initial wave, respondents were randomly assigned a list from the set of four lists and during the consequent four waves there were assigned a different list (McCammon, Fisher, Hassan, Faul, Rodgers, et al., 2022). **Delayed Word Recall** has the same structure as immediate word recall. In this task, respondents are asked to recall the same list of words once more, after spending several minutes on answering other survey questions. Word recall tests are widely used as measures of episodic memory frequently administered to patients with alzheimer's disease (see, e.g., Dixon and Frias, 2014; Runge, 2015).

Both of the word recall variables being measures of the same conceptual variable (episodic memory) perhaps explains the similar trends that they display. Of all the measurements of cognitive capacity, word recall variables have the sharpest decline

over the age span in our model, and as with other measurements, the decline is larger for women than for men, with the caveat that our data are conditional on survival.

Finally, **Self-Rated Memory**, is our last measure of cognitive capacity and is based on respondents' self-assessed rating of their general memory status. The values range from 1 (poor) to 5 (excellent). Self-Rated Memory displays a moderate decline in both genders, which has a somewhat more pronounced trend among men.

3.2.3 Exercise and Cognitive Stimulation

We use **Vigorous, Moderate and Light Activities** as measures for investment in physical health. Each of these survey questions asks respondents how often they do vigorous (running, jogging, cycling, etc.), moderate (gardening, cleaning the car, walking at moderate pace, dancing, stretching) and light/mildly energetic (vacuuming, laundry, home repair), respectively. Up until the sixth wave (year 2002) respondents were only asked if they do vigorous activities at least three times a week. Starting from wave seven, this questionnaire item was replaced by the three activity questions that we use in our study. Figures 3.2.1e and 3.2.1f show that with age people do less of all types of physical activities, with largely similar trends for men and women.

To obtain measures for cognitive stimulation, we utilized the CAMS survey which allowed us to construct measures of time respondents spend on different cognitively stimulating activities. Among these, our first measurement of cognitive stimulation is **Reading** that counts weekly hours spent on reading books, newspapers, or magazines. The association between reading and cognitive decline has been studied in psycho-medical literature, and reading has been found to be positively associated with hampered cognitive decline (Chang, Wu, and Hsiung, 2021). In Figures 3.2.1g and 3.2.1h we see that Reading has declining trend among women and is rather stable among men.

The second variable in Figures 3.2.1g and 3.2.1h is **Listening to Music**, and it measures how many hours weekly respondents listen to music. The effects of music listening on cognitive functioning of at-risk patients have been studied in psycho-medical literature, and listening to music has been found to be beneficial for cognitive functioning (see, e.g., Särkämö, Tervaniemi, Laitinen, Forsblom, Soinila, et al., 2008; Särkämö and Soto, 2012). As with most measurements of cognitive stimulation, we observe a declining age trend for Listening to Music both among men and women.

Our last variables for cognitive stimulation are **Stimulating Hobbies** and **Communication** which summarize how many hours respondents spend weekly on various hobbies that may be expected to stimulate cognition, and the weekly hours spent on interacting with others, respectively. Stimulating Hobbies aggregates the survey variables that ask how many hours respondents spend on: 1) playing cards or

solving jigsaw puzzles, 2) singing or playing instruments, 3) doing arts and crafts, and 4) going to movies or lectures. We construct the Communication variable as the sum of hours spent on visiting with others in person and communication via letters/phone/email. Looking at Figures 3.2.1g and 3.2.1h, Communication has similarly declining trend among men and women, whereas Stimulating Hobbies has a steeper slope for women and than for men.

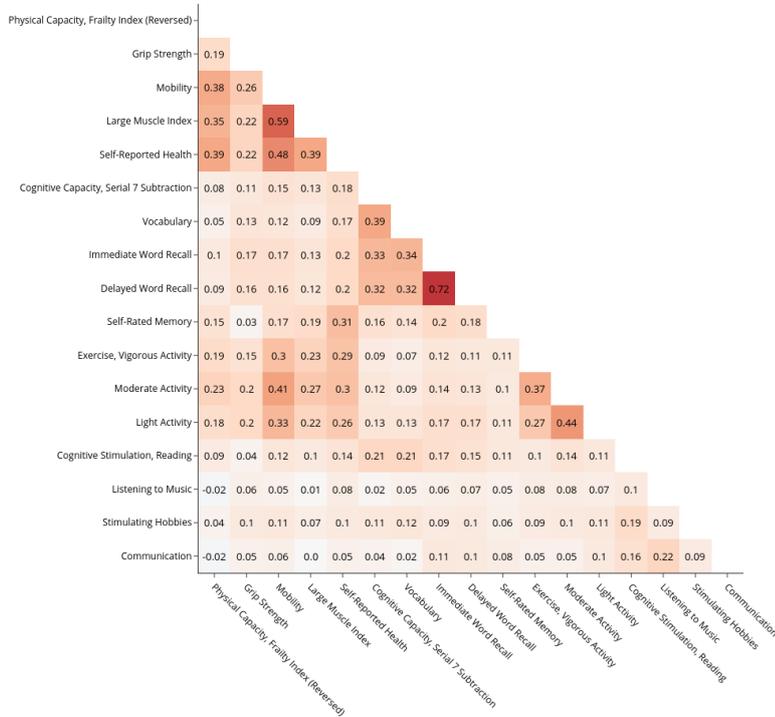
3.2.4 Raw Correlations in the Data

Figures 3.2.2 and 3.2.3 show correlation matrices for women and men, respectively. Each figure contains two panels. The upper panels show within-period correlations until age 79, the lower panels do the same ages 80 and above. We show the lower triangular part of the correlation matrix. We leave out the indicator for being alive because we only measure the other variables whenever it is one. In addition to showing the numbers, we color the matrix' elements such that a correlation of 1 is dark red, 0 is white, and -1 is dark blue. Scaling is linear on both sides of the origin. Variables are ordered by factor, which we include in the label of the first measure pertaining to it. The measures in the first five rows and columns—from the reversed frailty index until self-reported health—load on physical capacity. The subsequent block of five rows and columns load on cognitive capacity. In the lower part of the matrix, exercise and cognitive stimulation load on three and four measures, respectively.

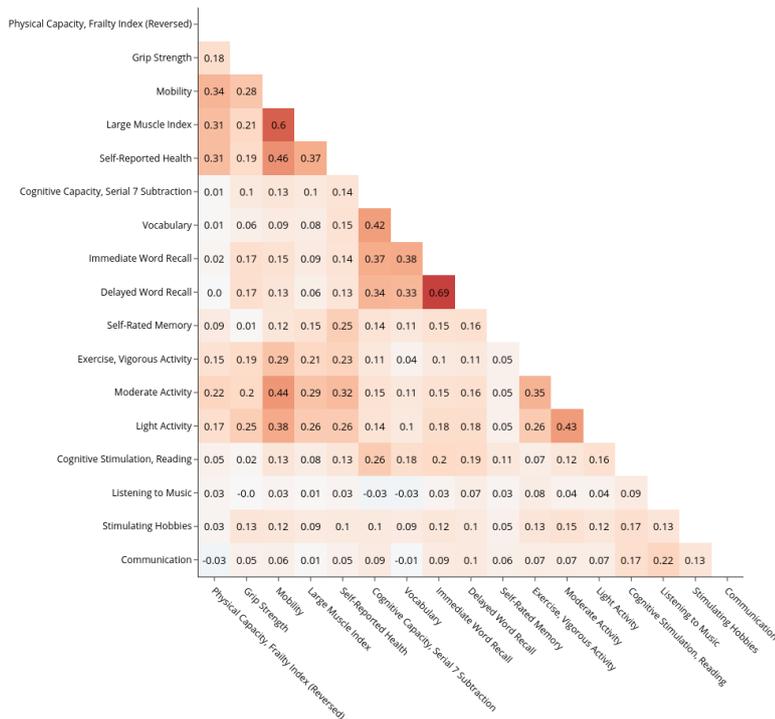
Several patterns are visually apparent in all four correlation matrices. First of all, the blocks of measures pertaining to each factor are clearly visible as having substantial cross-correlation throughout. For example, the first four entries in the first columns are the correlations of the reversed frailty index with the other measures loading on physical capacity. Across all four panels, correlations are at least 0.3 with the exception of the correlation of reversed frailty and grip strength, which is at least 0.1 throughout.

Similarly, the triangle with correlations for measurements pertaining to cognitive capacity—with the three corners (Serial 7 Subtraction, Vocabulary), (Serial 7 Subtraction, Self-Rated Memory), and (Delayed Word Recall, Self-Rated Memory)—has distinctly dark colors throughout. Unsurprisingly, correlations are particularly large between the two word recall tasks. The three correlations between the various types of physical activity are high throughout. The six elements to the bottom right to the matrix contain the correlations among the measures loading on cognitive stimulation. Among all factors, these have the weakest within-factor correlations with values ranging from 0.09 to 0.25. This is not very surprising as the variables do cover a much wider range of activities than, say, the various activity levels that load on exercising.

A second salient feature is that almost all elements are positive. This implies that it is important to model physical and cognitive capacity jointly with each other

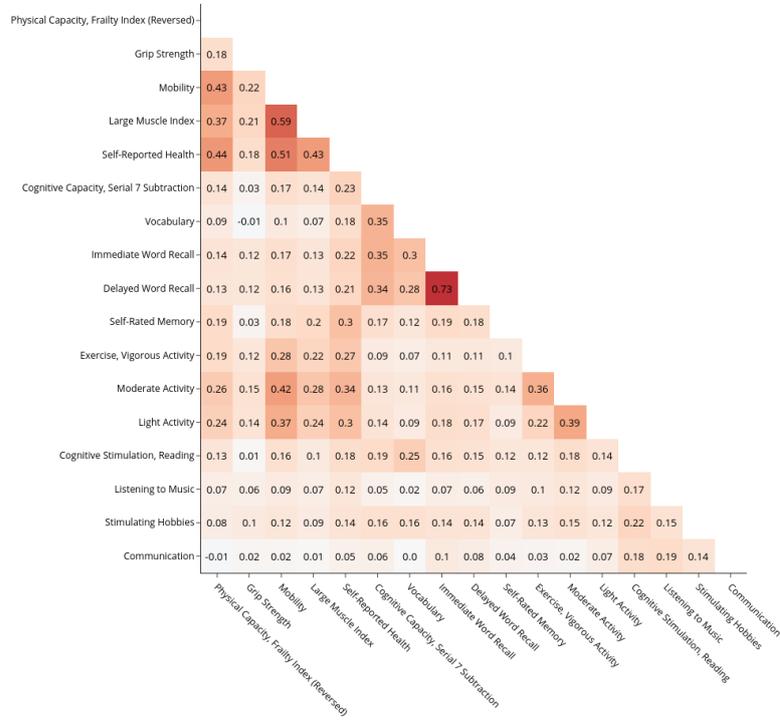


(a) Aged below 80

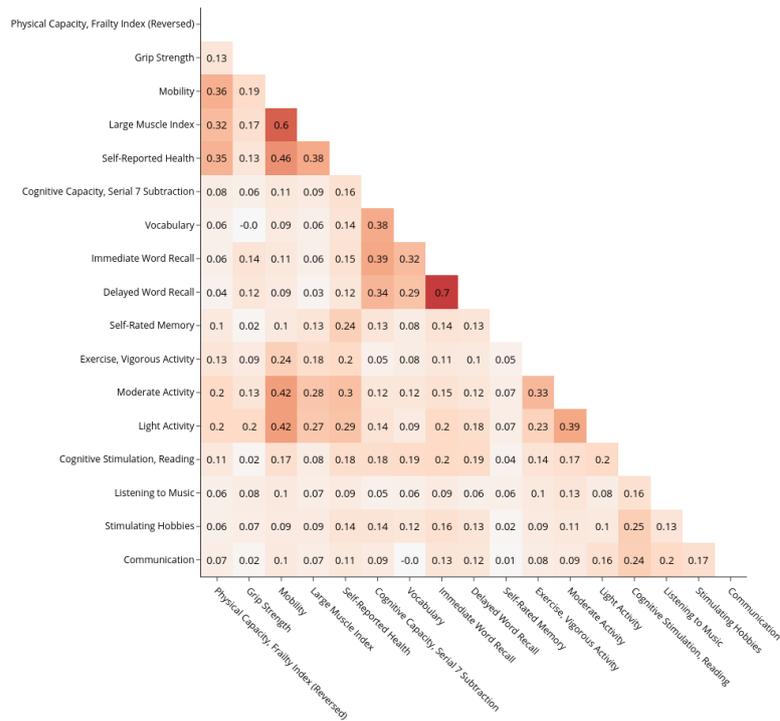


(b) Aged 80 and above

Figure 3.2.2. Cross factor measurement correlations (female).



(a) Aged below 80



(b) Aged 80 and above

Figure 3.2.3. Cross factor measurement correlations (male).

and with the two types of investments. This being written, there are clear level differences. Maybe unsurprisingly, the largest correlations are between measures of exercise and those of physical capacity. Most measures of cognitive capacity are substantially and positively related to variables measuring physical capacity and exercise, respectively. The correlation patterns are somewhat more mixed when it comes to cognitive stimulation and the other three factors.

This is related to our third broad observation: While the general patterns noted so far hold up across age groups and genders, there are some important differences. For example, the correlations of grip strength with other health measures are higher among women than among men, particularly at higher ages. Correlation patterns of individual measures pertaining to cognitive stimulation and cognitive capacity are quite distinct among men and women, particularly at older ages. For example, among individuals aged 80 and above, reading and serial 7 subtraction have a correlation of 0.27 among women whereas it is 0.18 among men. Among women in this age group, listening to music is slightly negatively correlated with serial 7 subtraction and vocabulary scores. For men, the same correlations are small and positive.

While these patterns are informative, the $2 \times 2 \times 153$ numbers in Figures 3.2.2 and 3.2.3 are clearly too many to make sense of directly – and the matrices already reduce the 13 periods we observe in our data to 2. In the next section, we outline a framework that constructs latent variables for our four factors and which allows us to interpret their joint evolution.

3.2.5 Example Transitions

Before going to the formal model, we show a few exemplary trajectories of physical and cognitive capacity. Figure 3.2.4 shows the trajectories of 500 randomly sampled individuals from our dataset. Dots at the end of a trajectory mean that that person died in the next period. Trajectories that do not end in a dot are from individuals whose death was not observed, either because they dropped out of the sample or are still alive in the last wave.

The highlighted lines are hand-picked examples of individuals that had a physical capacity close to the 90th percentile, but very different trajectories afterwards. The blue line shows a person that had a strong decline in physical capacity over two periods and then passed away. The yellow line shows a person with a very volatile trajectory in both physical and cognitive capacity. The red line shows an individual who had a bad health shock at some point but recovered and enjoyed a high level of physical capacity for many years. The right panel of the figure shows the cognitive capacity of the same individuals. All three lines show fluctuations around a robust declining trend.

The plot illustrates that vastly different trajectories of physical and cognitive capacities are possible even for people with similar starting conditions in terms of

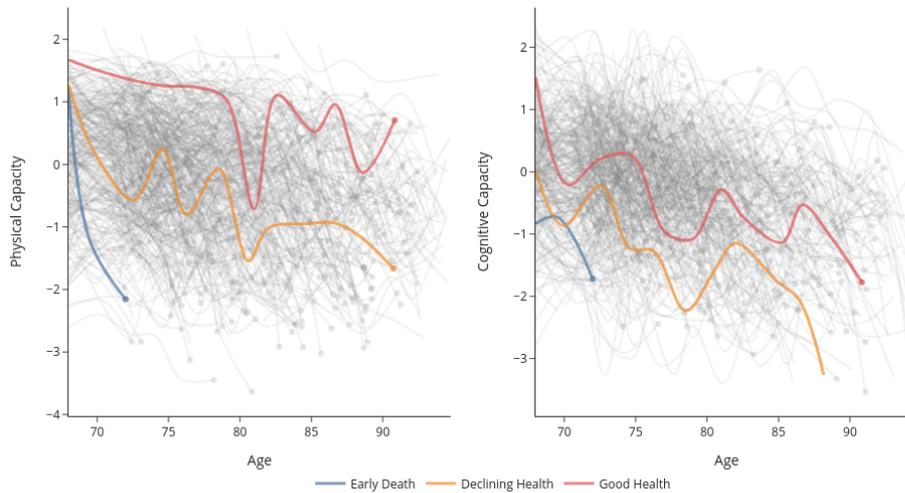


Figure 3.2.4. Trajectories for decline of health and cognitive capacity

physical capacity. Answering the question whether such differences can be explained or are the product of random shocks requires a more rigorous approach.

3.3 Model

3.3.1 The Technology of Aging

Analyzing the joint evolution of physical and cognitive capacity and the effect physical exercise and cognitive stimulation have on both poses many econometric challenges.

1. As discussed in the previous section, there are many potential observed variables to measure each concept we analyze. In order to make the results interpretable, their dimensionality has to be reduced.
2. All observed variables are subject to measurement error, which is potentially large in many cases.
3. Physical and cognitive capacity, exercise, and cognitive stimulation are dynamically intertwined in the sense that each of them has a potential effect on all others. For example, exercise should improve physical capacity. Conversely, it may well be that the cost of exercise might be higher at low levels of physical capacity because physiotherapy is less enjoyable than a walk in nature.
4. The relationships between variables might change over time.

The Technology of Skill Formation (Cunha and Heckman, 2007; Cunha, Heckman, and Schennach, 2010) is an econometric framework that emerged to deal with very similar challenges in the context of skill formation during childhood. It distinguishes observed variables—for example an IQ test—from latent factors such as cognitive and non-cognitive skills. The technology is the law of motion of latent factors over multiple discrete time periods. Observed variables are stochastic functions of one or more latent factors. In addition to the latent factors of interest, the framework allows for observed or latent investments such as parental investments in skills or schooling.

To account for the multitude of potential effects, each latent factor may depend on lagged values of itself and all other latent factors. The law of motion of the latent factors is usually nonlinear. This is necessary to allow for different productivity of investments at different levels of skills. Moreover, it allows for dynamic complementarity, i.e., the fact that earlier investments may increase the productivity of later investments (Cunha and Heckman, 2007).

The Technology of Skill Formation maps perfectly on our setting. Instead of cognitive and non-cognitive skills, our Technology of Aging models physical and cognitive capacity. Instead of parental investments, we have exercise and cognitive stimulation. While we separate investments into a physical and cognitive component, we allow each investment factor to influence both latent capacities.

Transition Functions

We assume the following law of motion of our latent factors:

$$\begin{aligned}
 x_{1,t+1} &= \beta_{1,t} + \sum_{i=1}^4 \gamma_{1,t,i} x_{t,i} + \sum_{i=1}^4 \sum_{j=1}^i \delta_{1,t,i,j} x_{t,i} x_{j,t} + \eta_{1,t} \\
 x_{2,t+1} &= \beta_{2,t} + \sum_{i=1}^4 \gamma_{2,t,i} x_{t,i} + \sum_{i=1}^4 \sum_{j=1}^i \delta_{2,t,i,j} x_{t,i} x_{j,t} + \eta_{2,t} \\
 x_{3,t+1} &= \beta_{3,t} + \sum_{i \in \{1,2,3\}} \gamma_{3,t,i} x_{t,i} + \eta_{3,t} \\
 x_{4,t+1} &= \beta_{4,t} + \sum_{i \in \{1,2,4\}} \gamma_{4,t,i} x_{t,i} + \eta_{4,t}
 \end{aligned} \tag{3.3.1}$$

Where x_1 , x_2 , x_3 , and x_4 are physical capacity, cognitive capacity, exercise, and cognitive stimulation, respectively. β , γ and δ denote the technology parameters to be estimated. η denotes a stochastic shock.

The first two equations in (3.3.1) mean that physical and cognitive capacity follow a flexible functional form containing all lagged factors, their squares, and their interaction terms. This is known as the translog function in the skill formation literature (because skills are typically assumed to be measured in logs, not levels) and has been used by, for example, Agostinelli and Wiswall (2016a). The translog

function allows for dynamic complementarity but does not assume it. While this functional form is not a standard economic production function, we interpret it as a flexible approximation to an arbitrary underlying production function in the spirit of a nonparametric series estimator.

The bottom two equations in (3.3.1) relate to exercise and cognitive stimulations, respectively. Both investment factors are assumed to depend on their own lagged values along with the lagged values of physical and cognitive capacity.

Measurement System

We assume the measurement equations to be linear with an additively separable and normally distributed error term. All of them thus have the following form:

$$y_{\ell,t} = \alpha_{\ell,t} + \sum_{i=1}^4 h_{\ell,t,i} x_{t,i} + \epsilon_{\ell,t} \quad (3.3.2)$$

where $y_{\ell,t}$ denotes the ℓ^{th} measurement in period t , α is the intercept of the measurement equation and h are factor loadings. In the empirical application we only have measurements that load on just one factor, so that for all measurements, three out of the potentially four loadings $h_{\ell,t}$ are zero by construction. Subject to identification requirements outlined in Cunha, Heckman, and Schennach (2010), this could easily be relaxed.

In typical applications of the Technology of Skill Formation, the number and type of available measurement variables varies strongly across periods. This is because any test score that is applicable to very young children would not work for older children. In our case, the measurements stay the same across periods and most of them can be assumed to be time-invariant, i.e. to have the same loading, intercept, and standard deviation of measurement error in each period.

3.3.2 Identification and Interpretation of Parameters

The econometric model implied by the Technology of Skill Formation is a Structural Equation Model or dynamic latent factor model. Linear Structural equation models are widely used since the 1970ies to study relationships between latent and observable variables. However, standard identification results and software for Structural Equation Models are not applicable to our setting because they usually require linearity assumptions or put restrictions on the connectedness of the underlying causal graph, which go beyond those encoded in our system (3.3.1).

Cunha, Heckman, and Schennach (2010) provide general nonparametric identification results for nonlinear dynamic latent factor models. The exact conditions for identification depend on the assumptions one is willing to put on the nature of measurement error. Typically, having at least two dedicated continuous measurements for each latent factor in each period is sufficient to identify an arbitrary production

function under mild conditions. Doing so requires normalizations of location and scale in each period because latent factors do not have a natural unit of measurement.

A subsequent literature (Agostinelli and Wiswall, 2016b; Freyberger, 2021) has shown that much fewer normalizations are required when empirical applications assume the popular constant-elasticity-of-substitution (CES) form, which implies restrictions on the location and scale of its outputs (see Appendix 3.C.1 for details). Our specification of the production function (3.3.1) does not impose any such restrictions. However, as discussed previously, we have at least one age invariant measurement for each latent factor. We always use such measurements for normalizations, which pin down the location and scale of each corresponding factor in all periods.

The lack of natural units for the latent factors and the requirement for normalizations also poses challenges for the interpretation of the results. In short: any outcome that depends on transformations of measurements outside of the model, the choice of the measurement being normalized, or the values of the normalized parameters cannot be interpreted without further information. For details and a more formal definition see Freyberger (2021).

In practical applications, different ways of dealing with this have emerged. Cunha and Heckman (2008) and Cunha, Heckman, and Schennach (2010) propose to anchor the latent factors in terms of observable cardinal variables. For example, they anchor cognitive and non-cognitive skills in terms of years of schooling, wages or the probability to commit a criminal offense. For each anchoring outcome, they re-estimate the model to obtain estimated production function parameters in terms of anchored factors. Attanasio, Meghir, and Nix (2020) do not have access to adult outcomes. Instead they communicate the variables that were normalized and state that results have to be interpreted with respect to the normalizations. Del Bono, Kinsler, and Pavan (2020) propose to simply standardize the variance of the latent factors in logs. This allows for statements such as increasing investment by 1 % increases skills by x %. While this is invariant to any normalization of location and scale in the measurement system, the approach is only valid if one defines that skills are measured in logs not levels. Due to the ordinality of skills, this is a valid but arbitrary definition and thus the approach falls short of its goal to be completely objective. Freyberger (2021) proposes to translate inputs and outputs of the production functions into ranks. This is invariant to any normalization of location and scale, assumptions on whether latent factors are measured in levels or logs and transformations of the measurements outside of the model.

We acknowledge that there is no single natural scale for latent factors and thus see value in all of the above approaches. For example, translating everything to ranks is a natural way of solving a problem that is caused by ordinality. Moreover, it makes the results completely invariant to many decisions made by the econometrician. However, it might not be as interpretable as anchoring approaches. For ex-

ample, it destroys any time trend that was present in the measurements. To address the shortcomings of any single method, we thus use a combination of all of them.

We standardize age invariant measures with respect to their mean and standard deviation at age 68. We estimate the parameters of the production function, normalizing one age-invariant measure for each factor in period zero. The normalized measures are the reversed Frailty Index, Serial 7 Subtraction, Moderate Activity, and Reading. This preserves the time trend in the measurement variables and means that our estimated parameters and the time trend can roughly be interpreted in terms of standard deviations at age 68. For reference, we also show the marginal distributions of each latent factor and the joint distributions of each factor pair at multiple ages (see 3.D.3).

3.3.3 Estimation

Multiple estimators for nonlinear dynamic latent factor models are available. Agostinelli and Wiswall (2016a) estimate the first period factor loadings from ratios of covariances between measurements. To estimate production function parameters, they subsequently employ an iterative IV approach. Their method is very tractable; it comes at the cost of statistical efficiency. Our own experiments on simulated data suggest that it works well for models with few periods but becomes imprecise if there are ten or more periods, especially when the correlation between latent factors is high.

Attanasio, Cunha, and Jervis (2019) use linear regression on Bartlett factor scores with a correction approach. This estimator is computationally very attractive. However, it does not deal well with missing observations. Several of our variables are not contained in the core HRS questionnaire; they are available for subsets of individuals at different points in time. Because of this, the estimator of Attanasio, Cunha, and Jervis (2019) is unsuitable for our application.

Attanasio, Meghir, and Nix (2020) first estimate the distribution of the latent factors as a mixture of normal distributions and then estimate the parameters of the production functions on a simulated sample from that distribution. This approach is computationally harder than the two previous ones but simpler than the maximum likelihood estimator by Cunha, Heckman, and Schennach (2010). The required assumptions are the same as for the likelihood estimator.

Cunha, Heckman, and Schennach (2010) use a maximum likelihood estimator. For computational tractability, they use nonlinear Kalman Filters to factorize the likelihood function into a product of conditional likelihoods. This estimator is computationally more difficult than the others. In its original formulation, numerical stability is often compromised. However, the estimator is statistically efficient and it can deal well with observations that are missing at random.

We derive a mathematically equivalent but numerically stable version of the likelihood estimator used by Cunha, Heckman, and Schennach (2010). Our version

replaces standard filters by square-root Kalman filters (Prvan and Osborne, 1988; van der Merwe and Wan, 2001), which are numerically more robust. The computational cost is similar to the original approach. The details of the original and the reformulated estimator as well as the exact assumptions required for estimation are described in Appendix 3.B.

To account for mortality, we add a dummy variable for being alive as an additional measurement of physical capacity. This is analogous to a linear probability model of survival. Thus, the estimated health state of survivors is adjusted upwards, while the health state of everyone who has passed away is adjusted downwards compared to a state estimation that ignores mortality. In future work we plan to replace the linear probability model of mortality by a Probit model.

A flexible implementation of the new estimator can be found in the Python package `skillmodels` (Gabler, 2022a). It uses JAX (Bradbury, Frostig, Hawkins, Johnson, Leary, et al., 2018) for just in time compilation and automatic differentiation. This reduces the computational cost drastically. We use `estimagic` (Gabler, 2022b) for numerical optimization and the calculation of standard errors. To generate good start values for the optimization, we first decompose the model into four single factor model with much fewer free parameters. In a second step we estimate a linear model. In the third step we estimate the full nonlinear model. We use `pytask` (Raabe, 2020) and the Templates for Reproducible Research Projects in Economics (Gaudecker, 2019) to automate our research project and to parallelize many tasks. The full estimation takes approximately four hours on a laptop.

3.4 Results

We present our results in three stages. First, we describe the measurement system. Next, we describe broad patterns for the transition equations. Finally, we dig deeper into the dynamic effects of changing factors along their distribution.

3.4.1 Measurement System

Table 3.4.1 shows exemplary parameter estimates of the measurement system. The first panel shows the parameters that we constrain to be time-invariant. The three panels below display time-varying parameters of the system at ages 70, 80, and 90. We show loadings and standard deviations for women and men, respectively. Tables 3.D.1–3.D.8 in Appendix 3.D.1 show the complete set of parameter estimates, including the intercepts. Remember from Section 3.2 that we scale all measures—except for dummy measuring vital status, which retains its natural form—to have mean zero and unit variance in the initial period.

For the measurements loading on physical capacity, we normalize the reversed frailty index to have intercept zero and unit loading. We also restrict the parameters relating to mobility, the large muscle index, and self-reported health to be time-

Table 3.4.1. Loadings and Measurement Standard Deviations

Age	Factor	Measurement	Female		Male	
			Loading	Meas. Std.	Loading	Meas. Std.
All	Physical Capacity	Frailty Index (Reversed)	1.000	0.707*** (0.001)	1.000	0.796*** (0.002)
		Mobility	1.228*** (0.005)	0.766*** (0.003)	1.331*** (0.007)	0.750*** (0.003)
		Large Muscle Index	0.929*** (0.005)	0.750*** (0.002)	1.032*** (0.006)	0.761*** (0.003)
		Self-Reported Health	0.950*** (0.004)	0.765*** (0.002)	0.963*** (0.006)	0.793*** (0.003)
	Cognitive Capacity	Serial 7 Subtraction	1.000	0.890*** (0.003)	1.000	0.907*** (0.004)
		Vocabulary	0.839*** (0.013)	0.923*** (0.004)	0.960*** (0.016)	0.900*** (0.004)
		Immediate Word Recall	1.801*** (0.015)	0.583*** (0.003)	1.684*** (0.016)	0.599*** (0.003)
		Delayed Word Recall	1.805*** (0.014)	0.595*** (0.002)	1.648*** (0.015)	0.605*** (0.003)
	Exercise	Vigorous Activity	0.682*** (0.010)	0.809*** (0.004)	0.741*** (0.012)	0.814*** (0.005)
		Moderate Activity	1.000	0.794*** (0.004)	1.000	0.816*** (0.004)
		Light Activity	1.076*** (0.012)	0.933*** (0.004)	0.927*** (0.013)	0.861*** (0.004)
	Cognitive Stimulation	Reading	1.000	0.780*** (0.006)	1.000	0.683*** (0.007)
		Listening to Music	0.512*** (0.010)	0.980*** (0.006)	0.229*** (0.010)	1.004*** (0.007)
		Stimulating Hobbies	0.578*** (0.011)	0.925*** (0.005)	0.375*** (0.012)	0.969*** (0.005)
		Communication	0.523*** (0.010)	0.999*** (0.005)	0.325*** (0.011)	0.989*** (0.006)
	70	Physical Capacity	Alive	0.042*** (0.011)	0.303*** (0.039)	0.058*** (0.013)
Grip Strength			0.489*** (0.042)	0.933*** (0.015)	0.578*** (0.053)	0.978*** (0.020)
	Cognitive Capacity	Self-Rated Memory	0.576*** (0.031)	0.961*** (0.009)	0.626*** (0.035)	0.937*** (0.011)
80	Physical Capacity	Alive	0.091*** (0.023)	0.353*** (0.047)	0.089*** (0.031)	0.367*** (0.068)
		Grip Strength	0.367*** (0.052)	0.882*** (0.021)	0.571*** (0.061)	0.891*** (0.023)
	Cognitive Capacity	Self-Rated Memory	0.470*** (0.038)	1.013*** (0.012)	0.589*** (0.048)	0.988*** (0.015)
90	Physical Capacity	Alive	0.137* (0.081)	0.425*** (0.133)	0.204* (0.121)	0.430*** (0.128)
		Grip Strength	0.357*** (0.099)	0.736*** (0.032)	0.508*** (0.120)	0.766*** (0.055)
	Cognitive Capacity	Self-Rated Memory	0.459*** (0.097)	1.081*** (0.026)	0.386*** (0.120)	1.080*** (0.038)

Note: ***p<0.01,**p<0.05,*p<0.1

invariant – all of these have fairly similar time trends as seen in Figure 3.2.1 (note that mobility has a steeper trend than the others, but making the measurement system time-varying did not change results). All four measurement have similar factor loadings in the 0.93–1.33 range and the standard deviation in their measurement errors is very similar, too (0.75–0.8). The correlations between these four measurements are high throughout in the 0.6–0.85 range (see the correlation matrices in Section 3.D.2 of the Appendix).

We leave the measurement systems for vital status and grip strength unrestricted across age groups. The standard deviation of measurement error in grip strength decreases over time; the loadings decrease for females and stay roughly constant for males. In sum, this means that the correlation between grip strength and the latent factor representing physical capacity stays constant with age for women at 0.3 and increases for men from 0.35 to 0.45. The loading on vital status increases for both genders. Due to the fact that the dummy for being alive has its natural scale, the coefficient has a meaningful interpretation in terms of survival probabilities. At age 70, the interquartile range of physical capacity is 0.95 for women and 0.78 for men (see Appendix Section 3.D.3). Changing physical capacity from its first to its third quartile thus increases the probability of survival by $0.95 \times 4.2\% = 4\%$ for women and $0.78 \times 5.8\% = 4.5\%$ for men. At age 80, the interquartile ranges are just below 1 and the loadings of 0.09 for both genders directly measure changes in survival chances as one moves across the outer quartiles. The same is true at age 90 for men ($\Delta_{\text{survival}} = 0.2$), for women the distribution is less dispersed at that age and an interquartile range of 0.8 implies a increase in survival probabilities of 11%. This is in line with the intuition that physical capacity is more predictive of death at older ages, as deterioration of overall health becomes a more important cause of death than fairly sudden shocks such as cancer or heart attacks (Gill, Gahbauer, Han, and Allore, 2010).

For measures pertaining to cognitive capacity, we normalize the results from the serial 7 subtraction task to have intercept zero and unit loading. This measure along with the vocabulary score and the two word recall tasks are restricted to have the same factor loading and measurement error variance across all ages. Serial 7 subtraction and the vocabulary score look very similar in terms of loading and measurement error. For the word recall tasks, loadings are substantially higher and measurement errors are lower than this. Consequently, all correlations between these measures and the cognitive capacity factor are high throughout – around 0.5 for serial 7 subtraction and vocabulary; exceeding 0.8 for the word recall tasks. The measurement system of self-rated memory is allowed to vary with age. For both genders, its loading is estimated to be about 0.6 initially and decreases over time. The standard deviation of measurement error is around unity, with a slightly increasing trend. Consequently, the correlation of self-rated memory with cognitive capacity is declining with age which is consistent with Huang and Maurer (2019)

Given the similarity of our measurements for exercise, it is unsurprising that all three of them load substantially on the underlying factor. Moderate activity—the normalized measurement—has the largest correlation with the exercise factor at all ages. The correlation of vigorous activity and exercise declines over time whereas light activity goes the other direction. Both of these trends are more pronounced among women than among men.

Among the measurements loading on cognitive stimulation, we normalize the parameters on the time spent reading. This is also the dominant one among the four measurements with a standard deviation of its error around 0.78 (women) and 0.68 (men) and correlations with the factor exceeding 0.7 throughout. The errors on the other three measurements are between 0.9 and 1; their loadings are estimated to be around 0.5 for women and 0.3 for men. For women, these coefficients translates into correlations with the cognitive stimulation factor of around 0.4, which are roughly stable over time. Among men, they start from a level around 0.2-0.3. While communication activities maintain a constant correlation with the factor, listening to music or pursuing stimulating hobbies have hardly any relation left with it by age 90.

In sum, the measurements show a high correlation with the factors they are supposed to identify. For many measurements, it is sensible to restrict the model parameters to be time invariant and we do so. Measurements that are allowed to be changing with age vary in a way that makes sense in the light of prior literature. Differences between genders are not dramatic, but large enough to command separate estimation. Having established these direct relations to the data, we now turn to the core contribution of our paper: The joint evolution of physical and cognitive capacity and the impact of exercise and cognitive stimulation.

3.4.2 Transition Equations

The translog production functions for physical and cognitive capacity have many parameters. In total, we have 15 coefficients per factor, which needs to be multiplied with four age groups (or “stages” in the terminology of Cunha, Heckman, and Schennach, 2010) and two genders. Furthermore, the parameters do not have intuitive interpretations without referring to precise values of the four factors in our model. We thus refrain from listing the parameters in the main text and relegate them to Tables 3.D.9–3.D.16 in Appendix 3.D.4. We note that the vast majority of parameters is very precisely estimated. The set of model parameters is completed with the initial distribution of states and the standard deviation of period-by-period innovations, which we relegate to Appendix 3.D.5.

As a first pass, Figure 3.4.1 shows transition equations for physical capacity (first row of each subfigure referring to women and men, respectively) and cognitive capacity as a function of the input factors. Each of the sixteen panels contains four lines, one for each age group or stage. Input factors are kept at their median except

for the one on the x-axis, which is varied from the 1st to the 99th percentile of its distribution in the respective age group.

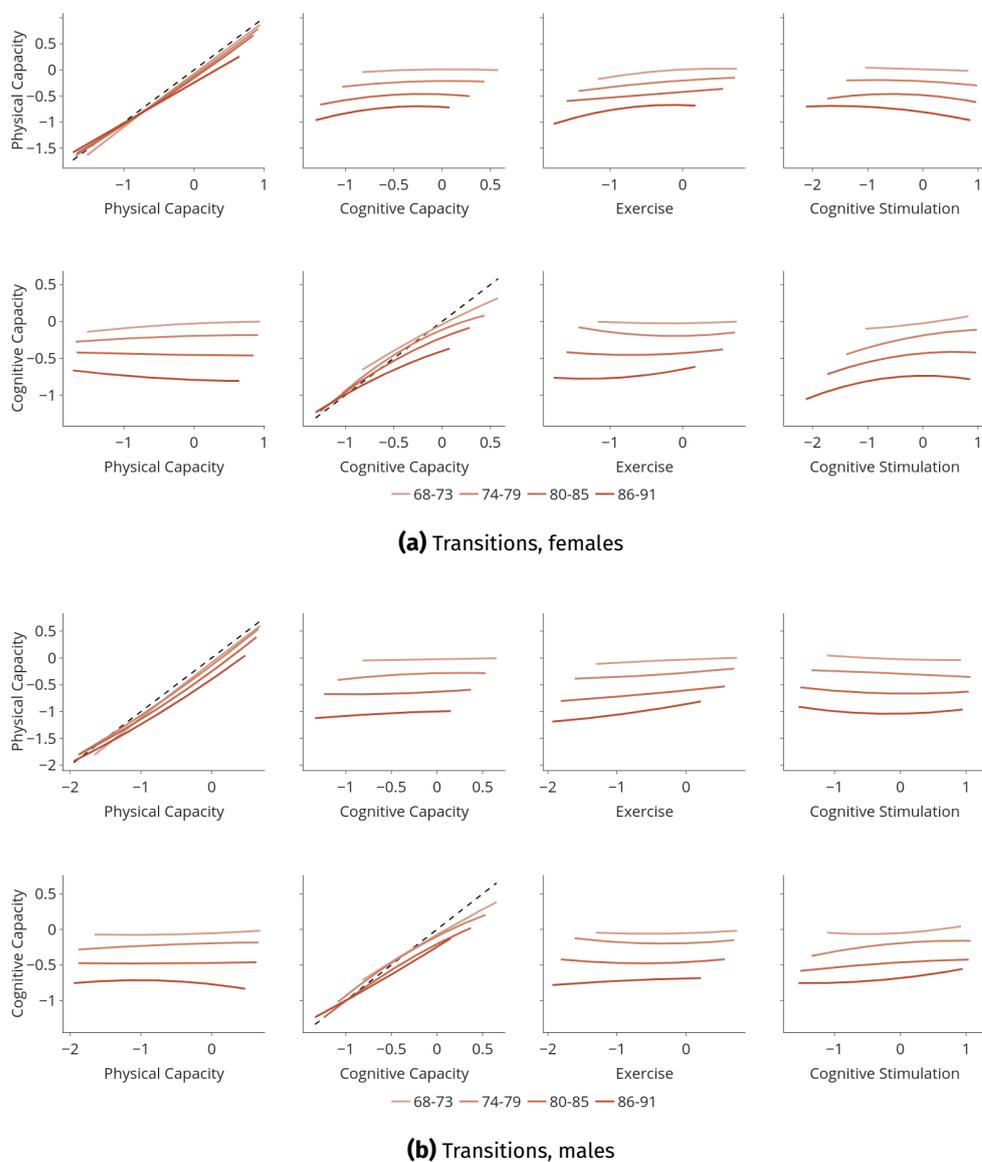


Figure 3.4.1. Next period states as a function of current states, other factors evaluated at the median

The top left panel in Figure 3.4.1a thus shows the result of the following thought experiment: Conditional on current age, what is a woman's expected value of physical capacity in two years as a function of her current physical capacity while fixing cognitive capacity, exercise, and cognitive stimulation at their median values. The results show that there is a high degree of persistence in all age groups. For the upper part of the distribution of physical capacity, the lines are below the 45°-line (the

distributions at ages 70, 80, and 90 are shown in Appendix 3.D.3, Figures 3.D.7–3.D.9; as a rough guide to interpret the first panel of Figure 3.4.1a, the first quartile at age 90 has a value of -1.2). The transition function is below 45° -line everywhere in the youngest age group, which has the steepest slope throughout. This means that at median levels of cognitive capacity, exercise, and cognitive stimulation, physical capacity will unambiguously decline in expectation regardless of the initial level. In contrast, for very low values of physical capacity at older ages, there would be some mean reversion – if all other factors were at their median. Of course, cognitive capacity is not a (direct) choice and there might be substantial costs to reaching median levels of exercise or cognitive stimulation if physical capacity is very low, for example.

Increased cognitive capacity is associated with a slightly more favorable evolution of physical capacity. For example, changing cognitive capacity from its first quartile (-0.63) to its third quartile (-0.17) at age 80 is associated with an increase of age-82 physical capacity of 0.02 units or just under 2 percentiles. The corresponding effects of increased exercise are positive as well and tend to be larger. The same interquartile move for exercise at age 80 (from -0.81 to -0.06) leads to an increase of physical capacity by 0.16 units, which corresponds to almost 5 percentiles. The effects of cognitive stimulation on the dynamics of physical capacity are often slightly negative at median levels of physical capacity, cognitive capacity, and exercise.

The second row of Figure 3.4.1a shows the corresponding effects for the evolution of cognitive capacity. We start with the second panel, which contains the own-effects. They are much less persistent than the own-effects for physical capacity as evident by the flatter slopes at all ages. The four lines are also further apart except at the very bottom of the distribution of cognitive capacity. This means that at almost any level of cognitive capacity, the dynamics are worse for higher ages, provided all other factors are at their median.

The first panel in the second row of Figure 3.4.1a displays modestly positive effects of physical capacity in the lower age groups; these become zero for higher ages and, in the highest age group, turn out to be negative at very low levels of physical capacity. Exercise has mostly positive effects on the evolution of cognitive capacity at median values of other states with an exception being in the lower half of the exercise distribution during women’s upper seventies. Finally, cognitive stimulation has positive effects almost everywhere. Note that the lines are visually misleading to some extent because of the long left tail of cognitive stimulation. For example, the first quartile at age 80 is -0.66 , the slope is steepest to the left of it. Moving cognitive stimulation to its third quartile at 0.14 has hardly any effect.

Figure 3.4.1b shows the same set of transition functions for men. Again, the broad patterns are fairly similar to women, but there are some important differences. For example, physical capacity is deteriorating more quickly for ages 74 and beyond across the entire distribution of current physical capacity; only at the very bottom of the distribution there is some sign of mean reversion. For the own-effects

of physical capacity, there is a similar pattern to what we noted for the own effects of cognitive capacity among women: At almost any level of physical capacity, the dynamics are worse for higher ages, provided all other factors are at their median. In contrast, for cognitive capacity, the same effect is somewhat less pronounced than for women; the lower two age group and the upper two age groups look much more similar to each other there. The signs and magnitudes we noted for the off-diagonal elements generally hold up, although some curvatures appear markedly different. These mostly concern the tails of the distributions, however.

3.4.3 Dynamic Effects Over Several Periods

A major benefit of our dynamic model over multiple periods is that it can be used to evaluate the dynamic effects of interventions through various channels. For example, a positive relation between exercise and cognitive capacity in the cross-section does not mean much because it is not the snapshot that matters, but the history of processes that has led there. In this section, we highlight a few examples of how the distributions of factors change in several years' time when we exogenously manipulate the factors measuring investments, i.e., exercise or cognitive stimulation.

Tables 3.4.2 and 3.4.3 contain the effects of one possible set of such exercises for women and men, respectively. In the baseline scenario, we fix all factors at their age-80 medians. The next row shows the age-86 quantiles the factors are expected to end up at. We then change exercise to its first quartile at age 80, leaving all other factors at their median and letting all of them evolve according to the estimated transition equations until age 86. We repeat this exercise for setting the exercise factor to its third quartile at age 80. The last two panels do the same for cognitive stimulation. We do not take into account that mortality might be affected by the experiment – all effects are conditional on the corresponding individual in the data still being alive at age 86.

The main takeaway from the baseline exercise is that even when fixing everything at the median, there can be large expected changes just a few years down the road. For women, physical capacity is expected to be at the 45th percentile at age 86, whereas cognitive capacity would be expected at its 64th percentile. Hardly any change would be expected in the quantiles of exercise or cognitive stimulation. In stark contrast, for men there would be large drops in the expected quantiles of physical capacity, exercise, and cognitive stimulation along with a tiny drop in the quantile of cognitive capacity.

Due to the high persistence in exercise (for women, see Table 3.D.13 or Figure 3.D.13 in the Appendix; the numbers for men follow directly after those), changing at age 80 essentially means setting it to the same quantile all three periods. Doing so has a large effect on physical capacity (drops by 6-7 percentiles); cognitive capacity and cognitive stimulation barely change. The effect of increasing exercise to its third quartile is almost symmetric for women; the improvement is only three

Table 3.4.2. 6-year-ahead effects of changing exercise or cognitive stimulation, females

Scenario	Age	Physical Capacity	Cognitive Capacity	Exercise	Cognitive Stimulation
Baseline	80	0.50	0.50	0.50	0.50
	86	0.45	0.64	0.49	0.51
Exercise low	80	0.50	0.50	0.25	0.50
	86	0.39	0.63	0.28	0.51
Exercise high	80	0.50	0.50	0.75	0.50
	86	0.50	0.65	0.67	0.52
Cognitive Stimulation low	80	0.50	0.50	0.50	0.25
	86	0.46	0.52	0.48	0.29
Cognitive Stimulation high	80	0.50	0.50	0.50	0.75
	86	0.40	0.71	0.49	0.72

Table 3.4.3. 6-year-ahead effects of changing exercise or cognitive stimulation, males

Scenario	Age	Physical Capacity	Cognitive Capacity	Exercise	Cognitive Stimulation
Baseline	80	0.50	0.50	0.50	0.50
	86	0.34	0.48	0.39	0.38
Exercise low	80	0.50	0.50	0.25	0.50
	86	0.27	0.49	0.23	0.37
Exercise high	80	0.50	0.50	0.75	0.50
	86	0.37	0.50	0.54	0.39
Cognitive Stimulation low	80	0.50	0.50	0.50	0.25
	86	0.35	0.43	0.39	0.19
Cognitive Stimulation high	80	0.50	0.50	0.50	0.75
	86	0.33	0.53	0.40	0.60

percentiles for men. Note that, by age 86, exercise has reverted to its 54th percentile, so less of an effect might be expected, too.

Fixing cognitive stimulation at its first quartile reduces cognitive capacity by 12 percentiles for women and by 5 percentiles for men. Interestingly, the larger effect for women occurs despite the fact that at age 86, cognitive stimulation is expected to be at its 29th percentile for women compared to the 19th percentile for men. Conversely, increasing cognitive stimulation substantially improves cognition for both genders. It also has a detrimental effect on women's health whereas there is no effect for men.

3.5 Conclusions and Outlook

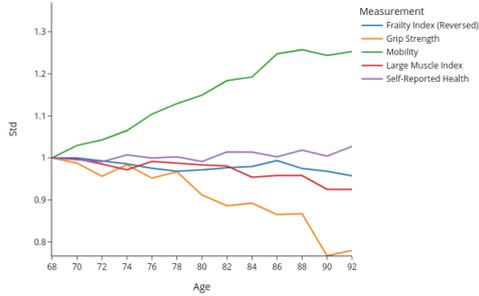
We adapt a nonlinear dynamic latent factor framework that was developed for skill formation of children to study the physical and cognitive decline between ages 68 and 93. To this end, we incorporate mortality into the model. The model is estimated with a rich set of measures from the Health and Retirement Study.

We document a large amount of measurement error in all observed variables. While most measurements have a high correlation with the latent factor they measure, no single measurement is a good enough proxy to use in isolation. A dynamic latent factor model is therefore a good fit for this setting. Having a rich set of time invariant measurements for each latent factor, lets us overcome some of the challenges related to the interpretability of latent factors. To make our results even more interpretable we also present them in terms of population ranks and use survival probabilities to anchor physical capacity.

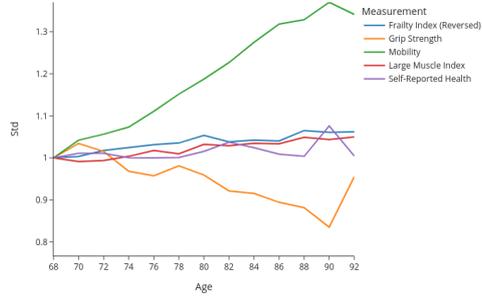
We find that, despite a strong decline in means for physical and cognitive capacity, the rank order of these latent factors is remarkably stable over periods. Nevertheless, physical and cognitive capacity can be influenced by investments until very high ages. Cognitive stimulation is a specific investment into cognitive capacity. Physical exercise has a larger effect on physical capacity and a small effect on cognitive capacity.

We leave a few extensions of our approach for future work. Besides expanding the sampling period by another wave, we want to add mental health as a separate latent factor that is different from cognitive and physical capacity but can influence both. As a robustness check we want to replace the linear probability model of mortality by a probit model. This requires the addition of nonlinear measurement equations to the model. To address any concerns of endogeneity of investments, we will use a control function approach similar to recent skill formation papers (Agostinelli and Wiswall, 2016a; Attanasio, Meghir, and Nix, 2020) as endogeneity correction. Finally we will use the model to simulate the effect of different investment policies and use Shapley decompositions to attribute the dynamic of investments over multiple periods to different channels of transmission.

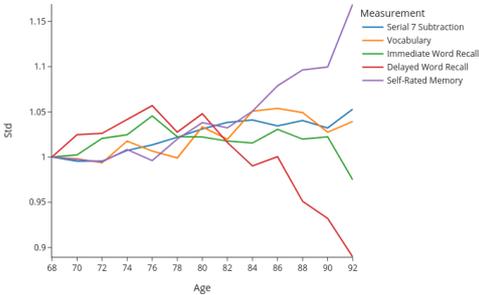
Appendix 3.A Additional Background on the Data and Measurements



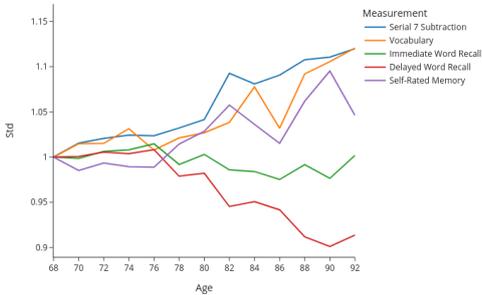
(a) Physical capacity, females



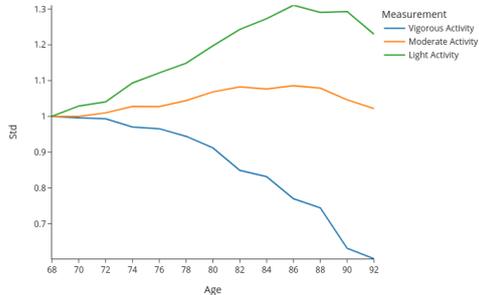
(b) Physical capacity, males



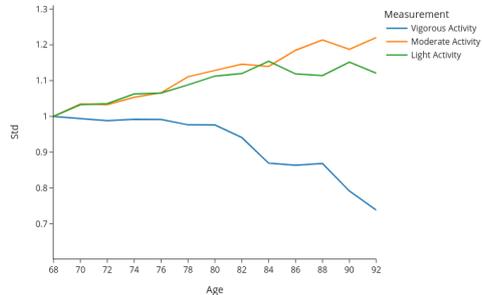
(c) Cognitive capacity, females



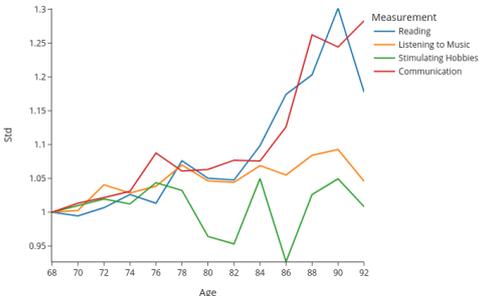
(d) Cognitive capacity, males



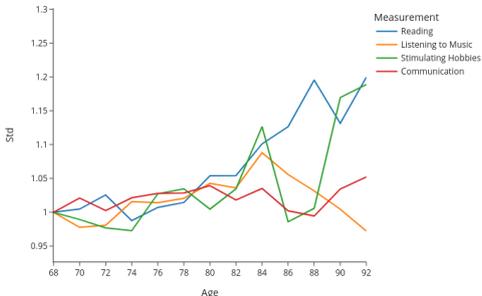
(e) Exercise, females



(f) Exercise, males



(g) Cognitive Stimulation, females



(h) Cognitive Stimulation, males

Figure 3.A.1. Standard deviation of measurements by age

Appendix 3.B The Maximum Likelihood Estimator

3.B.1 State Estimation

3.B.1.1 Preliminaries

To discuss the econometric approach used in this paper and potential alternatives it is convenient to express the model in state space notation.

To do so, let $\mathbf{x}_t \in \mathcal{R}^N$ denote the vector of latent factors (i.e. physical capacity, cognitive capacity, physical exercise and cognitive stimulation) in period t .

Similarly, let $\mathbf{y}_t \in \mathcal{R}^{L_t}$ denote the vector of all observable measurements in period t .

Then the transition function of the latent factors can be written as:

$$\mathbf{x}_{t+1} = F_t(\mathbf{x}_t) + \boldsymbol{\eta}_t \quad (3.B.1)$$

where $\boldsymbol{\eta}_t$ is a vector of error terms with η_t^j on the j^{th} position. Let \mathbf{Q}_t denote the covariance matrix of $\boldsymbol{\eta}_t$

The linear measurement system can be written as:

$$\mathbf{y}_t = \mathbf{H}_t \mathbf{x}_t + \boldsymbol{\epsilon}_t \quad (3.B.2)$$

where \mathbf{H}_t is a matrix of coefficients known as factor loadings and $\boldsymbol{\epsilon}_t$ is a vector of measurement errors with $\epsilon_{t,l}$ on the l^{th} position. Let \mathbf{R}_t denote the covariance matrix of $\boldsymbol{\epsilon}_t$.

Equations 3.B.1 and 3.B.2 define a state space model. Equation 3.B.1 is called transition equation. Equation 3.B.2 is called measurement equation. The vector \mathbf{x}_t is called the state of the system. The matrices \mathbf{Q}_t and \mathbf{R}_t are called process noise and measurement noise, respectively.

To see why it was handy to rewrite the technology of skill formation in state form, assume for a moment that the transition function F_t (including parameters) as well as the matrices \mathbf{H}_t , \mathbf{Q}_t and \mathbf{R}_t are known for all $t \in T$ but the state vectors \mathbf{x}_t are unknown and have to be estimated from measurements \mathbf{y}_t . This problem is known as optimal state estimation, which is a well researched topic in physics and engineering.

To efficiently estimate the state vector in period t , an estimator should not only use measurements from this period, but also take the information from all previous measurements into account. For linear systems, Kalman filters are the method of choice for state estimation (Kalman, 1960). For nonlinear systems, several nonlinear variants of the Kalman filter have been developed. Kalman filters treat the state of a system itself as random vector. Therefore, they are sometimes classified as Bayesian filters.

Kalman filters consist of a predict and an update step. They are initialised with an initial estimate for the mean $\bar{\mathbf{x}}_0$ and covariance matrix \mathbf{P}_0 of the distribution of the state vector. Then, in each period, the new measurements are incorporated to update the mean and covariance matrix of the state vector. After that, the transition equation is used to predict the mean and covariance matrix of the state vector in the next period. This predicted state vector can then again be updated with measurements.

For the application of Kalman filters, the following assumptions must hold:

1. $\eta_t \sim \mathcal{N}(\mathbf{0}_N, \mathbf{Q}_t)$ where $\mathbf{0}_N$ denotes a vector of zeros of length N , \mathbf{Q}_t is a diagonal matrix.
2. The η_t^j are serially independent over all t .
3. $\epsilon_t \sim \mathcal{N}(\mathbf{0}_{L_t}, \mathbf{R}_t)$ where \mathbf{R}_t is a diagonal matrix.
4. The $\epsilon_{t,l}$ are serially independent over all t .
5. $\epsilon_{t,l}$ and η_t^j are independent of \mathbf{x}_t for all $t = 1, \dots, T$, $l = 1, \dots, L$ and each factor j .
6. The distribution of the state vector $p(\mathbf{x}_t)$ can be approximated by a mixture of normal distributions for all $t = 1, \dots, T$

Due to the assumption of a linear measurement system, the state vector can be estimated by combining the update step of a linear Kalman filter with the predict step of a nonlinear Kalman filter. For computational reasons, it will be convenient not to incorporate all measurements at once but to perform a separate update step for each measurement.

3.B.1.2 The Update Step of the Kalman Filter

The aim of the Kalman update is to efficiently combine information from measurements in the current period with previous measurements. To do so, the measurement function is used to convert the pre-update state vector into predicted measurements for the current period (equation 3.B.3). The difference between the predicted and actual measurements is called residual (equation 3.B.4). This residual, scaled by the so called Kalman gain, is then added to the pre-update state vector (equation 3.B.8). The Kalman gain is smaller if the variance of the measurement (calculated by equation 3.B.6) is large. This has the intuitive consequence that noisy measurements receive a low weight. The Kalman gain becomes larger if the pre-update covariance matrix has large diagonal entries (equation 3.B.5 and 3.B.7). Thus, measurements receive more weight if the pre-update state is known imprecisely due to bad initial values or a high process noise, for example. After the incorporation of the measurements, the state is always known with the same or more precision than before. This is reflected by subtracting a positive semi-definite matrix from the pre-update covariance matrix (equation 3.B.9).

Let $\bar{\mathbf{x}}_{t|y_{t,l}^-}$ denote the mean of the conditional distribution of the state vector given all measurements up to but not including the l^{th} measurement in period t . Let $\mathbf{P}_{t|y_{t,l}^-}$ denote the covariance matrix of this distribution. Let $\mathbf{h}_{t,l}$ denote the l^{th} row of \mathbf{H}_t . Let $r_{t,l,l}$ be the l^{th} diagonal element of \mathbf{R}_t . The update step that incorporates the l^{th} measurement into the estimate is given by the following equations:

$$\bar{y}_{t,l|y_{t,l}^-} = \mathbf{h}_{t,l}\bar{\mathbf{x}}_{t|y_{t,l}^-} \qquad \bar{y}_{t,l|y_{t,l}^-} = E(y_{t,l}|y_{t,l}^-) \quad (3.B.3)$$

$$\delta_{t,l} = y_{t,l} - \bar{y}_{t,l|y_{t,l}^-} \qquad \delta_{t,l} \text{ can be interpreted as residual} \quad (3.B.4)$$

$$\mathbf{f}_{t,l} = \mathbf{P}_{t|y_{t,l}^-}\mathbf{h}_{t,l}^T \qquad \mathbf{f}_{t,l} \text{ is an intermediate result} \quad (3.B.5)$$

$$\sigma_{t,l} = \mathbf{h}_{t,l}\mathbf{f}_{t,l} + r_{t,l,l} \qquad \sigma_{t,l} \text{ is the variance of } y_{t,l} \quad (3.B.6)$$

$$\mathbf{k}_{t,l} = \frac{1}{\sigma_{t,l}}\mathbf{f}_{t,l} \qquad \mathbf{k}_{t,l} \text{ is the (scaled) Kalman gain} \quad (3.B.7)$$

$$\bar{\mathbf{x}}_{t|y_{t,l}} = \bar{\mathbf{x}}_{t|y_{t,l}^-} + \mathbf{k}_{t,l}\delta_{t,l} \qquad \bar{\mathbf{x}}_{t|y_{t,l}} \text{ is the updated mean} \quad (3.B.8)$$

$$\mathbf{P}_{t|y_{t,l}} = \mathbf{P}_{t|y_{t,l}^-} - \frac{1}{\sigma_{t,l}}\mathbf{f}_{t,l}\mathbf{f}_{t,l}^T \qquad \mathbf{P}_{t|y_{t,l}} \text{ is the updated covariance matrix} \quad (3.B.9)$$

3.B.1.3 The Predict Step of the Kalman Filter

In linear systems, the mean and covariance matrix of the system can be propagated to the next period by simply applying the linear transition equation. With a nonlinear transition function, however, this is not possible, as $E(f(X)) \neq f(E(X))$ in general. For the nonlinear predict step, two basic options exist: The *extended Kalman filter* and the *unscented Kalman filter*. Cunha, Heckman and Schennach choose the unscented Kalman filter because it has been shown to be more reliable in a wide range of settings (Van Der Merwe, 2004).

The intuition of the predict step of the unscented Kalman filter is relatively simple: firstly, a deterministic sample of points in the state space, called sigma points (equation 3.B.10), and accompanying weights are chosen (equation 3.B.11). Usually these are $2N + 1$ points and weights, where N is the length of the state vector. Secondly, these sigma points are transformed using the true nonlinear transition equation. Thirdly, the weighted sample mean is used as estimate for the next period mean of the state vector (equation 3.B.12). Fourthly, the sum of the covariance matrix of the process noise and the weighted sample covariance of the transformed sigma points is used as estimate of the covariance matrix of the state vector (equation 3.B.13). Intuitively, the addition of the process noise accounts for the fact that the prediction always adds some uncertainty about the state of the system.

For the choice of sigma points and sigma weights, many different algorithms exist. All have in common that some form of matrix square root of the covariance

matrix of the state vector is taken. Two definitions of matrix square root exist: 1) A is a matrix square root of P if $P = AA$. 2) A is a matrix square root of P if $P = AA^T$. The matrix square root is not unique in general and some matrices do not have a square root. However, all symmetric positive semi-definite matrices, i.e. all valid covariance matrices, can be decomposed into $P = LL^T$ where L is lower triangular (Zhang, 1999). For the unscented Kalman filter, both definitions of matrix square root work. Below, the sigma point algorithm proposed by Julier and Uhlmann (1997), is presented without reference to a particular type of matrix square root:

Let $\kappa \in \mathbb{R}$ be a scaling parameter. Usually, κ is set to 2 if the distribution of the state vector is assumed to be normal. Let $P_{t|t}$ denote the covariance matrix of the state vector, conditional on all measurements up to and including period t . Define $S_{t|t} \equiv \sqrt{P_{t|t}}$ as the matrix square root of $P_{t|t}$ and let $s_{t,n}$ denote its n^{th} column.

Sigma points are calculated according to the following equations:

$$\begin{aligned} \chi_{t,n} &= \bar{x}_{t|t} && \text{for } n = 0 \\ \chi_{t,n} &= \bar{x}_{t|t} + \sqrt{N + \kappa} s_{t,n} && \text{for } n = 1, \dots, N \\ \chi_{t,n} &= \bar{x}_{t|t} - \sqrt{N + \kappa} s_{t,n} && \text{for } n = N + 1, \dots, 2N \end{aligned} \quad (3.B.10)$$

where $\chi_{t,n}$ is the n^{th} sigma point at period t that is calculated after incorporating all measurements of that period. The corresponding sigma weights are calculated as follows:

$$\begin{aligned} w_{t,n} &= \frac{\kappa}{N + \kappa} && \text{for } n = 0 \\ w_{t,n} &= \frac{1}{2(N + \kappa)} && \text{for } n = 1, \dots, 2N \end{aligned} \quad (3.B.11)$$

where $w_{t,n}$ is the n^{th} sigma weight. Define $\tilde{\chi}_{t,n} \equiv F_t(\chi_{t,n})$ where $F_t(\cdot)$ is defined as in equation 3.B.1. Then the predict step of the unscented Kalman filter is given by:

$$\bar{x}_{t+1|t} = \sum_{n=0}^{2N} w_{t,n} \tilde{\chi}_{t,n} \quad (3.B.12)$$

$$P_{t+1|t} = \left[\sum_{n=0}^{2N} w_{t,n} (\tilde{\chi}_{t,n} - \bar{x}_{t+1|t})(\tilde{\chi}_{t,n} - \bar{x}_{t+1|t})^T \right] + Q_t \quad (3.B.13)$$

3.B.2 The Likelihood Interpretation of the Kalman Filter

Of course, the parameters of the function F_t and the matrices H_t , Q_t and R_t are unknown in reality. However, they can be estimated by maximum likelihood. The direct maximization of the likelihood function would involve the evaluation of high dimensional integrals which is computationally very expensive (Cunha, Heckman, and Schennach, 2010). Instead, Kalman filters can be used to reduce the number of computations required for each evaluation of the likelihood function dramatically.

To see how, define θ as the vector with all estimated parameters of the model. Then, the likelihood contribution of individual i is given by:

$$\mathcal{L}(\theta | y_1, \dots, y_T) \equiv p_\theta(y_1, \dots, y_T) = \prod_{t=1}^T \prod_{l=1}^{L_t} p_\theta(y_{t,l} | y_{t,l}^-) \quad (3.B.14)$$

where $p_\theta(y_1, \dots, y_T)$ denotes the joint density of all measurements for individual i , conditional on the parameter vector θ and $p_\theta(y_{t,l} | y_{t,l}^-)$ is the density of the l^{th} measurement in period t , given all measurements up to but not including this measurement. The subscript i is again omitted for readability.

To see how this relates to the Kalman filter, recall that for each $t = 1, \dots, T$ and each $l = 1, \dots, L_t$, equation 3.B.3 calculates $\bar{y}_{t,l} | y_{t,l}^-$, i.e. the expected value of the l^{th} measurement in period t , conditional on all previous measurements. In addition, due to the normality and independence assumptions on the error terms and the factor distribution, $y_{t,l}$ is normally distributed around $\bar{y}_{t,l} | y_{t,l}^-$. Equation 3.B.6 can be used to calculate the variance $\sigma_{t,l}$ of this distribution. Thus, $p_\theta(y_{t,l} | y_{t,l}^-) = \phi_{\bar{y}_{t,l} | y_{t,l}^-, \sigma_{t,l}}(y_{t,l})$ where $\phi_{\mu, \sigma}(\cdot)$ is the density of a normal random variable with mean μ and variance σ .

A nice feature of the estimator based on this factorization of the likelihood function is that it can deal very well with missing observations. If measurement $y_{t,l}$ is missing for individual i , the corresponding update of the state vector is just skipped. More formally, this means that the missing measurement is integrated out from the likelihood function.

3.B.3 Numerical Stability

3.B.3.1 Numerical Challenges

While the Kalman filter based maximum likelihood estimator is statistically and computationally efficient, it is numerically unstable. The numerical instability caused by floating point imprecision is inherent to Kalman filters and has been discovered soon after Kalman published his original article. Since then, the precision of computers has increased enormously such that nowadays numerical problems are not a big issue for well specified Kalman filters. However, during the maximization of the like-

likelihood function the optimizer might pick parameter combinations that are far from leading to a well specified filter.

The numerical problems manifest themselves in two places:

1. In the update step, the subtraction in equation 3.B.9 can lead to negative diagonal elements in the updated covariance matrix of the state vector. While this is mathematically impossible in a well specified Kalman filter, numerical imprecisions and badly specified Kalman filters during the maximization process make it possible.
2. Even if the covariance matrix of the state vector has nonnegative diagonal entries, numerical imprecisions might render it not positive semi-definite. With this the existence of a matrix square root is not guaranteed, which can make the calculation of sigma points impossible.

Cunha, Heckman and Schennach mention the numerical problems in their supplementary material. To solve the first problem, they recommend to find good initial values for the maximization by first constraining some parameters and letting the code find good initial values for the others. For the second problem, they propose to set all off-diagonal elements of \mathbf{P} to zero before taking the square root, which then corresponds to taking the element wise square root of the diagonal elements. While this prevents the estimator from crashing, it is not standard practice in Kalman filtering and it is not guaranteed that an estimator based on this type of matrix square root produces reliable results.

3.B.3.2 Outline of the Solution

A better approach is to use a square root implementation of the Kalman filter. Many different square root Kalman filters exist. They are mathematically equivalent to normal Kalman filters but numerically more stable.

Instead of propagating the full covariance matrix of the state vector, square root Kalman filters propagate the square root of this matrix. This has three advantages:

1. It avoids overflow errors due to numbers with very small or large absolute values, as taking the square root makes large numbers smaller and small numbers larger.
2. By using a matrix square root \mathbf{A} of the type $\mathbf{P} = \mathbf{A}\mathbf{A}^T$, the problematic covariance matrix is guaranteed to be positive semi-definite (Zhang, 1999), i.e. a valid covariance matrix. In particular, its diagonal entries are sums of squared terms and, consequently, guaranteed to be nonnegative. This solves the first problem.
3. By choosing an appropriate pair of square root update and predict algorithms, taking matrix square roots can be completely avoided. This eliminates the second problem.

The computational requirements of square root filters are comparable to those of normal Kalman filters. In the nonlinear case, they are even lower. For a maximally

robust estimator, we use a pair of square root update and predict algorithms that completely avoid taking matrix square roots. The algorithm for the update was developed by Prvan and Osborne (Prvan and Osborne, 1988). The unscented square root predict step was proposed by Van Der Merwe and Wan (van der Merwe and Wan, 2001). Both propagate the transpose of a lower triangular matrix square root of the state covariance matrix.

3.B.3.3 The QR Decomposition of a Matrix

Both square root algorithms rely on a matrix factorization called QR decomposition. Note that in this subsection, \mathbf{Q} and \mathbf{R} do not denote the covariance matrices of the process and measurement noise but factors into which a matrix is decomposed.

\mathbf{QR} is called QR decomposition of an $m \times n$ matrix \mathbf{A} with $m \geq n$ if:

1. $\mathbf{A} = \mathbf{QR}$
2. \mathbf{Q} is an orthogonal $m \times m$ matrix
3. \mathbf{R} is an $m \times n$ matrix and the first n rows of \mathbf{R} form an upper triangular matrix and its remaining rows only contain zeros

The QR decomposition of a matrix always exists but is not unique. A useful property of the QR decomposition is that:

$$\mathbf{A}^T \mathbf{A} = (\mathbf{QR})^T \mathbf{QR} = \mathbf{R}^T \mathbf{Q}^T \mathbf{QR} = \mathbf{R}^T \mathbf{R} \quad (3.B.15)$$

where the last equality comes from the defining property of orthogonal matrices that $\mathbf{Q}^T \mathbf{Q} = \mathbf{Q} \mathbf{Q}^T = \mathbf{I}$, where \mathbf{I} denotes the identity matrix. Thus, the upper triangular part of \mathbf{R} is the transpose of a lower triangular matrix square root of $\mathbf{A}^T \mathbf{A}$. For convenience, let $qr(\mathbf{A})$ denote the QR decomposition of \mathbf{A} that only returns the upper triangular part of the matrix \mathbf{R} .

3.B.3.4 The Update Step of the Square-Root Kalman Filter

Let $\mathbf{S}_{t|y_{t,l}^-}$ be a lower triangular matrix square root of $\mathbf{P}_{t|y_{t,l}^-}$ and keep the rest of the notation as in section 3.B.1. Then, the square root update that incorporates the l^{th} measurement in period t is given by the following equations:

$\tilde{\mathbf{y}}_{t,l|y_{t,l}^-}$ and $\delta_{t,l}$ are calculated as in equation 3.B.3 and 3.B.4 respectively. Then the following intermediate results are calculated.

$$\mathbf{f}_{t,l}^* = \mathbf{S}_{t|y_{t,l}^-}^T \mathbf{h}_{t,l}^T \quad (3.B.16)$$

$$\mathbf{M}_{t,l} = \begin{bmatrix} \sqrt{r_{t,l,l}} & \mathbf{0}_N^T \\ \mathbf{f}_{t,l}^* & \mathbf{S}_{t|y_{t,l}^-}^T \end{bmatrix} \quad (3.B.17)$$

It can be shown that:

$$qr(\mathbf{M}_{t,l}) = \begin{bmatrix} \sqrt{\sigma_{t,l}} & \frac{1}{\sqrt{\sigma_{t,l}}} \mathbf{f}_{t,l}^T \\ \mathbf{0}_N & \mathbf{S}_{t|y_{t,l}}^T \end{bmatrix} \quad (3.B.18)$$

where $\mathbf{S}_{t|y_{t,l}}^T$ is the transpose of a lower triangular square root of the updated covariance matrix and $\mathbf{0}_N$ denotes a column vector of length N that is filled with zeros.

The matrix in equation 3.B.18 also contains $\mathbf{f}_{t,l}$ and $\sigma_{t,l}$ such that the Kalman gain can be calculated as in equation 3.B.7 and the mean of the state vector can be updated as in equation 3.B.8.

To see why equation 3.B.18 holds, define $\mathbf{U}_{t,l} \equiv qr(\mathbf{M}_{t,l})$ and partition it as follows:

$$\mathbf{U}_{t,l} = \begin{bmatrix} \mathbf{U}_{1,1} & \mathbf{U}_{1,2} \\ \mathbf{0} & \mathbf{U}_{2,2} \end{bmatrix} \quad (3.B.19)$$

where $\mathbf{U}_{1,1}$ is a scalar, $\mathbf{U}_{1,2}$ a row vector of length N , $\mathbf{0}$ a column vector of length N filled with zeros and $\mathbf{U}_{2,2}$ an upper triangular $N \times N$ matrix. Recall from the definition of $\mathbf{U}_{t,l}$ and equation 3.B.15 that $\mathbf{U}_{t,l}^T \mathbf{U}_{t,l} = \mathbf{M}_{t,l}^T \mathbf{M}_{t,l}$. Multiplying out both sides of this equality yields:

$$\begin{bmatrix} r_{t,l,l} + \mathbf{f}_{t,l}^{*T} \mathbf{f}_{t,l}^* & \mathbf{f}_{t,l}^{*T} \mathbf{S}_{t|y_{t,l}^-}^T \\ \mathbf{S}_{t|y_{t,l}^-} \mathbf{f}_{t,l}^* & \mathbf{S}_{t|y_{t,l}^-} \mathbf{S}_{t|y_{t,l}^-}^T \end{bmatrix} = \begin{bmatrix} \mathbf{U}_{1,1}^2 & \mathbf{U}_{1,1} \mathbf{U}_{1,2} \\ \mathbf{U}_{1,2}^T \mathbf{U}_{1,1} & \mathbf{U}_{1,2}^T \mathbf{U}_{1,2} + \mathbf{U}_{2,2}^T \mathbf{U}_{2,2} \end{bmatrix} \quad (3.B.20)$$

It is obvious from equation 3.B.6 and 3.B.16 that $\mathbf{U}_{1,1} = \sqrt{\sigma_{t,l}}$. Using this and noting that $\mathbf{f}_{t,l}^{*T} \mathbf{S}_{t|y_{t,l}^-}^T = \mathbf{f}_{t,l}^T$, where $\mathbf{f}_{t,l}$ is defined as in equation 3.B.5, one obtains that:

$$\mathbf{U}_{1,2} = \frac{\mathbf{f}_{t,l}^T}{\sqrt{\sigma_{t,l}}} \quad (3.B.21)$$

It remains to show that $\mathbf{U}_{2,2} = \mathbf{S}_{t|y_{t,l}}^T$. By noting that the bottom right element of the left hand side of equation 3.B.20 is, by definition, equal to the pre-update covariance matrix $\mathbf{P}_{t|y_{t,l}^-}$ and plugging in the value for $\mathbf{U}_{1,2}$, one obtains that:

$$\mathbf{U}_{2,2}^T \mathbf{U}_{2,2} = \mathbf{P}_{t|y_{t,l}^-} - \frac{1}{\sigma_{t,l}} \mathbf{f}_{t,l} \mathbf{f}_{t,l}^T = \mathbf{P}_{t|y_{t,l}} \quad (3.B.22)$$

where the last equality comes from equation 3.B.9. Thus $\mathbf{U}_{2,2}^T$ is a matrix square root of $\mathbf{P}_{t|y_{t,l}}$ and by the definition of the QR decomposition it is lower triangular, which completes the proof. Importantly, no part of the proof requires the lower triangular square roots of $\mathbf{P}_{t|y_{t,l}^-}$ or $\mathbf{P}_{t|y_{t,l}}$ to be unique or makes reference to a specific type of matrix square root.

3.B.3.5 The Predict Step of the Square-Root Kalman Filter

For the square root implementation of the unscented predict step in period t , firstly the sigma points are calculated as in equation 3.B.10, where this time $S_{t|t}$ is required to be a lower triangular matrix square root of $P_{t|t}$. Again, $\tilde{\mathcal{X}}_t$ denotes the $(2N + 1) \times N$ matrix of the transformed sigma points. The calculation of the predicted mean of the state vector remains the same as before (equation 3.B.12).

Define A_t as stacked matrix of weighted deviations of the sigma points from the predicted mean and the covariance matrix of the transition shocks:

$$A_t \equiv \begin{bmatrix} \sqrt{w_{t,0}}(\tilde{\mathcal{X}}_{t,0} - \bar{\mathbf{x}}_{t+1|t})^T \\ \cdots \\ \sqrt{w_{t,2n}}(\tilde{\mathcal{X}}_{t,2n} - \bar{\mathbf{x}}_{t+1|t})^T \\ \sqrt{Q_t} \end{bmatrix} \quad (3.B.23)$$

Then equation 3.B.13 can be rewritten as:

$$P_{t+1|t} = A_t^T A_t \quad (3.B.24)$$

and by the relation of the QR decomposition and the lower triangular matrix square root (equation 3.B.15) a lower triangular matrix square root of $P_{t+1|t}$ is given by $qr(A_t)^T$.

Appendix 3.C Detailed Model Setup

3.C.1 Background on Identification

Cunha, Heckman, and Schennach (2010) provide very general nonparametric Identification result for nonlinear dynamic latent factor models. The exact conditions for identification depend on the assumptions one is willing to put on the measurement error. However, having at least two dedicated measurements for each latent factor in each period is sufficient to identify an arbitrary production function under mild conditions. Since latent factors do not have a natural unit of measurement, the identification requires normalizations of location and scale. Thus, Cunha, Heckman, and Schennach (2010) normalize one loading of each factor in each period to 1 and one intercept of each factor in each period to 0. While the identification result works for arbitrary production functions, they use a parametric CES function in their empirical application.

Agostinelli and Wiswall (2016b) criticize the identification result by Cunha, Heckman, and Schennach (2010) to be flawed. They point out that the CES production function already puts a restriction on the scale and location of its output. Thus, normalization of scale and location are only required in the first period and

re-normalizations in each period are actually not normalizations but testable assumptions. Moreover, they show that under the implicit restrictions imposed by the CES production function, identification under a linear measurement system can be achieved with as little as one measurement per latent factor and period as long as there are at least two measurements in the first period.

Freyberger (2021) shows that the CES production function also imposes implicit restrictions on the relative scale of the latent factors and thus identification can be achieved if only the location and scale of a single factor are normalized in the first period.

While the critique by Agostinelli and Wiswall (2016b) that over-normalizations are detrimental is correct, it mostly applies to the empirical application and not the general identification result in Cunha, Heckman, and Schennach (2010) nor the maximum likelihood estimator used in the paper. The identification result states that latent factors have no natural scale and location that could be identified from data and thus their location and scale has to be fixed by restrictions imposed by the econometrician. Cunha, Heckman, and Schennach (2010) restrict factor loadings and intercepts but mention, that instead of factor loadings, the variances of measurement errors could be restricted. Of course, these restrictions are mutually exclusive and it would not be valid to restrict factor loadings and variances of measurement error at the same time. The main contribution of Agostinelli and Wiswall (2016b) is to point out that using restrictive functional forms for the production function is yet another way of fixing the location and scale of the latent factors.

Appendix 3.D Additional Tables and Figures for the Main Specification

3.D.1 Complete Set of Parameters of the Measurement System

Table 3.D.1. Intercepts, Loadings, and Measurement Standard Deviations for Physical Capacity, Females

		Intercept	Loading	Meas. Std.
All	Frailty Index (Reversed)	0.000	1.000	0.707*** (0.001)
	Mobility	-0.113*** (0.003)	1.228*** (0.005)	0.766*** (0.003)
	Large Muscle Index	0.005* (0.003)	0.929*** (0.005)	0.750*** (0.002)
	Self-Reported Health	-0.048*** (0.003)	0.950*** (0.004)	0.765*** (0.002)
70	Alive	0.897*** (0.103)	0.042*** (0.011)	0.303*** (0.039)
	Grip Strength	-0.126*** (0.027)	0.489*** (0.042)	0.933*** (0.015)
72	Alive	0.909*** (0.107)	0.045*** (0.011)	0.288*** (0.038)
	Grip Strength	-0.241*** (0.028)	0.396*** (0.042)	0.922*** (0.016)
74	Alive	0.902*** (0.097)	0.060*** (0.013)	0.301*** (0.036)
	Grip Strength	-0.292*** (0.030)	0.466*** (0.043)	0.935*** (0.018)
76	Alive	0.885*** (0.101)	0.073*** (0.018)	0.327*** (0.043)
	Grip Strength	-0.471*** (0.030)	0.368*** (0.049)	0.924*** (0.012)
78	Alive	0.879*** (0.103)	0.075*** (0.019)	0.339*** (0.046)
	Grip Strength	-0.540*** (0.033)	0.447*** (0.048)	0.924*** (0.019)
80	Alive	0.871*** (0.097)	0.091*** (0.023)	0.353*** (0.047)
	Grip Strength	-0.758*** (0.034)	0.367*** (0.052)	0.882*** (0.021)
82	Alive	0.870*** (0.112)	0.089*** (0.026)	0.359*** (0.054)
	Grip Strength	-0.789*** (0.037)	0.336*** (0.055)	0.861*** (0.020)
84	Alive	0.869*** (0.105)	0.110*** (0.030)	0.371*** (0.053)
	Grip Strength	-0.980*** (0.042)	0.334*** (0.061)	0.866*** (0.026)
86	Alive	0.856*** (0.122)	0.123*** (0.040)	0.391*** (0.067)
	Grip Strength	-0.997*** (0.046)	0.337*** (0.071)	0.839*** (0.028)
88	Alive	0.846*** (0.146)	0.129** (0.053)	0.406*** (0.086)
	Grip Strength	-1.191*** (0.060)	0.413*** (0.084)	0.827*** (0.035)
90	Alive	0.828*** (0.203)	0.137* (0.081)	0.425*** (0.133)
	Grip Strength	-1.105*** (0.062)	0.357*** (0.099)	0.736*** (0.032)
92	Alive	0.819*** (0.215)	0.168 (0.116)	0.443*** (0.148)
	Grip Strength	-1.362*** (0.083)	0.350*** (0.116)	0.746*** (0.048)

Note: ***p<0.01;**p<0.05;*p<0.1

Table 3.D.2. Intercepts, Loadings, and Measurement Standard Deviations for Physical Capacity, Males

		Intercept	Loading	Meas. Std.
All	Frailty Index (Reversed)	0.000	1.000	0.796 ^{***} (0.002)
	Mobility	-0.015 ^{***} (0.005)	1.331 ^{***} (0.007)	0.750 ^{***} (0.003)
	Large Muscle Index	0.042 ^{***} (0.004)	1.032 ^{***} (0.006)	0.761 ^{***} (0.003)
	Self-Reported Health	0.026 ^{***} (0.003)	0.963 ^{***} (0.006)	0.793 ^{***} (0.003)
70	Alive	0.901 ^{***} (0.092)	0.058 ^{***} (0.013)	0.303 ^{***} (0.035)
	Grip Strength	-0.056 (0.034)	0.578 ^{***} (0.053)	0.978 ^{***} (0.020)
72	Alive	0.907 ^{***} (0.083)	0.075 ^{***} (0.015)	0.298 ^{***} (0.030)
	Grip Strength	-0.294 ^{***} (0.034)	0.550 ^{***} (0.053)	0.959 ^{***} (0.020)
74	Alive	0.900 ^{***} (0.119)	0.061 ^{***} (0.017)	0.310 ^{***} (0.046)
	Grip Strength	-0.318 ^{***} (0.035)	0.497 ^{***} (0.057)	0.922 ^{***} (0.021)
76	Alive	0.876 ^{***} (0.129)	0.073 ^{***} (0.024)	0.344 ^{***} (0.059)
	Grip Strength	-0.506 ^{***} (0.037)	0.559 ^{***} (0.057)	0.898 ^{***} (0.020)
78	Alive	0.872 ^{***} (0.130)	0.081 ^{***} (0.027)	0.355 ^{***} (0.062)
	Grip Strength	-0.560 ^{***} (0.041)	0.553 ^{***} (0.059)	0.920 ^{***} (0.023)
80	Alive	0.866 ^{***} (0.135)	0.089 ^{***} (0.031)	0.367 ^{***} (0.068)
	Grip Strength	-0.737 ^{***} (0.043)	0.571 ^{***} (0.061)	0.891 ^{***} (0.023)
82	Alive	0.852 ^{***} (0.117)	0.136 ^{***} (0.043)	0.394 ^{***} (0.066)
	Grip Strength	-0.959 ^{***} (0.046)	0.468 ^{***} (0.065)	0.872 ^{***} (0.025)
84	Alive	0.868 ^{***} (0.130)	0.139 ^{***} (0.047)	0.387 ^{***} (0.068)
	Grip Strength	-1.042 ^{***} (0.052)	0.557 ^{***} (0.069)	0.842 ^{***} (0.027)
86	Alive	0.849 ^{***} (0.157)	0.140 ^{**} (0.062)	0.408 ^{***} (0.092)
	Grip Strength	-1.238 ^{***} (0.064)	0.488 ^{***} (0.085)	0.841 ^{***} (0.034)
88	Alive	0.856 ^{***} (0.154)	0.176 ^{**} (0.074)	0.418 ^{***} (0.090)
	Grip Strength	-1.283 ^{***} (0.071)	0.471 ^{***} (0.109)	0.826 ^{***} (0.045)
90	Alive	0.850 ^{***} (0.212)	0.204 [*] (0.121)	0.430 ^{***} (0.128)
	Grip Strength	-1.358 ^{***} (0.102)	0.508 ^{***} (0.120)	0.766 ^{***} (0.055)
92	Alive	0.765 ^{**} (0.312)	0.183 (0.218)	0.464 [*] (0.268)
	Grip Strength	-1.493 ^{***} (0.123)	0.684 ^{***} (0.166)	0.816 ^{***} (0.077)

Note: ***p<0.01;**p<0.05;*p<0.1

Table 3.D.3. Intercepts, Loadings, and Measurement Standard Deviations for Cognitive Capacity, Females

		Intercept	Loading	Meas. Std.
All	Serial 7 Subtraction	0.000	1.000	0.890*** (0.003)
	Vocabulary	0.043*** (0.006)	0.839*** (0.013)	0.923*** (0.004)
	Immediate Word Recall	-0.161*** (0.006)	1.801*** (0.015)	0.583*** (0.003)
	Delayed Word Recall	-0.189*** (0.006)	1.805*** (0.014)	0.595*** (0.002)
70	Self-Rated Memory	0.005 (0.014)	0.576*** (0.031)	0.961*** (0.009)
72	Self-Rated Memory	0.029** (0.015)	0.593*** (0.030)	0.955*** (0.009)
74	Self-Rated Memory	0.016 (0.015)	0.555*** (0.030)	0.973*** (0.009)
76	Self-Rated Memory	0.028* (0.017)	0.497*** (0.033)	0.968*** (0.010)
78	Self-Rated Memory	0.045** (0.019)	0.501*** (0.035)	0.992*** (0.011)
80	Self-Rated Memory	0.052** (0.022)	0.470*** (0.038)	1.013*** (0.012)
82	Self-Rated Memory	0.069** (0.027)	0.460*** (0.043)	1.010*** (0.013)
84	Self-Rated Memory	0.083** (0.032)	0.398*** (0.050)	1.035*** (0.015)
86	Self-Rated Memory	0.079* (0.041)	0.393*** (0.058)	1.063*** (0.018)
88	Self-Rated Memory	0.261*** (0.055)	0.549*** (0.075)	1.069*** (0.021)
90	Self-Rated Memory	0.210*** (0.074)	0.459*** (0.097)	1.081*** (0.026)
92	Self-Rated Memory	0.218** (0.110)	0.538*** (0.133)	1.145*** (0.040)
<i>Note:</i>			***p<0.01; **p<0.05; *p<0.1	

Table 3.D.4. Intercepts, Loadings, and Measurement Standard Deviations for Cognitive Capacity, Males

		Intercept	Loading	Meas. Std.
All	Serial 7 Subtraction	0.000	1.000	0.907*** (0.004)
	Vocabulary	0.048*** (0.008)	0.960*** (0.016)	0.900*** (0.004)
	Immediate Word Recall	-0.183*** (0.008)	1.684*** (0.016)	0.599*** (0.003)
	Delayed Word Recall	-0.200*** (0.008)	1.648*** (0.015)	0.605*** (0.003)
70	Self-Rated Memory	-0.041** (0.017)	0.626*** (0.035)	0.937*** (0.011)
72	Self-Rated Memory	-0.052*** (0.017)	0.560*** (0.034)	0.955*** (0.011)
74	Self-Rated Memory	-0.043** (0.017)	0.573*** (0.035)	0.949*** (0.011)
76	Self-Rated Memory	-0.039** (0.020)	0.527*** (0.040)	0.955*** (0.012)
78	Self-Rated Memory	-0.051** (0.022)	0.607*** (0.043)	0.972*** (0.013)
80	Self-Rated Memory	-0.002 (0.026)	0.589*** (0.048)	0.988*** (0.015)
82	Self-Rated Memory	-0.019 (0.034)	0.479*** (0.057)	1.033*** (0.018)
84	Self-Rated Memory	-0.019 (0.040)	0.520*** (0.063)	1.007*** (0.020)
86	Self-Rated Memory	-0.019 (0.046)	0.464*** (0.071)	0.992*** (0.022)
88	Self-Rated Memory	0.007 (0.065)	0.509*** (0.091)	1.035*** (0.028)
90	Self-Rated Memory	0.011 (0.089)	0.386*** (0.120)	1.080*** (0.038)
92	Self-Rated Memory	0.003 (0.125)	0.599*** (0.182)	1.011*** (0.049)

Note:

***p<0.01; **p<0.05; *p<0.1

Table 3.D.5. Intercepts, Loadings, and Measurement Standard Deviations for Exercise, Females

		Intercept	Loading	Meas. Std.
All	Vigorous Activity	−0.009 (0.006)	0.682*** (0.010)	0.809*** (0.004)
	Moderate Activity	0.000	1.000	0.794*** (0.004)
	Light Activity	−0.127*** (0.007)	1.076*** (0.012)	0.933*** (0.004)
<i>Note:</i>			***p<0.01; **p<0.05; *p<0.1	

Table 3.D.6. Intercepts, Loadings, and Measurement Standard Deviations for Exercise, Males

		Intercept	Loading	Meas. Std.
All	Vigorous Activity	−0.012** (0.006)	0.741*** (0.012)	0.814*** (0.005)
	Moderate Activity	0.000	1.000	0.816*** (0.004)
	Light Activity	−0.077*** (0.007)	0.927*** (0.013)	0.861*** (0.004)
<i>Note:</i>			***p<0.01; **p<0.05; *p<0.1	

Table 3.D.7. Intercepts, Loadings, and Measurement Standard Deviations for Cognitive Stimulation, Females

		Intercept	Loading	Meas. Std.
All	Reading	0.000	1.000	0.780*** (0.006)
	Listening to Music	-0.168*** (0.006)	0.512*** (0.010)	0.980*** (0.006)
	Stimulating Hobbies	-0.069*** (0.007)	0.578*** (0.011)	0.925*** (0.005)
	Communication	-0.062*** (0.006)	0.523*** (0.010)	0.999*** (0.005)
<i>Note:</i>		***p<0.01,**p<0.05,*p<0.1		

Table 3.D.8. Intercepts, Loadings, and Measurement Standard Deviations for Cognitive Stimulation, Males

		Intercept	Loading	Meas. Std.
All	Reading	0.000	1.000	0.683*** (0.007)
	Listening to Music	-0.175*** (0.007)	0.229*** (0.010)	1.004*** (0.007)
	Stimulating Hobbies	-0.012 (0.009)	0.375*** (0.012)	0.969*** (0.005)
	Communication	-0.083*** (0.007)	0.325*** (0.011)	0.989*** (0.006)
<i>Note:</i>		***p<0.01,**p<0.05,*p<0.1		

3.D.2 Correlations Between Measurements and Factors

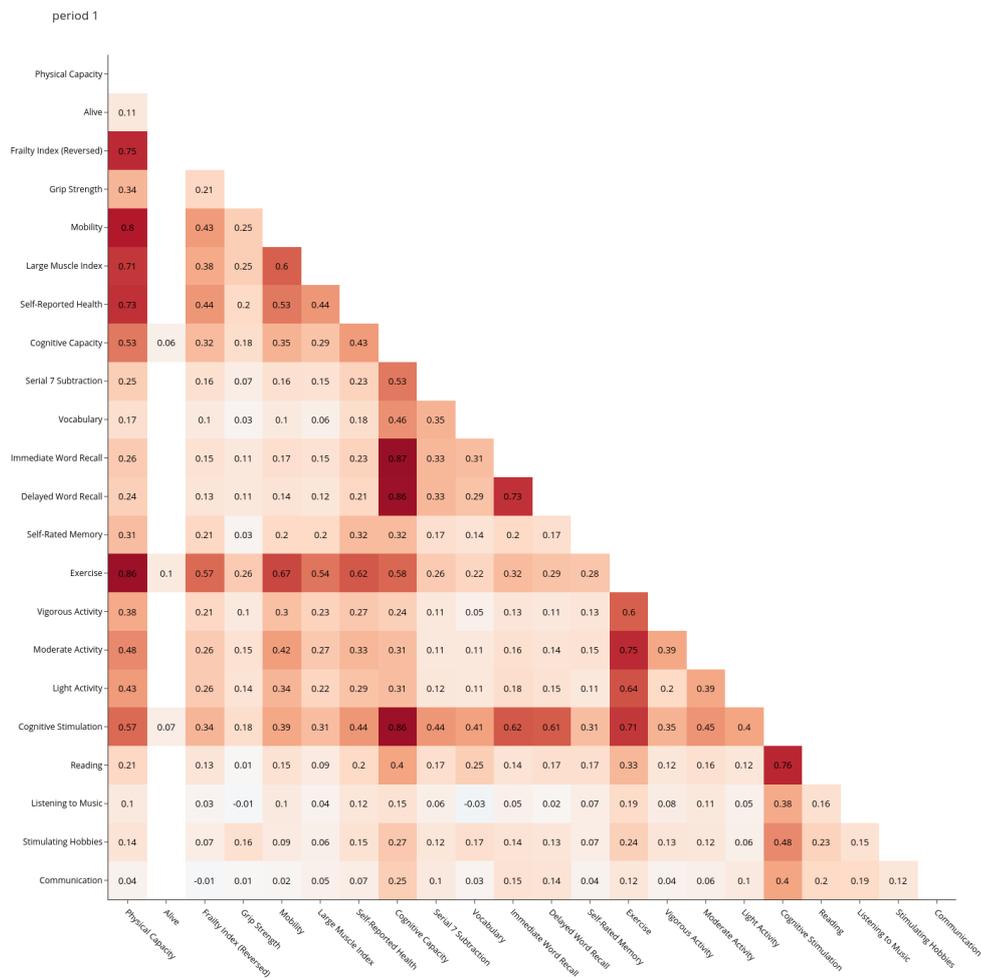


Figure 3.D.1. Correlations across implied factors and measurement correlations – females aged 70

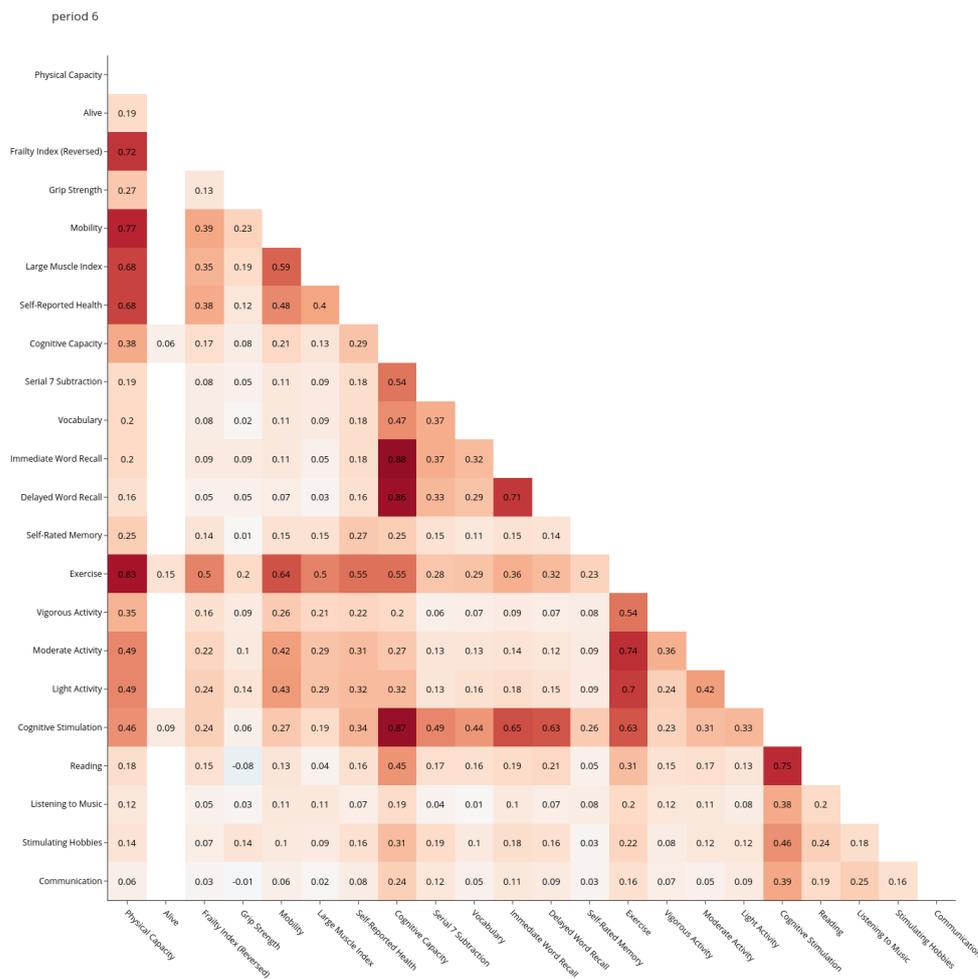


Figure 3.D.2. Correlations across implied factors and measurement correlations – females aged 80

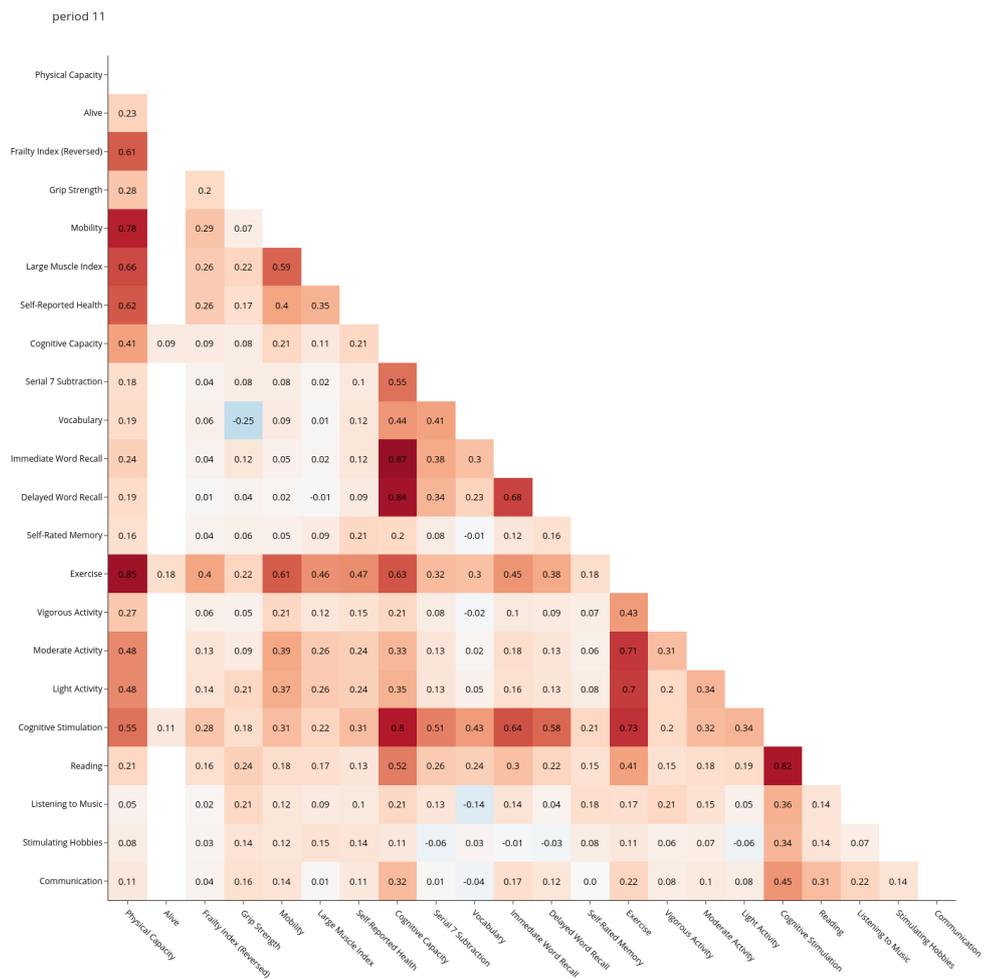


Figure 3.D.3. Correlations across implied factors and measurement correlations – females aged 90

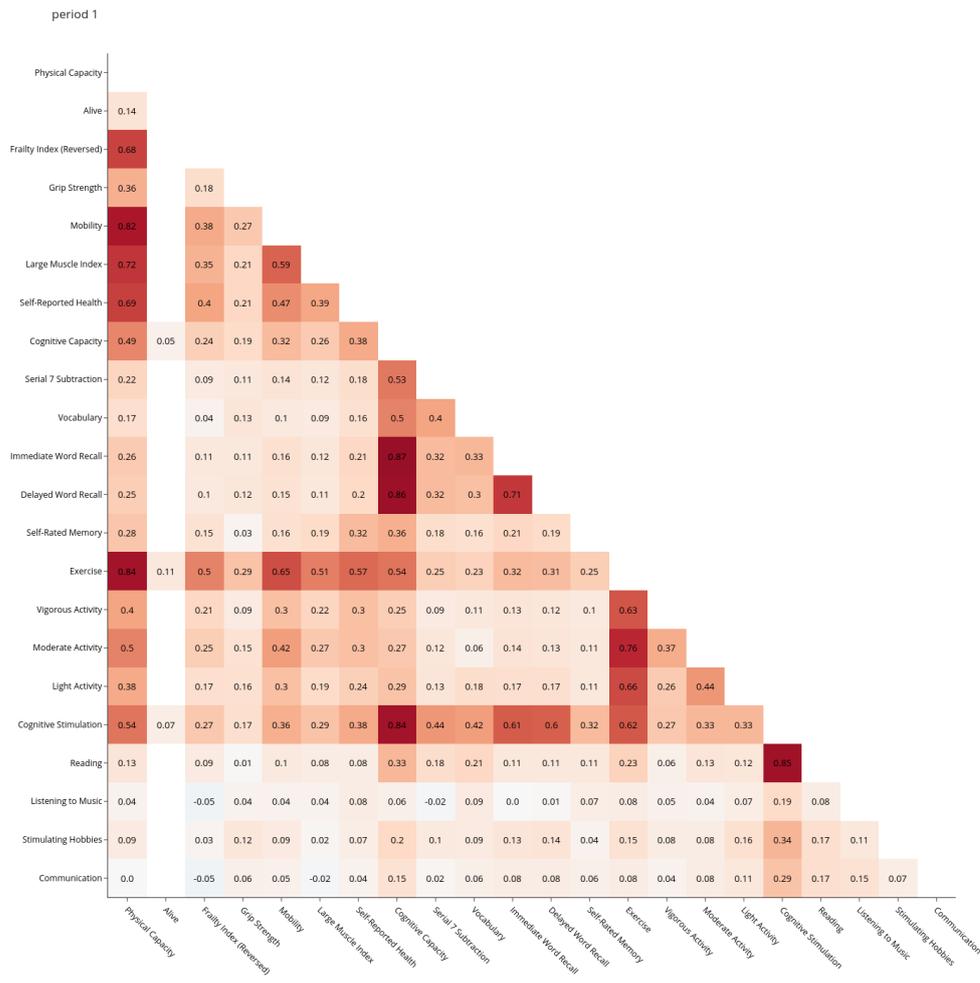


Figure 3.D.4. Correlations across implied factors and measurement correlations – males aged 70

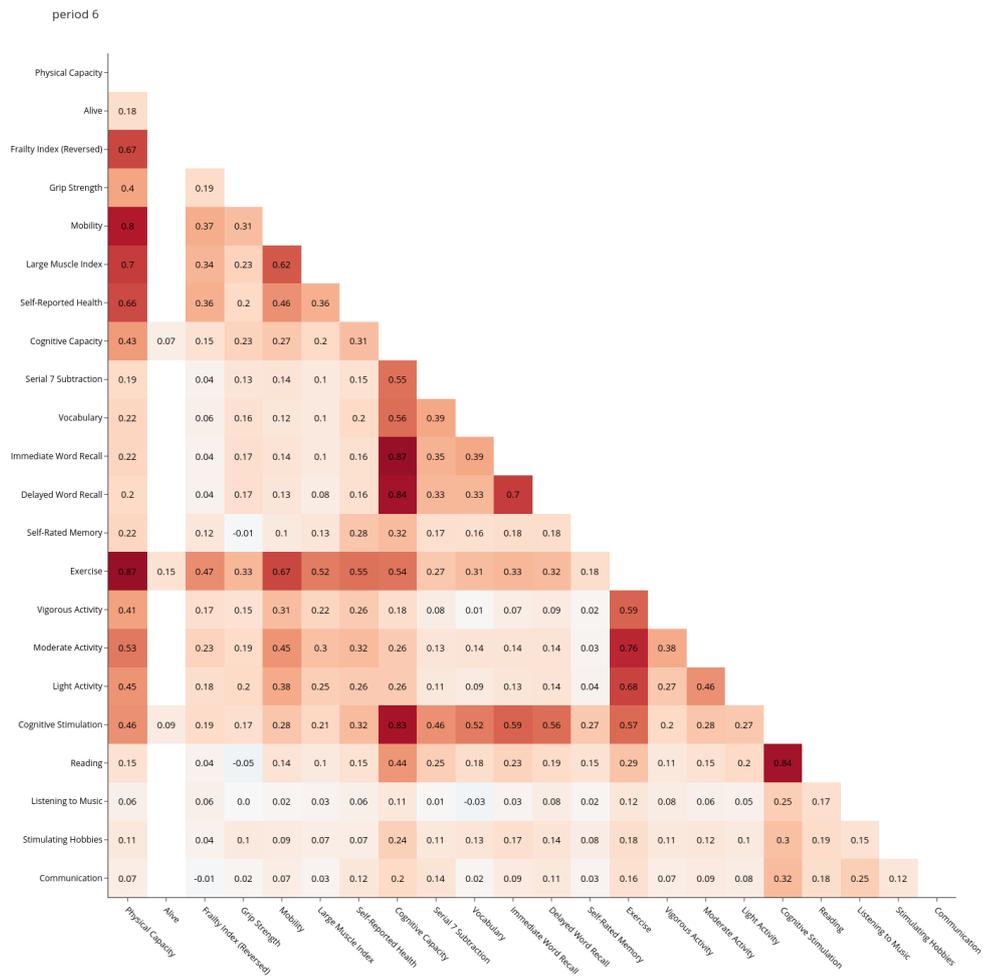


Figure 3.D.5. Correlations across implied factors and measurement correlations – males aged 80

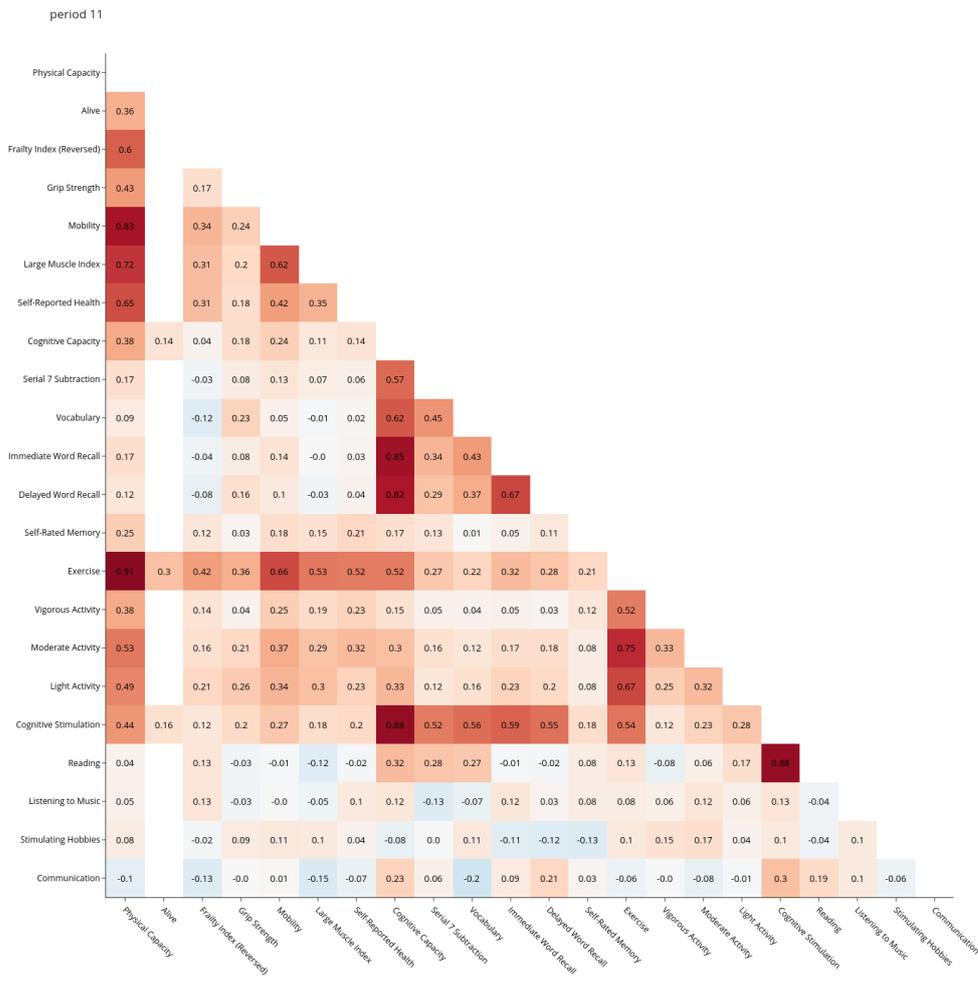


Figure 3.D.6. Correlations across implied factors and measurement correlations – males aged 90

3.D.3 Factor Distributions

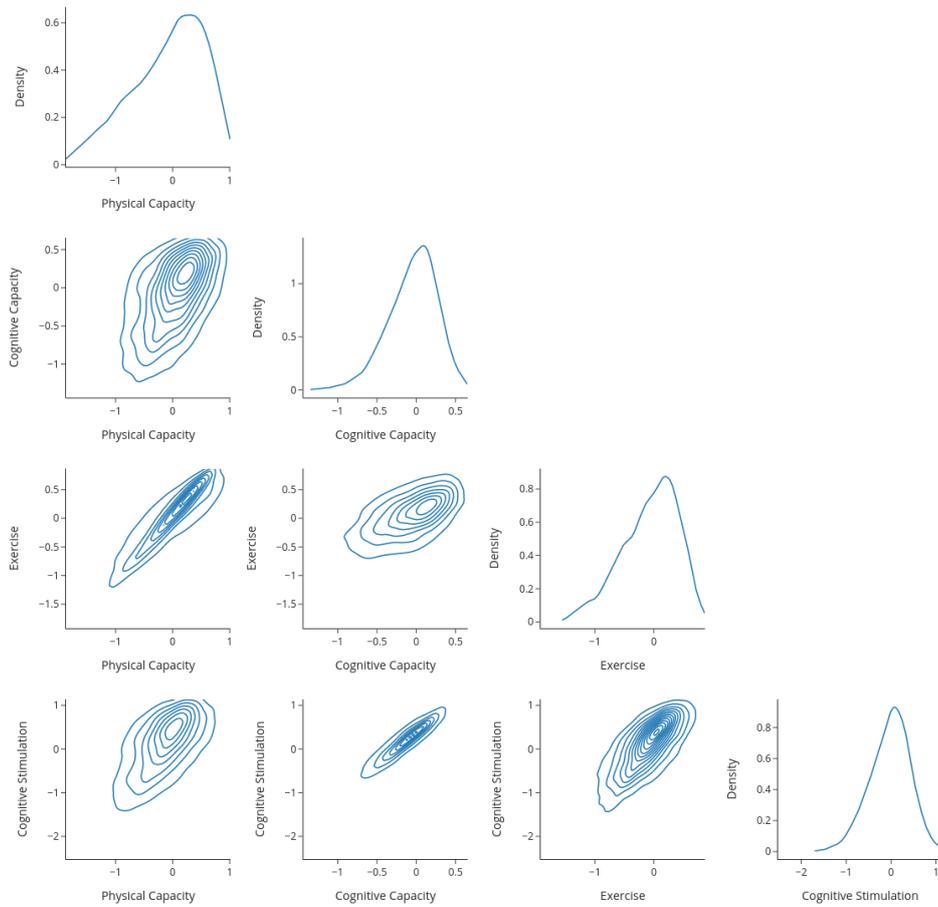


Figure 3.D.7. Factor distributions – females aged 70

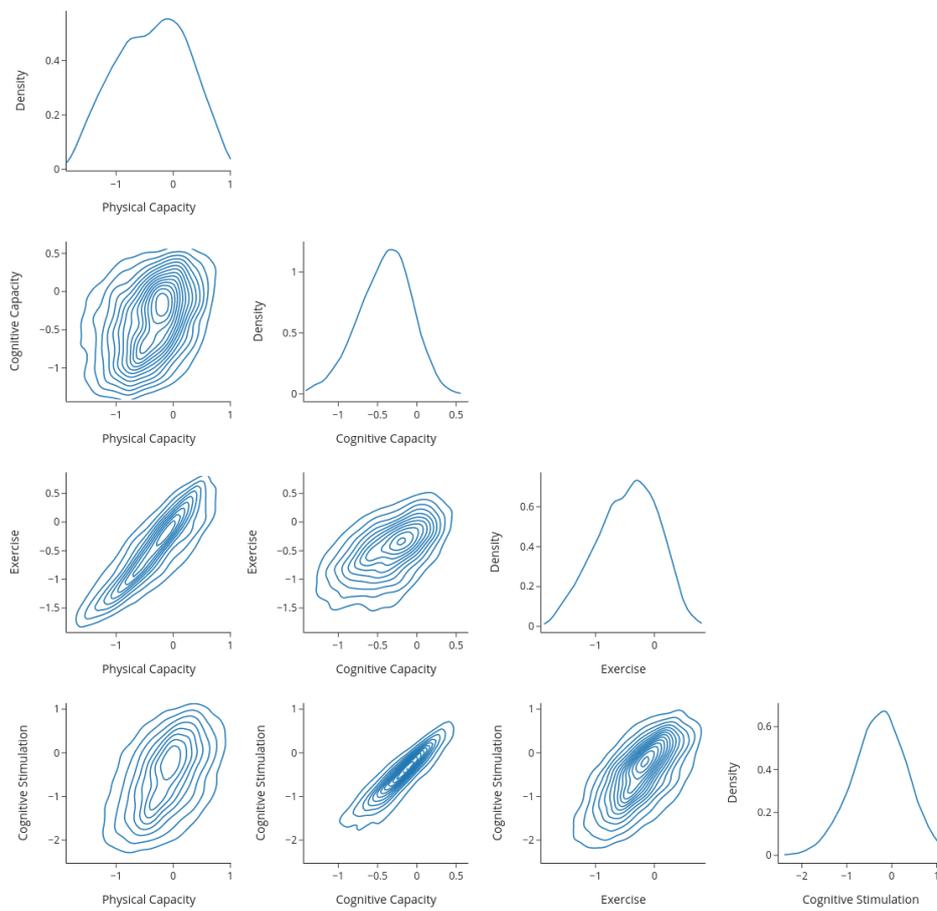


Figure 3.D.8. Factor distributions – females aged 80

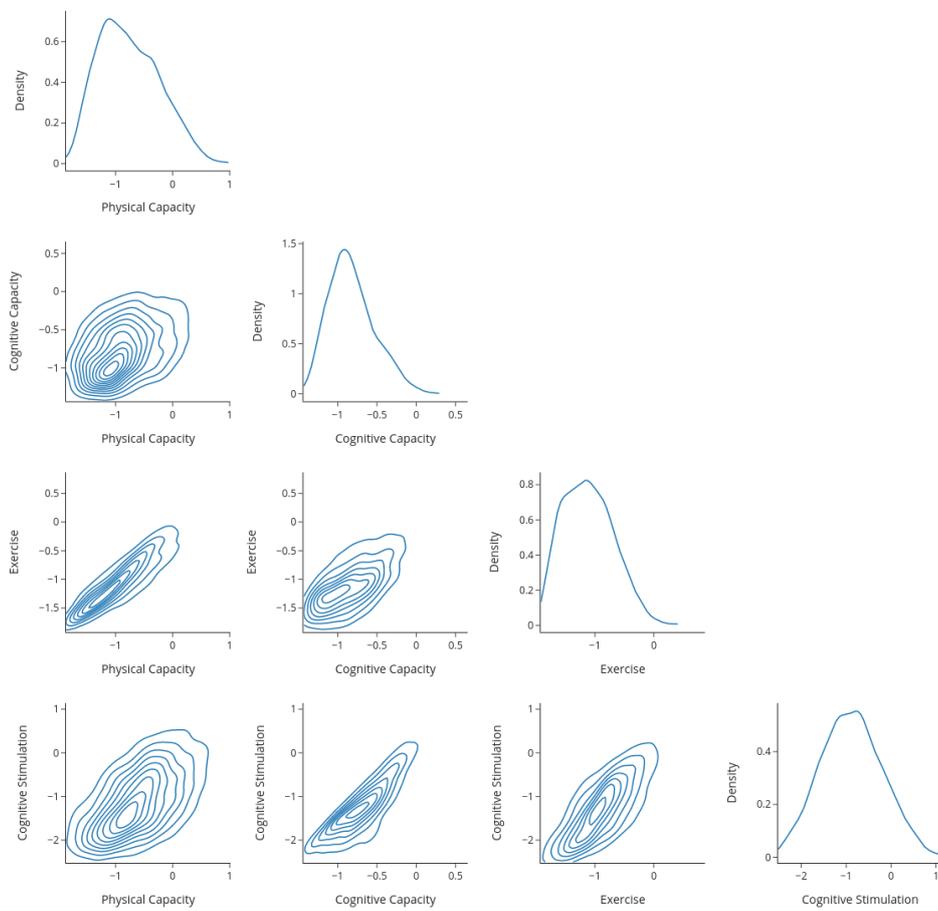


Figure 3.D.9. Factor distributions – females aged 90

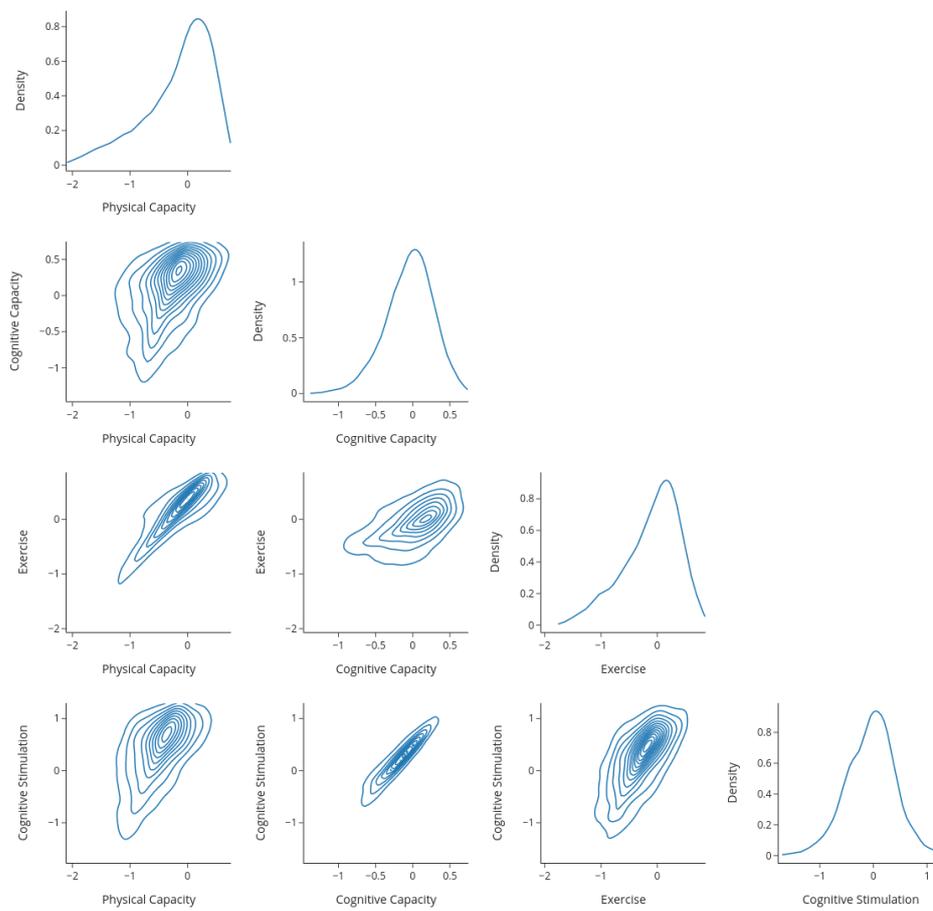


Figure 3.D.10. Factor distributions – males aged 70

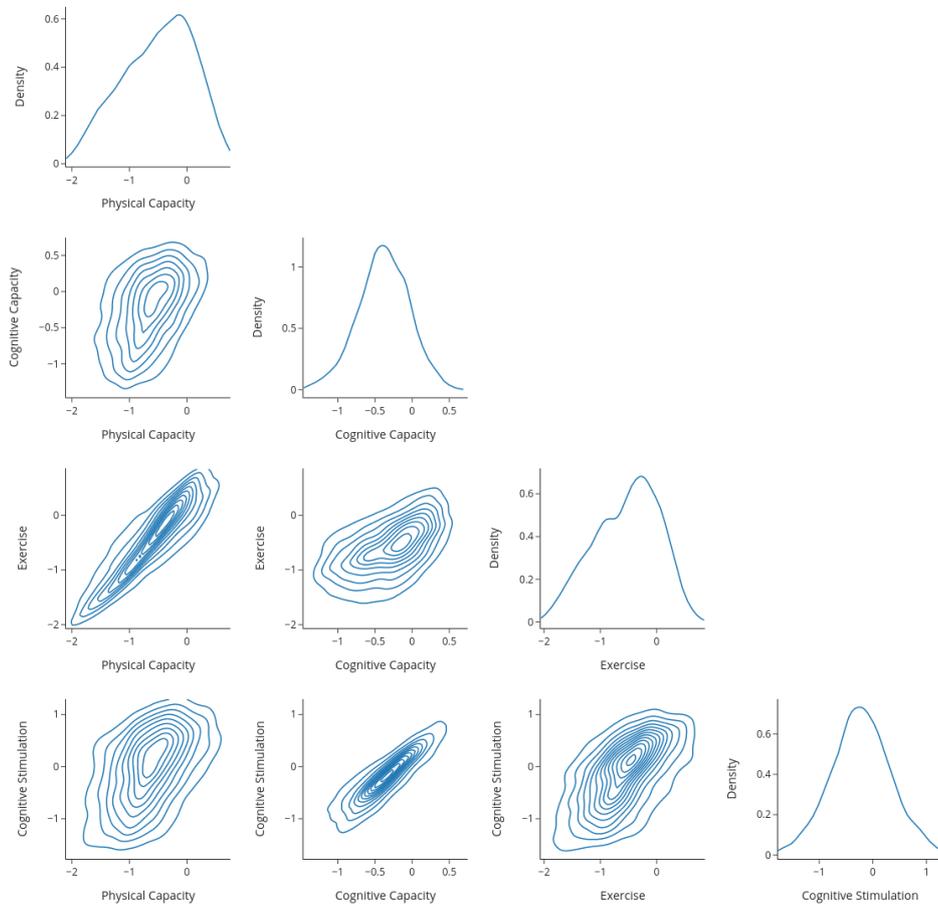


Figure 3.D.11. Factor distributions – males aged 80

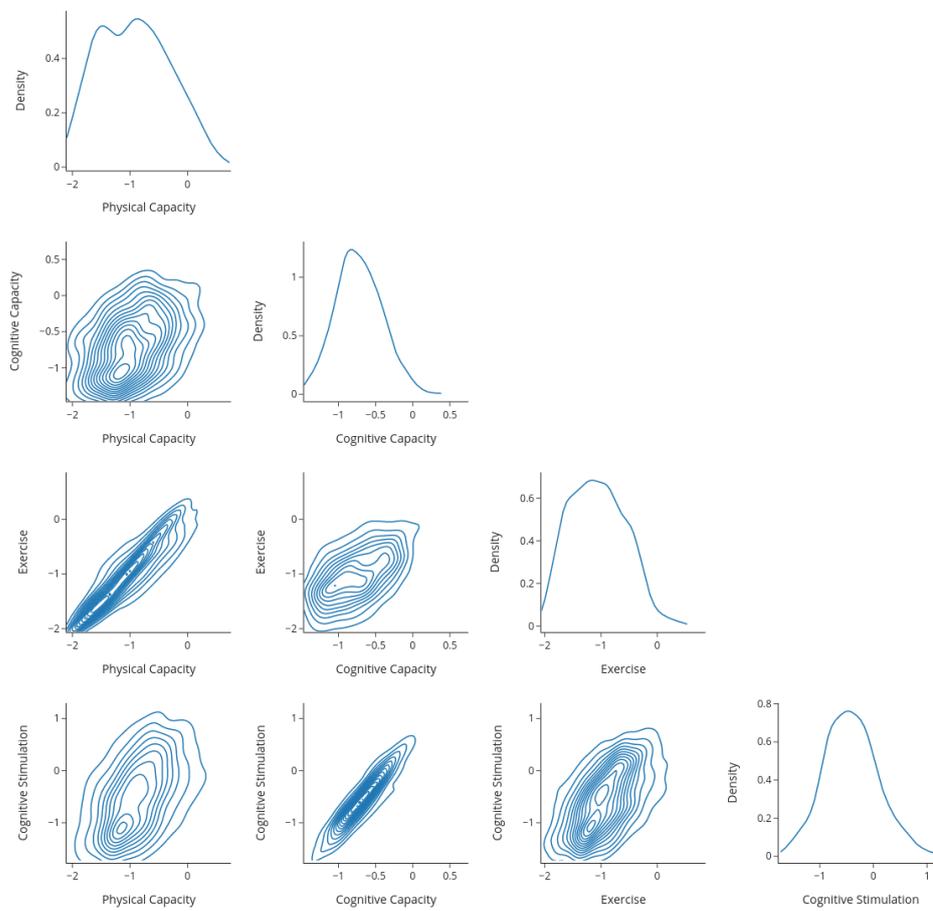


Figure 3.D.12. Factor distributions – males aged 90

3.D.4 Transition Equations

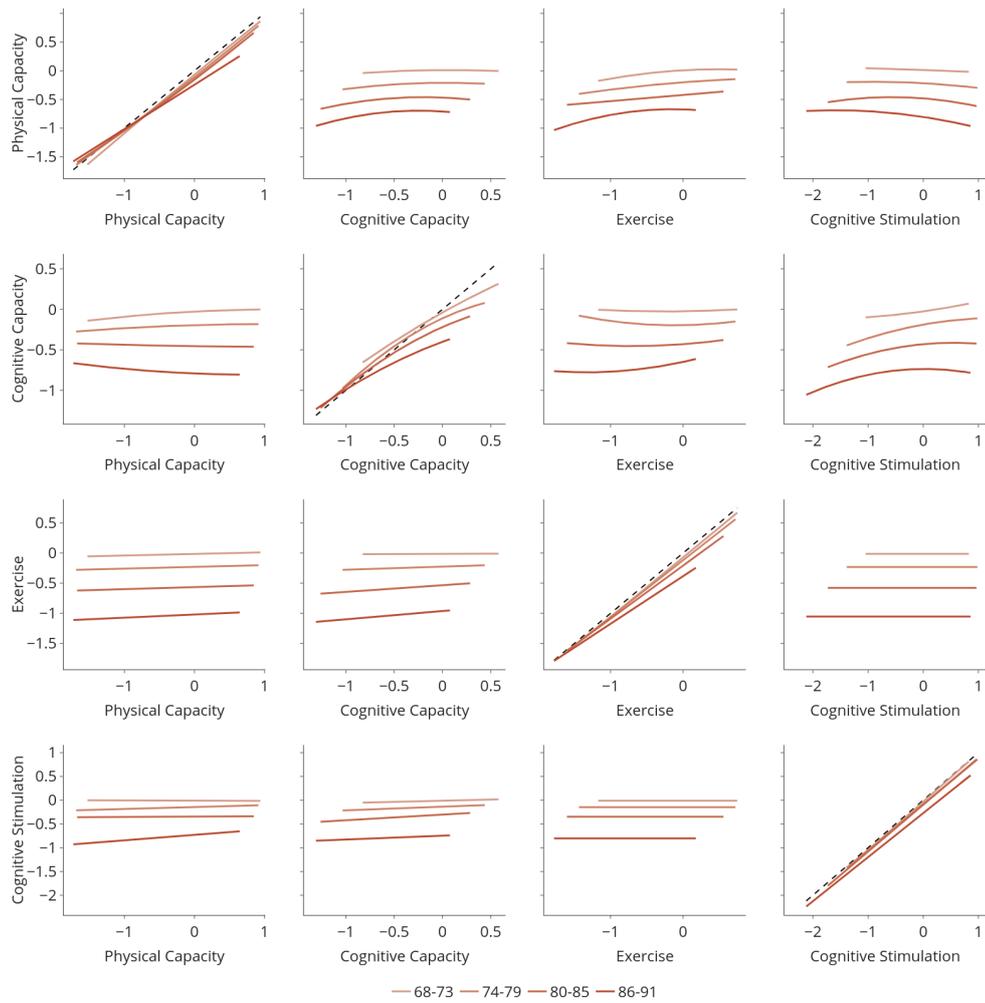


Figure 3.D.13. Transition equations for all factors (other factors evaluated at the median), females

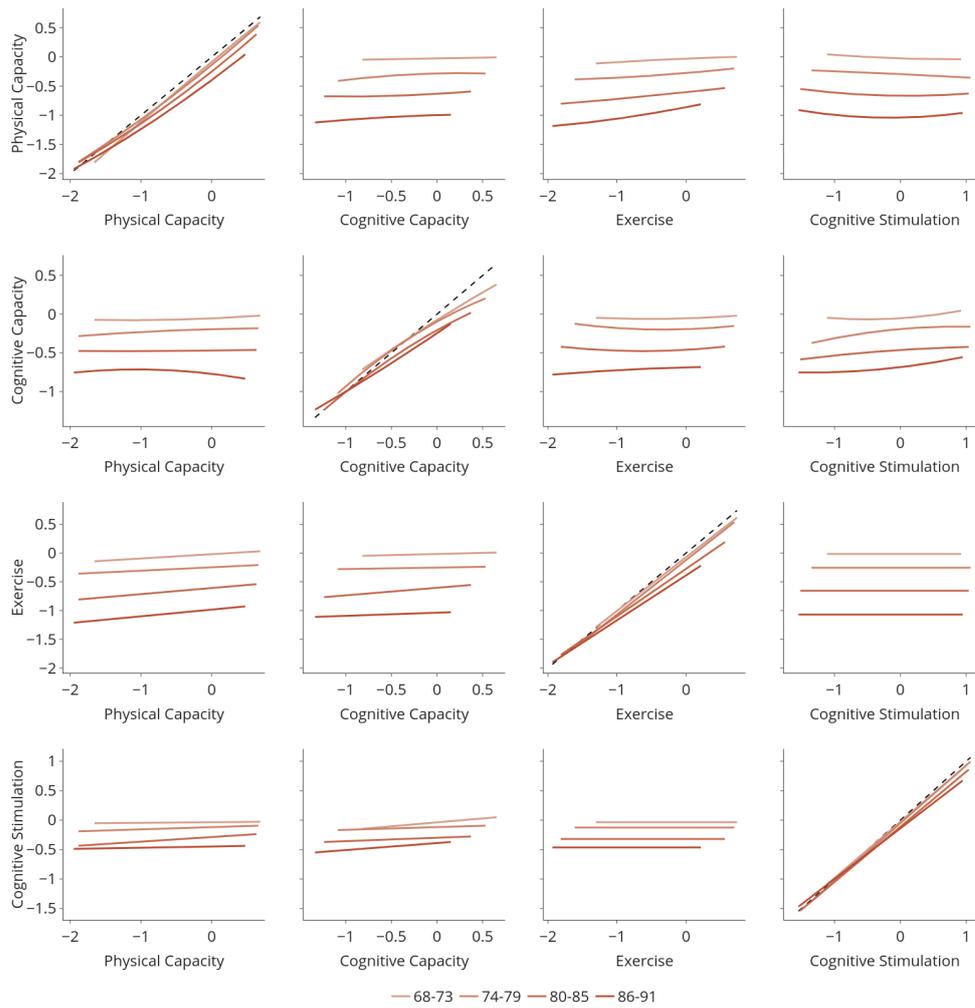


Figure 3.D.14. Transition equations for all factors (other factors evaluated at the median), males

Table 3.D.9. Transition Parameters for Physical Capacity, Females

	68-73	74-79	80-85	86-91
Physical Capacity	0.999** (0.006)	0.972** (0.008)	0.969** (0.014)	0.994** (0.035)
Cognitive Capacity	0.007 (0.008)	0.003 (0.011)	-0.086** (0.023)	-0.093 (0.069)
Exercise	0.066** (0.007)	0.074** (0.009)	0.054** (0.015)	-0.022 (0.043)
Cognitive Stimulation	-0.035** (0.006)	-0.025** (0.008)	0.039** (0.015)	0.021 (0.034)
Physical Capacity Squared	-0.015 (0.009)	0.030** (0.012)	0.041** (0.015)	0.002 (0.030)
Cognitive Capacity Squared	-0.066** (0.015)	-0.097** (0.019)	-0.180** (0.031)	-0.235** (0.067)
Exercise Squared	-0.074** (0.016)	-0.031* (0.016)	-0.001 (0.021)	-0.136** (0.040)
Cognitive Stimulation Squared	-0.004 (0.013)	-0.029** (0.012)	-0.066** (0.015)	-0.045** (0.021)
Physical Capacity × Cognitive Capacity	0.069** (0.018)	0.099** (0.021)	0.142** (0.032)	0.221** (0.065)
Physical Capacity × Exercise	0.075** (0.019)	0.024 (0.021)	0.014 (0.027)	0.161** (0.054)
Physical Capacity × Cognitive Stimulation	-0.074** (0.015)	-0.036** (0.016)	-0.080** (0.021)	-0.115** (0.036)
Cognitive Capacity × Exercise	-0.099** (0.024)	-0.211** (0.026)	-0.267** (0.040)	-0.254** (0.071)
Cognitive Capacity × Cognitive Stimulation	0.077** (0.022)	0.148** (0.025)	0.187** (0.037)	0.170** (0.060)
Exercise × Cognitive Stimulation	0.085** (0.024)	0.094** (0.020)	0.170** (0.027)	0.148** (0.043)
Constant	-0.072** (0.006)	-0.104** (0.007)	-0.120** (0.011)	-0.089** (0.025)

Note: ***p<0.01,**p<0.05,*p<0.1

Table 3.D.10. Transition Parameters for Physical Capacity, Males

	68-73	74-79	80-85	86-91
Physical Capacity	1.010** (0.008)	0.989** (0.011)	0.997** (0.023)	0.939** (0.062)
Cognitive Capacity	0.034** (0.009)	0.036** (0.013)	0.034 (0.035)	-0.031 (0.113)
Exercise	0.051** (0.006)	0.082** (0.009)	0.070** (0.020)	0.133* (0.070)
Cognitive Stimulation	-0.039** (0.007)	-0.037** (0.009)	-0.029 (0.022)	-0.048 (0.065)
Physical Capacity Squared	-0.020* (0.011)	0.057** (0.013)	0.073** (0.024)	0.062 (0.054)
Cognitive Capacity Squared	0.004 (0.018)	-0.074** (0.021)	0.047 (0.046)	-0.037 (0.112)
Exercise Squared	-0.012 (0.012)	0.019 (0.014)	0.009 (0.025)	0.032 (0.062)
Cognitive Stimulation Squared	0.024** (0.009)	-0.005 (0.011)	0.044** (0.021)	0.067* (0.035)
Physical Capacity × Cognitive Capacity	-0.010 (0.019)	0.024 (0.024)	0.131** (0.047)	0.120 (0.114)
Physical Capacity × Exercise	-0.006 (0.017)	-0.045** (0.020)	0.002 (0.039)	-0.019 (0.088)
Physical Capacity × Cognitive Stimulation	0.006 (0.015)	0.026 (0.017)	-0.074** (0.032)	-0.052 (0.071)
Cognitive Capacity × Exercise	-0.078** (0.019)	-0.073** (0.025)	-0.207** (0.050)	-0.152 (0.119)
Cognitive Capacity × Cognitive Stimulation	-0.026 (0.021)	0.076** (0.024)	-0.088* (0.050)	-0.088 (0.105)
Exercise × Cognitive Stimulation	0.047** (0.016)	0.011 (0.017)	0.101** (0.034)	0.027 (0.080)
Constant	-0.090** (0.007)	-0.124** (0.008)	-0.191** (0.016)	-0.246** (0.044)

Note: *** p<0.01, ** p<0.05, * p<0.1

Table 3.D.11. Transition Parameters for Cognitive Capacity, Females

	68-73	74-79	80-85	86-91
Physical Capacity	0.042** (0.008)	0.046** (0.009)	0.011 (0.013)	0.016 (0.028)
Cognitive Capacity	0.664** (0.010)	0.568** (0.012)	0.600** (0.019)	0.521** (0.052)
Exercise	0.013 (0.011)	-0.011 (0.011)	0.038** (0.016)	0.039 (0.037)
Cognitive Stimulation	0.100** (0.010)	0.151** (0.010)	0.129** (0.014)	0.194** (0.027)
Physical Capacity Squared	-0.019 (0.013)	-0.014 (0.014)	0.005 (0.016)	0.022 (0.024)
Cognitive Capacity Squared	-0.096** (0.018)	-0.274** (0.023)	-0.210** (0.027)	-0.150** (0.058)
Exercise Squared	0.025 (0.026)	0.067** (0.022)	0.047** (0.024)	0.072** (0.036)
Cognitive Stimulation Squared	0.025 (0.019)	-0.045** (0.016)	-0.058** (0.015)	-0.069** (0.019)
Physical Capacity × Cognitive Capacity	0.064** (0.021)	0.203** (0.024)	0.132** (0.033)	0.218** (0.058)
Physical Capacity × Exercise	0.030 (0.029)	-0.022 (0.026)	-0.034 (0.030)	-0.056 (0.046)
Physical Capacity × Cognitive Stimulation	-0.002 (0.022)	-0.070** (0.021)	-0.051** (0.022)	-0.081** (0.033)
Cognitive Capacity × Exercise	-0.044 (0.029)	-0.178** (0.030)	-0.056 (0.038)	-0.243** (0.072)
Cognitive Capacity × Cognitive Stimulation	0.015 (0.027)	0.227** (0.031)	0.199** (0.033)	0.262** (0.054)
Exercise × Cognitive Stimulation	-0.070** (0.034)	0.029 (0.028)	0.016 (0.029)	0.080* (0.042)
Constant	-0.045** (0.009)	-0.098** (0.010)	-0.163** (0.011)	-0.263** (0.020)

Note: ***p<0.01,**p<0.05,*p<0.1

Table 3.D.12. Transition Parameters for Cognitive Capacity, Males

	68-73	74-79	80-85	86-91
Physical Capacity	0.040** (0.010)	0.030** (0.012)	0.027 (0.017)	-0.041 (0.045)
Cognitive Capacity	0.738** (0.011)	0.668** (0.013)	0.699** (0.026)	0.760** (0.075)
Exercise	0.029** (0.010)	0.019* (0.011)	0.025 (0.016)	0.090* (0.053)
Cognitive Stimulation	0.059** (0.010)	0.099** (0.010)	0.099** (0.016)	0.020 (0.039)
Physical Capacity Squared	0.018 (0.017)	-0.012 (0.018)	0.006 (0.023)	-0.052 (0.044)
Cognitive Capacity Squared	-0.054** (0.018)	-0.177** (0.024)	-0.141** (0.040)	0.044 (0.088)
Exercise Squared	0.026 (0.021)	0.046** (0.020)	0.040* (0.023)	-0.013 (0.057)
Cognitive Stimulation Squared	0.054** (0.014)	-0.044** (0.014)	-0.017 (0.017)	0.037 (0.025)
Physical Capacity × Cognitive Capacity	0.043* (0.025)	0.062** (0.031)	0.071* (0.043)	-0.023 (0.087)
Physical Capacity × Exercise	-0.001 (0.030)	-0.008 (0.031)	-0.030 (0.034)	0.092 (0.081)
Physical Capacity × Cognitive Stimulation	-0.000 (0.022)	-0.028 (0.023)	0.006 (0.028)	0.006 (0.055)
Cognitive Capacity × Exercise	-0.039 (0.028)	-0.083** (0.031)	-0.033 (0.042)	0.016 (0.092)
Cognitive Capacity × Cognitive Stimulation	-0.068** (0.025)	0.162** (0.027)	0.114** (0.042)	-0.098 (0.076)
Exercise × Cognitive Stimulation	-0.023 (0.027)	0.031 (0.025)	-0.003 (0.029)	-0.031 (0.059)
Constant	-0.079** (0.010)	-0.082** (0.011)	-0.164** (0.014)	-0.209** (0.028)
<i>Note:</i>	*** p<0.01; ** p<0.05; * p<0.1			

Table 3.D.13. Transition Parameters for Exercise, Females

	68-73	74-79	80-85	86-91
Physical Capacity	0.026*** (0.010)	0.029*** (0.011)	0.035** (0.014)	0.053** (0.021)
Cognitive Capacity	0.006 (0.011)	0.050*** (0.011)	0.110*** (0.015)	0.138*** (0.027)
Exercise	0.990*** (0.014)	0.941*** (0.014)	0.880*** (0.018)	0.790*** (0.027)
Constant	-0.074*** (0.004)	-0.109*** (0.005)	-0.155*** (0.008)	-0.258*** (0.017)
<i>Note:</i>	***p<0.01;**p<0.05;*p<0.1			

Table 3.D.14. Transition Parameters for Exercise, Males

	68-73	74-79	80-85	86-91
Physical Capacity	0.075*** (0.012)	0.059*** (0.013)	0.106*** (0.021)	0.117*** (0.038)
Cognitive Capacity	0.038*** (0.013)	0.025* (0.014)	0.132*** (0.022)	0.056 (0.044)
Exercise	0.933*** (0.015)	0.945*** (0.015)	0.825*** (0.022)	0.782*** (0.047)
Constant	-0.078*** (0.005)	-0.111*** (0.006)	-0.178*** (0.011)	-0.266*** (0.025)
<i>Note:</i>	***p<0.01;**p<0.05;*p<0.1			

Table 3.D.15. Transition Parameters for Cognitive Stimulation, Females

	68-73	74-79	80-85	86-91
Physical Capacity	−0.006 (0.013)	0.041*** (0.014)	0.008 (0.020)	0.116*** (0.041)
Cognitive Capacity	0.050** (0.022)	0.076*** (0.023)	0.120*** (0.035)	0.081 (0.072)
Cognitive Stimulation	1.020*** (0.018)	0.962*** (0.017)	0.985*** (0.024)	0.927*** (0.046)
Constant	−0.033*** (0.007)	−0.071*** (0.009)	−0.046*** (0.014)	−0.144*** (0.035)
<i>Note:</i>	***p<0.01,**p<0.05,*p<0.1			

Table 3.D.16. Transition Parameters for Cognitive Stimulation, Males

	68-73	74-79	80-85	86-91
Physical Capacity	0.011 (0.018)	0.037** (0.019)	0.079** (0.032)	0.020 (0.080)
Cognitive Capacity	0.134*** (0.026)	0.046 (0.031)	0.057 (0.050)	0.119 (0.141)
Cognitive Stimulation	0.953*** (0.018)	0.982*** (0.019)	0.936*** (0.030)	0.858*** (0.067)
Constant	−0.033*** (0.009)	−0.039*** (0.011)	−0.056** (0.022)	−0.051 (0.070)
<i>Note:</i>	***p<0.01,**p<0.05,*p<0.1			

3.D.5 Distributions of Initial Factors and of Shocks to Factors

Table 3.D.17. Distribution of the initial states, females

Factor	Mean	Standard Deviation	Correlation with			
			Physical Capacity	Cognitive Capacity	Exercise	Cognitive Stimulation
Physical Capacity	0.19	0.68	1.00	0.35	0.66	0.36
Cognitive Capacity	0.11	0.46	0.35	1.00	0.32	0.52
Exercise	0.15	0.59	0.66	0.32	1.00	0.51
Cognitive Stimulation	0.08	0.68	0.36	0.52	0.51	1.00

Table 3.D.18. Distribution of the initial states, males

Factor	Mean	Standard Deviation	Correlation with			
			Physical Capacity	Cognitive Capacity	Exercise	Cognitive Stimulation
Physical Capacity	0.10	0.61	1.00	0.30	0.58	0.30
Cognitive Capacity	0.11	0.49	0.30	1.00	0.27	0.42
Exercise	0.12	0.64	0.58	0.27	1.00	0.33
Cognitive Stimulation	0.04	0.79	0.30	0.42	0.33	1.00

Table 3.D.19. Standard deviations of shocks

		Male	Female
68-73	Physical Capacity	0.094*** (0.008)	0.005 (0.103)
	Cognitive Capacity	0.292*** (0.005)	0.308*** (0.004)
	Exercise	0.236*** (0.012)	0.164*** (0.011)
	Cognitive Stimulation	0.155*** (0.028)	0.001 (2.725)
74-79	Physical Capacity	0.161*** (0.006)	0.159*** (0.005)
	Cognitive Capacity	0.283*** (0.005)	0.302*** (0.004)
	Exercise	0.261*** (0.012)	0.240*** (0.009)
	Cognitive Stimulation	0.190*** (0.026)	0.185*** (0.017)
80-85	Physical Capacity	0.231*** (0.009)	0.188*** (0.008)
	Cognitive Capacity	0.240*** (0.006)	0.274*** (0.005)
	Exercise	0.326*** (0.015)	0.275*** (0.012)
	Cognitive Stimulation	0.319*** (0.031)	0.225*** (0.024)
86-91	Physical Capacity	0.315*** (0.020)	0.227*** (0.012)
	Cognitive Capacity	0.250*** (0.013)	0.238*** (0.010)
	Exercise	0.372*** (0.028)	0.304*** (0.017)
	Cognitive Stimulation	0.471*** (0.059)	0.309*** (0.047)

Note: *** p<0.01, ** p<0.05, * p<0.1

Appendix 3.E Results for a Linearized Model

3.E.1 Measurement System

Table 3.E.1. Loadings and Measurement Standard Deviations for Physical Capacity, Females

		Intercept	Loading	Meas. Std.
All	Frailty Index (Reversed)	0.000	1.000	0.705*** (0.001)
	Mobility	-0.114*** (0.003)	1.222*** (0.005)	0.768*** (0.002)
	Large Muscle Index	0.005* (0.003)	0.926*** (0.005)	0.750*** (0.002)
	Self-Reported Health	-0.048*** (0.003)	0.947*** (0.004)	0.765*** (0.002)
70	Alive	0.897*** (0.101)	0.042*** (0.011)	0.303*** (0.038)
	Grip Strength	-0.125*** (0.027)	0.488*** (0.042)	0.933*** (0.015)
72	Alive	0.910*** (0.106)	0.045*** (0.011)	0.288*** (0.037)
	Grip Strength	-0.240*** (0.028)	0.395*** (0.042)	0.922*** (0.016)
74	Alive	0.902*** (0.096)	0.060*** (0.013)	0.301*** (0.036)
	Grip Strength	-0.291*** (0.030)	0.464*** (0.042)	0.936*** (0.018)
76	Alive	0.886*** (0.099)	0.073*** (0.018)	0.327*** (0.042)
	Grip Strength	-0.470*** (0.030)	0.367*** (0.048)	0.924*** (0.012)
78	Alive	0.879*** (0.101)	0.075*** (0.019)	0.339*** (0.045)
	Grip Strength	-0.540*** (0.033)	0.445*** (0.048)	0.924*** (0.019)
80	Alive	0.870*** (0.097)	0.091*** (0.022)	0.353*** (0.046)
	Grip Strength	-0.758*** (0.034)	0.365*** (0.052)	0.882*** (0.021)
82	Alive	0.871*** (0.109)	0.090*** (0.026)	0.359*** (0.053)
	Grip Strength	-0.789*** (0.036)	0.339*** (0.054)	0.860*** (0.020)
84	Alive	0.869*** (0.103)	0.110*** (0.030)	0.371*** (0.052)
	Grip Strength	-0.979*** (0.041)	0.336*** (0.060)	0.866*** (0.025)
86	Alive	0.855*** (0.124)	0.120*** (0.040)	0.391*** (0.069)
	Grip Strength	-0.999*** (0.046)	0.332*** (0.070)	0.840*** (0.028)
88	Alive	0.845*** (0.142)	0.128*** (0.051)	0.406*** (0.084)
	Grip Strength	-1.190*** (0.059)	0.415*** (0.082)	0.826*** (0.035)
90	Alive	0.826*** (0.204)	0.133* (0.080)	0.425*** (0.135)
	Grip Strength	-1.099*** (0.061)	0.371*** (0.097)	0.734*** (0.031)
92	Alive	0.816*** (0.228)	0.164 (0.120)	0.444*** (0.159)
	Grip Strength	-1.357*** (0.083)	0.356*** (0.115)	0.745*** (0.047)

Note: ***p<0.01; **p<0.05; *p<0.1

Table 3.E.2. Loadings and Measurement Standard Deviations for Physical Capacity, Males

		Intercept	Loading	Meas. Std.
All	Frailty Index (Reversed)	0.000	1.000	0.796*** (0.002)
	Mobility	-0.015*** (0.005)	1.330*** (0.007)	0.751*** (0.003)
	Large Muscle Index	0.043*** (0.004)	1.033*** (0.006)	0.761*** (0.003)
	Self-Reported Health	0.027*** (0.003)	0.964*** (0.006)	0.792*** (0.003)
70	Alive	0.901*** (0.093)	0.057*** (0.013)	0.303*** (0.035)
	Grip Strength	-0.055 (0.034)	0.578*** (0.053)	0.977*** (0.020)
72	Alive	0.907*** (0.083)	0.074*** (0.015)	0.298*** (0.030)
	Grip Strength	-0.294*** (0.034)	0.549*** (0.053)	0.959*** (0.020)
74	Alive	0.900*** (0.119)	0.061*** (0.017)	0.310*** (0.046)
	Grip Strength	-0.317*** (0.035)	0.499*** (0.057)	0.922*** (0.021)
76	Alive	0.876*** (0.129)	0.073*** (0.024)	0.344*** (0.059)
	Grip Strength	-0.505*** (0.036)	0.557*** (0.056)	0.898*** (0.020)
78	Alive	0.872*** (0.128)	0.081*** (0.026)	0.355*** (0.061)
	Grip Strength	-0.559*** (0.040)	0.552*** (0.058)	0.920*** (0.022)
80	Alive	0.866*** (0.132)	0.089*** (0.031)	0.367*** (0.066)
	Grip Strength	-0.736*** (0.042)	0.573*** (0.062)	0.891*** (0.023)
82	Alive	0.853*** (0.114)	0.136*** (0.042)	0.393*** (0.064)
	Grip Strength	-0.959*** (0.046)	0.466*** (0.064)	0.873*** (0.025)
84	Alive	0.869*** (0.127)	0.138*** (0.046)	0.387*** (0.067)
	Grip Strength	-1.040*** (0.052)	0.561*** (0.068)	0.842*** (0.027)
86	Alive	0.847*** (0.158)	0.137** (0.061)	0.408*** (0.092)
	Grip Strength	-1.237*** (0.063)	0.491*** (0.083)	0.841*** (0.033)
88	Alive	0.858*** (0.145)	0.179** (0.071)	0.416*** (0.083)
	Grip Strength	-1.280*** (0.069)	0.480*** (0.107)	0.824*** (0.044)
90	Alive	0.851*** (0.201)	0.204* (0.116)	0.429*** (0.120)
	Grip Strength	-1.361*** (0.097)	0.503*** (0.114)	0.767*** (0.053)
92	Alive	0.765** (0.317)	0.183 (0.220)	0.464* (0.271)
	Grip Strength	-1.487*** (0.120)	0.683*** (0.162)	0.817*** (0.076)
Note:		***p<0.01;**p<0.05;*p<0.1		

Table 3.E.3. Loadings and Measurement Standard Deviations for Cognitive Capacity, Females

		Intercept	Loading	Meas. Std.
All	Serial 7 Subtraction	0.000	1.000	0.890 ^{***} (0.003)
	Vocabulary	0.044 ^{***} (0.006)	0.840 ^{***} (0.013)	0.923 ^{***} (0.004)
	Immediate Word Recall	-0.161 ^{***} (0.006)	1.799 ^{***} (0.014)	0.584 ^{***} (0.003)
	Delayed Word Recall	-0.189 ^{***} (0.006)	1.803 ^{***} (0.014)	0.595 ^{***} (0.002)
70	Self-Rated Memory	0.005 (0.014)	0.577 ^{***} (0.031)	0.961 ^{***} (0.009)
72	Self-Rated Memory	0.029 ^{**} (0.014)	0.595 ^{***} (0.030)	0.954 ^{***} (0.009)
74	Self-Rated Memory	0.016 (0.015)	0.562 ^{***} (0.030)	0.972 ^{***} (0.009)
76	Self-Rated Memory	0.028 [*] (0.017)	0.496 ^{***} (0.032)	0.968 ^{***} (0.010)
78	Self-Rated Memory	0.046 ^{**} (0.019)	0.501 ^{***} (0.035)	0.992 ^{***} (0.011)
80	Self-Rated Memory	0.054 ^{**} (0.022)	0.480 ^{***} (0.038)	1.012 ^{***} (0.012)
82	Self-Rated Memory	0.069 ^{**} (0.027)	0.460 ^{***} (0.043)	1.009 ^{***} (0.013)
84	Self-Rated Memory	0.082 ^{**} (0.032)	0.396 ^{***} (0.050)	1.035 ^{***} (0.015)
86	Self-Rated Memory	0.079 ^{**} (0.040)	0.393 ^{***} (0.058)	1.063 ^{***} (0.018)
88	Self-Rated Memory	0.261 ^{***} (0.054)	0.549 ^{***} (0.074)	1.069 ^{***} (0.021)
90	Self-Rated Memory	0.213 ^{***} (0.074)	0.463 ^{***} (0.096)	1.080 ^{***} (0.026)
92	Self-Rated Memory	0.215 ^{**} (0.108)	0.532 ^{***} (0.131)	1.146 ^{***} (0.040)
<i>Note:</i>			*** p<0.01; ** p<0.05; * p<0.1	

Table 3.E.4. Loadings and Measurement Standard Deviations for Cognitive Capacity, Males

		Intercept	Loading	Meas. Std.
All	Serial 7 Subtraction	0.000	1.000	0.907*** (0.004)
	Vocabulary	0.048*** (0.007)	0.962*** (0.015)	0.900*** (0.004)
	Immediate Word Recall	-0.184*** (0.008)	1.683*** (0.015)	0.600*** (0.003)
	Delayed Word Recall	-0.201*** (0.008)	1.647*** (0.015)	0.607*** (0.003)
70	Self-Rated Memory	-0.041** (0.016)	0.627*** (0.035)	0.937*** (0.011)
72	Self-Rated Memory	-0.052*** (0.017)	0.563*** (0.034)	0.955*** (0.011)
74	Self-Rated Memory	-0.043** (0.017)	0.579*** (0.035)	0.948*** (0.011)
76	Self-Rated Memory	-0.039** (0.019)	0.528*** (0.039)	0.955*** (0.012)
78	Self-Rated Memory	-0.050** (0.022)	0.610*** (0.043)	0.971*** (0.013)
80	Self-Rated Memory	-0.001 (0.026)	0.596*** (0.048)	0.988*** (0.015)
82	Self-Rated Memory	-0.019 (0.034)	0.478*** (0.056)	1.033*** (0.018)
84	Self-Rated Memory	-0.018 (0.039)	0.520*** (0.062)	1.007*** (0.020)
86	Self-Rated Memory	-0.018 (0.045)	0.465*** (0.069)	0.992*** (0.021)
88	Self-Rated Memory	0.007 (0.063)	0.511*** (0.088)	1.035*** (0.027)
90	Self-Rated Memory	0.013 (0.086)	0.391*** (0.117)	1.080*** (0.037)
92	Self-Rated Memory	0.006 (0.124)	0.602*** (0.180)	1.011*** (0.048)

Note: *** p<0.01; ** p<0.05; * p<0.1

Table 3.E.5. Loadings and Measurement Standard Deviations for Exercise, Females

		Intercept	Loading	Meas. Std.
All	Vigorous Activity	−0.009 (0.006)	0.683*** (0.010)	0.808*** (0.004)
	Moderate Activity	0.000	1.000	0.794*** (0.004)
	Light Activity	−0.127*** (0.007)	1.077*** (0.012)	0.933*** (0.004)
<i>Note:</i>		***p<0.01; **p<0.05; *p<0.1		

Table 3.E.6. Loadings and Measurement Standard Deviations for Exercise, Males

		Intercept	Loading	Meas. Std.
All	Vigorous Activity	−0.012** (0.006)	0.742*** (0.012)	0.813*** (0.005)
	Moderate Activity	0.000	1.000	0.816*** (0.004)
	Light Activity	−0.078*** (0.007)	0.927*** (0.013)	0.861*** (0.004)
<i>Note:</i>		***p<0.01; **p<0.05; *p<0.1		

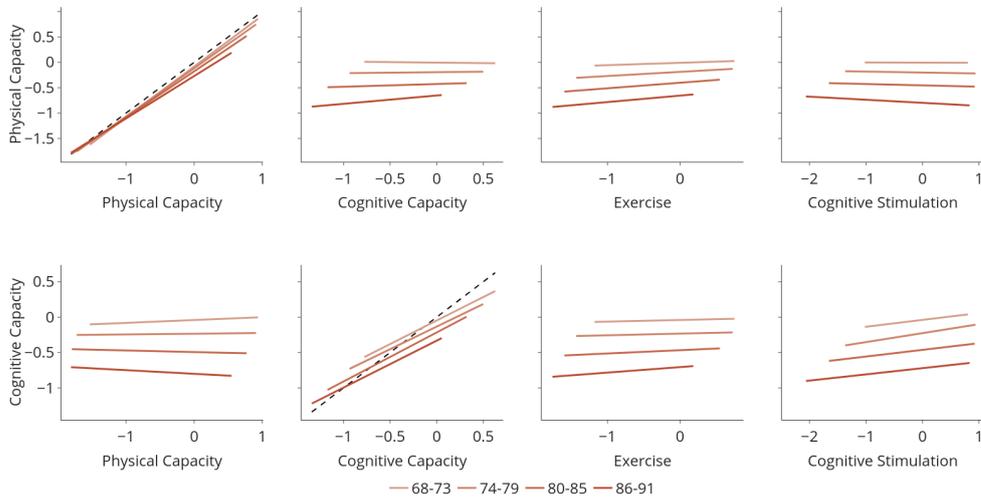
Table 3.E.7. Loadings and Measurement Standard Deviations for Cognitive Stimulation, Females

		Intercept	Loading	Meas. Std.
All	Reading	0.000	1.000	0.769*** (0.006)
	Listening to Music	-0.168*** (0.006)	0.498*** (0.010)	0.981*** (0.006)
	Stimulating Hobbies	-0.068*** (0.007)	0.564*** (0.011)	0.926*** (0.005)
	Communication	-0.062*** (0.006)	0.513*** (0.010)	0.999*** (0.005)
<i>Note:</i>		***p<0.01; **p<0.05; *p<0.1		

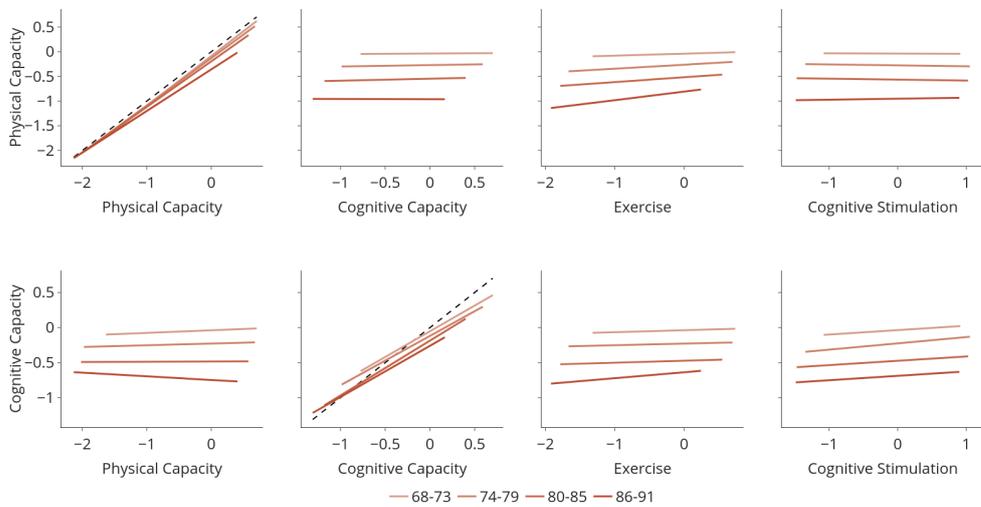
Table 3.E.8. Loadings and Measurement Standard Deviations for Cognitive Stimulation, Males

		Intercept	Loading	Meas. Std.
All	Reading	0.000	1.000	0.674*** (0.007)
	Listening to Music	-0.175*** (0.007)	0.223*** (0.010)	1.005*** (0.007)
	Stimulating Hobbies	-0.011 (0.009)	0.369*** (0.011)	0.970*** (0.005)
	Communication	-0.082*** (0.007)	0.320*** (0.010)	0.990*** (0.006)
<i>Note:</i>		***p<0.01; **p<0.05; *p<0.1		

3.E.2 Transition Equations



(a) Transitions, females



(b) Transitions, males

Figure 3.E.1. Transition equations (other factors evaluated at the median)

Table 3.E.9. Transition Parameters for Physical Capacity, Females

	68-73	74-79	80-85	86-91
Physical Capacity	1.000*** (0.006)	0.950*** (0.007)	0.905*** (0.009)	0.835*** (0.017)
Cognitive Capacity	-0.020*** (0.007)	0.020** (0.009)	0.053*** (0.015)	0.164*** (0.029)
Exercise	0.046*** (0.007)	0.081*** (0.008)	0.109*** (0.010)	0.128*** (0.019)
Cognitive Stimulation	-0.003 (0.007)	-0.019** (0.008)	-0.026** (0.011)	-0.061*** (0.017)
Constant	-0.086*** (0.002)	-0.105*** (0.004)	-0.117*** (0.007)	-0.090*** (0.018)
<i>Note:</i>	***p<0.01; **p<0.05; *p<0.1			

Table 3.E.10. Transition Parameters for Physical Capacity, Males

	68-73	74-79	80-85	86-91
Physical Capacity	1.010*** (0.007)	0.957*** (0.008)	0.922*** (0.013)	0.842*** (0.030)
Cognitive Capacity	0.009 (0.009)	0.028** (0.011)	0.038** (0.019)	-0.003 (0.042)
Exercise	0.039*** (0.006)	0.081*** (0.007)	0.098*** (0.013)	0.173*** (0.030)
Cognitive Stimulation	-0.004 (0.007)	-0.018** (0.008)	-0.020 (0.014)	0.020 (0.029)
Constant	-0.090*** (0.003)	-0.115*** (0.005)	-0.139*** (0.010)	-0.200*** (0.029)
<i>Note:</i>	***p<0.01; **p<0.05; *p<0.1			

Table 3.E.11. Transition Parameters for Cognitive Capacity, Females

	68-73	74-79	80-85	86-91
Physical Capacity	0.040** (0.007)	0.011 (0.008)	-0.023** (0.010)	-0.052** (0.015)
Cognitive Capacity	0.664** (0.010)	0.639** (0.011)	0.693** (0.014)	0.664** (0.026)
Exercise	0.023** (0.010)	0.024** (0.011)	0.047** (0.012)	0.078** (0.018)
Cognitive Stimulation	0.097** (0.010)	0.126** (0.010)	0.095** (0.011)	0.087** (0.016)
Constant	-0.054** (0.003)	-0.122** (0.004)	-0.180** (0.006)	-0.248** (0.013)
<i>Note:</i>	***p<0.01; **p<0.05; *p<0.1			

Table 3.E.12. Transition Parameters for Cognitive Capacity, Males

	68-73	74-79	80-85	86-91
Physical Capacity	0.037** (0.009)	0.024** (0.010)	0.003 (0.012)	-0.052** (0.024)
Cognitive Capacity	0.733** (0.011)	0.704** (0.013)	0.782** (0.017)	0.730** (0.033)
Exercise	0.028** (0.010)	0.023** (0.010)	0.028** (0.012)	0.084** (0.026)
Cognitive Stimulation	0.063** (0.010)	0.089** (0.010)	0.061** (0.013)	0.063** (0.023)
Constant	-0.056** (0.004)	-0.105** (0.004)	-0.160** (0.007)	-0.203** (0.017)
<i>Note:</i>	***p<0.01; **p<0.05; *p<0.1			

Table 3.E.13. Transition Parameters for Exercise, Females

	68-73	74-79	80-85	86-91
Physical Capacity	0.027*** (0.010)	0.030*** (0.011)	0.027** (0.014)	0.052** (0.021)
Cognitive Capacity	0.005 (0.010)	0.040*** (0.011)	0.101*** (0.016)	0.115*** (0.027)
Exercise	0.991*** (0.014)	0.941*** (0.014)	0.886*** (0.018)	0.802*** (0.026)
Constant	-0.073*** (0.004)	-0.111*** (0.005)	-0.158*** (0.008)	-0.261*** (0.016)
<i>Note:</i>	***p<0.01;**p<0.05;*p<0.1			

Table 3.E.14. Transition Parameters for Exercise, Males

	68-73	74-79	80-85	86-91
Physical Capacity	0.073*** (0.012)	0.063*** (0.013)	0.106*** (0.020)	0.117*** (0.037)
Cognitive Capacity	0.038*** (0.013)	0.022 (0.014)	0.133*** (0.022)	0.045 (0.042)
Exercise	0.934*** (0.015)	0.942*** (0.015)	0.820*** (0.021)	0.786*** (0.046)
Constant	-0.078*** (0.005)	-0.111*** (0.006)	-0.180*** (0.011)	-0.268*** (0.024)
<i>Note:</i>	***p<0.01;**p<0.05;*p<0.1			

Table 3.E.15. Transition Parameters for Cognitive Stimulation, Females

	68-73	74-79	80-85	86-91
Physical Capacity	−0.010 (0.013)	0.024* (0.014)	−0.000 (0.021)	0.101** (0.043)
Cognitive Capacity	0.041* (0.023)	0.097*** (0.024)	0.115*** (0.038)	0.109 (0.076)
Cognitive Stimulation	1.030*** (0.018)	0.960*** (0.017)	0.980*** (0.025)	0.922*** (0.049)
Constant	−0.037*** (0.007)	−0.064*** (0.009)	−0.057*** (0.015)	−0.141*** (0.034)

Note: ***p<0.01,**p<0.05;*p<0.1

Table 3.E.16. Transition Parameters for Cognitive Stimulation, Males

	68-73	74-79	80-85	86-91
Physical Capacity	−0.007 (0.018)	0.033* (0.019)	0.067** (0.032)	0.025 (0.080)
Cognitive Capacity	0.143*** (0.026)	0.059* (0.030)	0.048 (0.049)	0.110 (0.138)
Cognitive Stimulation	0.950*** (0.018)	0.974*** (0.019)	0.939*** (0.032)	0.843*** (0.067)
Constant	−0.034*** (0.009)	−0.036*** (0.011)	−0.062*** (0.023)	−0.059 (0.069)

Note: ***p<0.01,**p<0.05;*p<0.1

Table 3.E.17. Standard deviations of shocks

		Male All	Male All Linear	Female All	Female All Linear
68-73	Physical Capacity	0.094*** (0.008)	0.111*** (0.006)	0.005 (0.103)	0.038*** (0.014)
	Cognitive Capacity	0.292*** (0.005)	0.294*** (0.004)	0.308*** (0.004)	0.308*** (0.004)
	Exercise	0.236*** (0.012)	0.235*** (0.012)	0.164*** (0.011)	0.158*** (0.011)
	Cognitive Stimulation	0.155*** (0.028)	0.171*** (0.026)	0.001 (2.725)	0.001 (2.840)
74-79	Physical Capacity	0.161*** (0.006)	0.181*** (0.005)	0.159*** (0.005)	0.183*** (0.004)
	Cognitive Capacity	0.283*** (0.005)	0.287*** (0.005)	0.302*** (0.004)	0.309*** (0.004)
	Exercise	0.261*** (0.012)	0.264*** (0.012)	0.240*** (0.009)	0.243*** (0.009)
	Cognitive Stimulation	0.190*** (0.026)	0.198*** (0.025)	0.185*** (0.017)	0.192*** (0.017)
80-85	Physical Capacity	0.231*** (0.009)	0.252*** (0.007)	0.188*** (0.008)	0.219*** (0.006)
	Cognitive Capacity	0.240*** (0.006)	0.243*** (0.006)	0.274*** (0.005)	0.279*** (0.004)
	Exercise	0.326*** (0.015)	0.328*** (0.015)	0.275*** (0.012)	0.273*** (0.012)
	Cognitive Stimulation	0.319*** (0.031)	0.326*** (0.031)	0.225*** (0.024)	0.231*** (0.024)
86-91	Physical Capacity	0.315*** (0.020)	0.341*** (0.014)	0.227*** (0.012)	0.268*** (0.010)
	Cognitive Capacity	0.250*** (0.013)	0.254*** (0.009)	0.238*** (0.010)	0.258*** (0.007)
	Exercise	0.372*** (0.028)	0.372*** (0.027)	0.304*** (0.017)	0.295*** (0.018)
	Cognitive Stimulation	0.471*** (0.059)	0.487*** (0.056)	0.309*** (0.047)	0.313*** (0.049)

Note: *** p<0.01; ** p<0.05; * p<0.1

References

- Agostinelli, Francesco, and Matthew Wiswall.** 2016a. “Estimating the Technology of Children’s Skill Formation.” Working Paper 22442. National Bureau of Economic Research. [145, 148, 157]
- Agostinelli, Francesco, and Matthew Wiswall.** 2016b. “Identification of Dynamic Latent Factor Models: The Implications of Re-Normalization in a Model of Child Development.” Working Paper 22441. National Bureau of Economic Research. [147, 168, 169]
- Amengual, Dante, Jesús Bueren, and Julio A Crego.** 2021. “Endogenous health groups and heterogeneous dynamics of the elderly.” *Journal of Applied Econometrics* 36 (7): 878–97. [134]
- Attanasio, Orazio, Flávio Cunha, and Pamela Jervis.** 2019. “Subjective Parental Beliefs. Their Measurement and Role.” NBER Working Papers 26516. National Bureau of Economic Research. [148]
- Attanasio, Orazio, Costas Meghir, and Emily Nix.** 2020. “Human Capital Development and Parental Investment in India.” *Review of Economic Studies* 87 (6): 2511–41. eprint: <https://academic.oup.com/restud/article-pdf/87/6/2511/34133061/rdaa026.pdf>. [147, 148, 157]
- Baker, Michael, Mark Stabile, and Catherine Deri.** 2004. “What do self-reported, objective, measures of health measure?” *Journal of human Resources* 39 (4): 1067–93. [131, 133]
- Ball, Karlene, Daniel Berch, Karin Helmers, Jared Jobe, Mary Leveck, Michael Marsiske, John Morris, George Rebok, David Smith, Sharon Tennstedt, Frederick Unverzagt, and Sherry Willis.** 2002. “Effects of Cognitive Training Interventions With Older Adults: A Randomized Controlled Trial.” *JAMA: the journal of the American Medical Association* 288 (12): 2271–81. [130]
- Bound, John, Charles Brown, and Nancy Mathiowetz.** 2001. “Measurement error in survey data.” In *Handbook of econometrics*. Vol. 5, Elsevier, 3705–843. [131]
- Bradbury, James, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, and Qiao Zhang.** 2018. *JAX: composable transformations of Python+NumPy programs*. Version 0.3.13. [149]
- Chang, Yu-Hung, I-Chien Wu, and Chao A. Hsiung.** 2021. “Reading activity prevents long-term decline in cognitive function in older people: evidence from a 14-year longitudinal study.” *International Psychogeriatrics* 33 (1): 63–74. [139]
- Clouston, Sean A. P., Paul Brewster, Diana Kuh, Marcus Richards, Rachel Cooper, Rebecca Hardy, Marcie S. Rubin, and Scott M. Hofer.** 2013. “The Dynamic Relationship Between Physical Function and Cognition in Longitudinal Aging Cohorts.” *Epidemiologic Reviews* 35 (1): 33–50. eprint: <https://academic.oup.com/epirev/article-pdf/35/1/33/7287926/mxs004.pdf>. [130]
- Contoyannis, Paul, Andrew M. Jones, and Nigel Rice.** 2004. “The dynamics of health in the British Household Panel Survey.” *Journal of Applied Econometrics* 19 (4): 473–503. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jae.755>. [132]

- Crimmins, Eileen M.** 2020. "Social hallmarks of aging: Suggestions for geroscience research." *Ageing research reviews* 63: 101136. [130, 134]
- Cunha, Flavio, and James Heckman.** 2007. "The Technology of Skill Formation." *American Economic Review* 97 (2): 31–47. [145]
- Cunha, Flavio, and James J. Heckman.** 2008. "Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation." *Journal of human resources* 43 (4): 738–82. [147]
- Cunha, Flavio, James J. Heckman, and Susanne Schennach.** 2010. "Estimating the technology of cognitive and noncognitive skill formation." *Econometrica* 78 (3): 883–931. [130, 131, 133, 145–148, 152, 164, 168, 169]
- Cutler, David M, and Elizabeth Richardson.** 1997. "Measuring the health of the US population." *Brookings papers on economic activity. Microeconomics* 1997: 217–82. [132]
- Del Bono, Emilia, Josh Kinsler, and Ronni Pavan.** 2020. "A Note on the Importance of Normalizations in Dynamic Latent Factor Models of Skill Formation." IZA Discussion Papers 13714. Institute of Labor Economics (IZA). [147]
- Dixon, Roger A., and Cindy M. de Frias.** 2014. "Cognitively elite, cognitively normal, and cognitively impaired aging: Neurocognitive status and stability moderate memory performance." *Journal of Clinical and Experimental Neuropsychology* 36 (4): 418–30. [138]
- Dowd, Jennifer Beam, and Anna Zajacova.** 2007. "Does the predictive power of self-rated health for subsequent mortality risk vary by socioeconomic status in the US?" *International journal of epidemiology* 36 (6): 1214–21. [133]
- Dowd, Jennifer Beam, and Anna Zajacova.** 2010. "Does self-rated health mean the same thing across socioeconomic groups? Evidence from biomarker data." *Annals of epidemiology* 20 (10): 743–49. [133]
- Fatarone, Maria A., Evelyn F. O'Neill, Nancy Doyle Ryan, Karen M. Clements, Guido R. Solares, Miriam E. Nelson, Susan B. Roberts, Joseph J. Kehayias, Lewis A. Lipsitz, and William J. Evans.** 1994. "Exercise Training and Nutritional Supplementation for Physical Frailty in Very Elderly People." *New England Journal of Medicine* 330 (25): 1769–75. eprint: <https://doi.org/10.1056/NEJM199406233302501>. PMID: 8190152. [130]
- Freyberger, Joachim.** 2021. "Normalizations and misspecification in skill formation models." Working Paper. [147, 169]
- Gabler, Janos.** 2022a. *A Python Library to Estimate Nonlinear Dynamic Latent Factor Models*. [149]
- Gabler, Janos.** 2022b. *A Python Tool for the Estimation of large scale scientific models*. [149]
- Gaudecker, Hans-Martin von.** 2019. "Templates for Reproducible Research Projects in Economics." [149]
- Gill, Thomas M., Evelyn A. Gahbauer, Ling Han, and Heather G. Allore.** 2010. "Trajectories of Disability in the Last Year of Life." *New England Journal of Medicine* 362 (13): 1173–80. eprint: <https://doi.org/10.1056/NEJMoa0909087>. PMID: 20357280. [151]
- Health and Retirement Study.** 2022a. "Cross-Wave Imputation of Cognitive Functioning Measures 1992-2018." [138]

- Health and Retirement Study.** 2022b. "HRS Public Survey Data." [134, 135]
- Health and Retirement Study.** 2022c. "RAND HRS Products." [135]
- Heckman, James J, and Stefano Mosso.** 2014. "The Economics of Human Development and Social Mobility." *Annu. Rev. Econ.* 6 (1): 689–733. [129]
- Heiss, Florian.** 2011. "Dynamics of self-rated health and selective mortality." *Empirical Economics* 40 (1): 119–40. [132]
- Hosseini, Roozbeh, Karen A. Kopecky, and Kai Zhao.** 2022. "The evolution of health over the life cycle." *Review of Economic Dynamics* 45: 237–63. [130, 131, 133, 135, 136]
- Huang, Zhiyong, and Jürgen Maurer.** 2019. "Validity of Self-Rated Memory Among Middle-Aged and Older Chinese Adults: Results From the China Health and Retirement Longitudinal Study (CHARLS)." *Assessment* 26 (8): 1582–93. eprint: <https://doi.org/10.1177/1073191117741188>. PMID: 29126348. [151]
- Idler, Ellen L., and Yael Benyamini.** 1997. "Self-Rated Health and Mortality: A Review of Twenty-Seven Community Studies." *Journal of Health and Social Behavior* 38 (1): 21–37. [132]
- Jenicek, Milos, Robert Cleroux, and Michel Lamoureux.** 1979. "Principal component analysis of four health indicators and construction of a global health index in the aged." *American Journal of Epidemiology* 110 (3): 343–49. [133]
- Julier, Simon J., and Jeffrey K. Uhlmann, editors.** 1997. *New extension of the Kalman filter to non-linear systems.* International Society for Optics and Photonics. [163]
- Jürges, Hendrik.** 2007. "True health vs response styles: exploring cross-country differences in self-reported health." *Health Economics* 16 (2): 163–78. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/hec.1134>. [132]
- Kalman, Rudolph Emil.** 1960. "A new approach to linear filtering and prediction problems." 0021-9223, [160]
- Kapteyn, Arie, James Banks, Mark Hamer, James P Smith, Andrew Steptoe, Arthur Van Soest, Anemarie Koster, and Saw Htay Wah.** 2018. "What they say and what they do: comparing physical activity across the USA, England and the Netherlands." *J Epidemiol Community Health* 72 (6): 471–76. [131]
- Karzmark, Peter.** 2000. "8." *Internation journal of geriatric psychiatry* Validity of the serial seven procedure (15): [138]
- Kasper, Judith D, Kitty S Chan, and Vicki A Freedman.** 2017. "Measuring physical capacity: an assessment of a composite measure using self-report and performance-based items." *Journal of aging and health* 29 (2): 289–309. [131]
- Latham, Kenzie, and Chuck W. Peek.** 2012. "Self-Rated Health and Morbidity Onset Among Late Midlife U.S. Adults." *Journals of Gerontology: Series B* 68 (1): 107–16. eprint: <https://academic.oup.com/psychsocgerontology/article-pdf/68/1/107/1694524/gbs104.pdf>. [132]
- Lindeboom, Maarten, and Eddy Van Doorslaer.** 2004. "Cut-point shift and index shift in self-reported health." *Journal of health economics* 23 (6): 1083–99. [133]

- Maurer, Jürgen, Roger Klein, and Francis Vella.** 2011. "Subjective health assessments and active labor market participation of older men: evidence from a semiparametric binary choice model with nonadditive correlated individual-specific effects." *Review of Economics and Statistics* 93 (3): 764–74. [133]
- McCammon, Ryan J., Gwenith G. Fisher, Halimah Hassan, Jessica D. Faul, Willard L. Rodgers, and David R. Weir.** 2022. "Health and Retirement Study – Imputation of Cognitive Functioning Measures: 1992–2018 Data Description." Working paper. Version 7.0. Survey Research Center University of Michigan. [138]
- McFadden, Daniel.** 2008. "Human capital accumulation and depreciation." *Applied Economic Perspectives and Policy* 30 (3): 379–85. [129]
- Montero-Odasso, Manuel, Joe Verghese, Olivier Beauchet, and Jeffrey M Hausdorff.** 2012. "Gait and cognition: a complementary approach to understanding brain function and the risk of falling." *Journal of the American Geriatrics Society* 60 (11): 2127–36. [130]
- Nakazato, Yuichi, Tomoko Sugiyama, Rena Ohno, Hirofumi Shimoyama, Diana L Leung, Alan A Cohen, Riichi Kurane, Satoru Hirose, Akihisa Watanabe, and Hiromi Shimoyama.** 2020. "Estimation of homeostatic dysregulation and frailty using biomarker variability: a principal component analysis of hemodialysis patients." *Scientific Reports* 10 (1): 1–12. [133]
- Niccoli, Teresa, and Linda Partridge.** 2012. "Ageing as a Risk Factor for Disease." *Current Biology* 22 (17): R741–R752. [130]
- Poterba, James M, Steven F Venti, and David A Wise.** 2017. "The asset cost of poor health." *Journal of the Economics of Ageing* 9: 172–84. [133]
- Prvan, Tania, and M. R. Osborne.** 1988. "A Square-Root Fixed-Interval Discrete-Time Smoother." *Journal of the Australian Mathematical Society. Series B. Applied Mathematics* 30 (1): 57–68. [149, 166]
- Raabe, Tobias.** 2020. "A Python tool for managing scientific workflows." [149]
- Rockwood, Kenneth, and Arnold Mitnitski.** 2007. "Frailty in relation to the accumulation of deficits." *Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 62 (7): 722–27. [133]
- Runge, Shannon K.** 2015. "Word Recall: Cognitive Performance Within Internet Surveys." *JMIR Mental Health* 2 (2): [138]
- Salthouse, Timothy.** 2010. "Selective review of aging." *Journal of the International Neuropsychological Society : JINS* 16 (09): 754–60. [131]
- Salthouse, Timothy.** 2012. "Consequences of age-related cognitive declines." *Annual review of psychology* 63: 201. [131]
- Särkämö, Teppo, and David Soto.** 2012. "Music listening after stroke: beneficial effects and potential neural mechanisms." *Annals of the New York Academy of Sciences* 1252 (1): 266–81. eprint: <https://nyaspubs.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1749-6632.2011.06405.x>. [139]
- Särkämö, Teppo, Mari Tervaniemi, Sari Laitinen, Anita Forsblom, Seppo Soinila, Mikko Mikkonen, Taina Autti, Heli M. Silvennoinen, Jaakko Erkkilä, Matti Laine, Isabelle Peretz, and Marja Hi-**

- etanen.** 2008. "Music listening enhances cognitive recovery and mood after middle cerebral artery stroke." *Brain* 131 (3): 866–76. eprint: <https://academic.oup.com/brain/article-pdf/131/3/866/907374/awn013.pdf>. [139]
- Schiele, Valentin, and Hendrik Schmitz.** 2021. "Understanding cognitive decline in older ages: The role of health shocks." *Ruhr Economic Papers*, (919): [134]
- Steiber, Nadia.** 2016. "Strong or Weak Handgrip? Normative Reference Values for the German Population across the Life Course Stratified by Sex, Age, and Body Height." *PLOS One* 10 (11): [136]
- Van Der Merwe, Rudolph.** 2004. "Sigma-Point Kalman Filters for Probabilistic Inference in Dynamic State-Space Models." Doctoral dissertation. OGI School of Science & Engineering at Oregon Health & Science University. [162]
- van der Merwe, Rudolph, and Eric A. Wan, editors.** 2001. *The square-root unscented Kalman filter for state and parameter-estimation*. IEEE. [149, 166]
- World Bank.** 2018. *World development report 2019: The changing nature of work*. The World Bank. [129]
- Zhang, Fuzhen.** 1999. *Matrix Theory: Basic Results and Techniques*. Universitext (Berlin. Print). Springer. [163, 165]