ESSAYS ON TECHNOLOGICAL CHANGE AND INEQUALITY

Leon Hubert Huetsch

A DISSERTATION

in

Economics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2024

Supervisor of Dissertation

Dirk Krueger

Walter H. and Leonore C. Annenberg Professor in the Social Sciences and Professor of Economics

Graduate Group Chairperson

David Dillenberger, Professor of Economics

Dissertation Committee

Jesús Fernández-Villaverde, Professor of Economics

Joachim Hubmer, Assistant Professor of Economics

ESSAYS ON TECHNOLOGICAL CHANGE AND INEQUALITY

ACKNOWLEDGEMENT

I'm highly indebted to my committee and advisors, Dirk Krueger, Joachim Hubmer, Jesús Fernández-Villaverde, Alan Taylor and Alexander Ludwig, who supported me and my research continuously. I'm particularly thankful to Dirk Krueger who always took the time to speak to me about research, carefully read and scrutinize my writing, and to ensure that the Department supported me financially throughout my graduate studies. My interest in the distributional consequences of structural change largely originated from his mentorship. I also thank Víctor Ríos-Rull for always making time to speak to me, for his critical comments and for challenging my views.

I would like to express my deepest gratitude to my many colleagues and friends who made life in Philadelphia interesting, fun and the Ph.D. easier. They provided invaluable support over the years and made me a happier and more open-minded person. Among them Min, Daniel, Javiera, Marlon, Aaron, Kathleen, Yoshiki, Sara and Agustin. I would like to especially acknowledge Marko, Priyanka, and Joao, who became close friends, who are among the most interesting people I have met, and who taught me a lot about Economics and more about life.

Finally, and most importantly, I thank my family and Naomi. They were my foundation.

ABSTRACT

ESSAYS ON TECHNOLOGICAL CHANGE AND INEQUALITY

Leon Hubert Huetsch

Dirk Krueger

This thesis consists of three chapters and studies the macroeconomics of technological change and inequality, in particular the distributional consequences of technological and structural change.

In the first chapter, I ask how labor adjustment costs, specifically in the form of unionization, shape the evolution of wages and employment of workers exposed to labor replacement by automation. I argue that, by raising adjustment costs, unions generate intergenerational redistribution by shifting the impact from existing, older to incoming, younger cohorts, and further generate aggregate effects by accelerating overall labor reallocation from automating to non-automating occupations. The reason is that labor adjustment costs incentivize firms to adjust through hiring rather than layoffs, and to reduce labor in anticipation of future adoption. Using variation across local labor markets in the U.S. since 1980, I document that unionization among exposed workers is associated with greater wage and employment decline among young relative to older workers, and with accelerated overall employment decline. I then develop an overlapping generations model of technological change and unionization that rationalizes the empirical findings through the impact of union-imposed labor adjustment costs on firms' choice how to transform their workforce over time when gradually adopting automation. Within automating occupations, unions reduce the welfare cost of automation of older workers along the transition by up to 4% of permanent consumption while raising the welfare costs of cohorts entering during the transition by up to 2%. Incoming workers endogenously respond to automation by entering non-adopting occupations. The union effect spills over into non-adopting occupations as the accelerated labor reallocation depresses wages there.

The second chapter, joint with Dirk Krueger and Alexander Ludwig, poses the question what caused the increase in life expectancy since 1800 and the rapid growth of a modern health sector during the

20th century in the United States. We document the evolution of life expectancy over the last two centuries and the emergence of the modern health sector in the 20th century in the U.S. We then provide a quantitative theory of the joint dynamics of income growth, the modern health sector, and the increase in life expectancy over the last two centuries to explain the documented facts. The theory is built on the insight that the demand for health increases over time as individuals become richer and older, which in turn sparks a reallocation of resources towards the production and innovation of health goods. Households are initially too poor to demand health goods, and life expectancy is stagnant. As income grows, fueled by technological progress, households start consuming basic health goods, life expectancy starts to rise, and directed technological progress eventually, with a delay of 100 years, leads to the emergence of a modern health sector. We find that rising household demand for health accounts for one-third of the observed increase in the relative price of health goods, while two-thirds are accounted for falling input prices in the modern health sector relative to the final goods sector, driven by technological progress. Moreover, modern health goods have accounted for roughly 30% of the increase in life expectancy at age 20 since 1940, which translates into 3.3 additional expected years of life.

In the third chapter, I study how much rising risk in labor earnings matter for wealth inequality and welfare. I answer this question by introducing higher-order earnings risk consistent with recent empirical findings into a benchmark heterogeneous-agent model. I show that higher-order earnings dynamics in the form of left-skewness and excess kurtosis strengthen the precautionary savings motive, leading to greater consumption inequality and lower wealth inequality. The earnings dynamics are partially passed through to the consumption of poor households who are willing to pay up to 1.7% of permanent consumption to avoid higher-order earnings risk. Methodologically, I develop a new General Polynomial Chaos Expansion approach, a global solution method to solve for the aggregate dynamics of this class of models, and demonstrate that it increases efficiency relative to previous methods. I extend the baseline method to allow for time-varying base distributions, which is particularly useful in economic settings in which the cross-sectional household distribution at times moves far away from the ergodic distribution. I then apply the extension by introducing time-varying earnings risk into the benchmark model.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1

TECHNOLOGICAL CHANGE AND UNIONS: AN INTERGENERATIONAL CONFLICT WITH AGGREGATE IMPACT

1.1. Introduction

Technological change improves productivity and standards of living but creates winners and losers among workers. One of the most prominent technological changes in recent decades is the adoption of automation technologies, which boost productivity but can temporarily disrupt labor markets for transitional generations through worker displacement and reduced earnings (Graetz and Michaels (2018), Humlum (2021), Acemoglu and Restrepo (2020)). Increased adoption of automation technologies, such as industrial robots and artificial intelligence, has spurred an active literature studying the labor market impact on workers and discussing appropriate policy responses, most notably taxing automation (Beraja and Zorzi (2022), Costinot and Werning (2023)).

The existing literature on the labor market impact of automation assumes that firms can freely adjust their workforce, abstracting from labor adjustment frictions that firms face. Yet, the adoption of automation technologies entails substantial labor adjustment, making such frictions especially potent during the transition. Moreover, it is well documented that labor market institutions and employment protection laws that generate labor adjustment frictions are empirically associated with reduced employment flows and increased capital deepening (Bassanini and Duval (2006)), as well as with raising unemployment particularly during times of economic turbulence (Ljungqvist and Sargent (1998, 2008)).

In this paper, I ask how labor adjustment costs that firms face shape the impact of labor-replacing technological change on workers across different generations by studying the effect of unions on the evolution of employment and wages of workers who are substitutable with automation technologies. I argue that, by raising labor adjustment costs, unions shape the transformation of workforces in two ways. First, labor adjustment costs incentivize firms to adjust through hiring rather than layoffs, thereby shifting the adverse labor market impact from existing, older to incoming, younger

cohorts and generating intergenerational redistribution effects. Second, labor adjustment costs incentive firms to shrink their workforce in anticipation of future automation adoption in order to smooth labor adjustment over time, thereby accelerating overall labor reallocation and generating aggregate employment effects.

To support this argument, I start by providing empirical evidence showing that unionization among workers exposed to automation is associated with, first, larger employment and wage decline among young relative to older workers and, second, accelerated overall employment decline among these exposed workers. To further strengthen the empirical evidence, I build a simple theoretical model that illustrates the effect of labor adjustment costs on the evolution of the age composition of workers during a labor-replacing technological transition and show that the empirical findings are consistent with the predictions of the model. I then develop a quantitative dynamic equilibrium model of unionization and technological transitions to first show that the proposed mechanism can quantitatively account for the documented intergenerational redistribution and aggregate employment effects of unionization. I then use the model to study how unionization, and labor adjustment costs more broadly, shapes the intergenerational distribution effects of automation in terms of welfare.

In the empirical exercise, I focus on the wages and employment of workers in routine manual occupations, which have been particularly exposed to labor-replacing technologies, such as automation (Goos et al. (2014)). I combine data from several sources to exploit variation in unionization and the evolution of worker outcomes within such occupations across local labor markets in the U.S. since 1980.[1] First, I find that unionization is associated with a greater fall in employment and wages among young workers entering the labor market relative to older workers, consistent with insider-outsider dynamics.[2] In particular, comparing a low-unionized to a high-unionized labor market implies that the routine manual employment share of young workers below the age of 30 falls by an additional 11% and their wage falls by an additional 9% relative to the average routine

[1] I define a local labor market in the baseline analysis as a Metropolitan Statistical Area (MSA) and check that results hold at the state level.

[2] See e.g. Carruth and Oswald (1987) who theoretically introduce insider-outsider dynamics into a standard union model.

manual wage during the first 10 years of the transition between 1980 and 1990.[3] As a result, the routine manual workforce in more unionized labor markets becomes older relative to less unionized labor markets, and I show that this relative aging persists throughout the technological transition. Second, unionization is associated with an accelerated decline in overall routine manual employment while not significantly affecting the long-run decline. In particular, I document a greater employment decline in high-unionized labor markets early in the transition, and a subsequent slow catch-up of employment decline in less unionized labor markets from 2000 onward. By 2020, the gap has mostly closed.

Motivated by these findings, I develop a quantitative dynamic equilibrium model of unionization and endogenous technological change that demonstrates that the interaction of union-imposed convex labor adjustment costs and gradual technology adoption over time can jointly rationalize the documented distributional and aggregate effects of unions. First, adjustment costs give rise to a static effect by incentivizing firms to replace their workforce through reduced hiring rather than through costly layoffs when adopting automation. Second, there is a dynamic effect. In the context of expected gradual technology adoption, firms smooth their labor adjustment along the transition by shrinking their workforce preemptively today in order to avoid adjustment costs in the future. Intuitively, any worker not hired in 1980 is a worker the firm will not have to lay off when more automation technologies are adopted later. This dynamic effect is not dependent on but particularly strong in the context of convex labor adjustment costs as convexity generates an additional incentive to smooth labor adjustments over time. As a result, the dynamic mechanism gives rise to accelerated overall employment decline in routine occupations in high relative to low unionized labor markets. Similar in spirit to the extensive literature on labor market institutions and economic turbulence, the mechanism here builds on the interaction of adjustment costs and expectations about future technology adoption in driving current labor demand.[4]

To answer my research question, I use the model as a measurement device to quantify the im-

[3] Here I define a low-unionized and a high-unionized labor market as the average MSA at the 10th and 90th percentile of routine manual unionization, respectively.

[4] See, for example, Blanchard and Summers (1986), Ljungqvist and Sargent (1998, 2008).

pact of automation and unionization on labor market outcomes and life-cycle consumption paths of different cohorts of exposed workers during the automation transition. At its core, the model combines three building blocks that make it a suitable quantitative framework for that objective: 1) firms combine routine and non-routine occupations, they endogenously automate the routine occupations over time; 2) routine workers are represented by a labor union that raises labor adjustment costs and endogenously sets their wages; and 3) overlapping generations of workers make occupational choices and accumulate occupation-specific human capital, which makes switching occupations costly, more so when old. The rate of unionization in the model is parameterized by the level of exogenous labor adjustment costs, consistent with empirical evidence that unions raise firing costs.[5] Moreover, labor adjustment costs determine the union's ability to impose wage premia by reducing the elasticity of labor demand of firms, making them an intuitive measure of unionization in the model. The union sets separate wages for routine workers of different ages as they are imperfect substitutes in production due to skill accumulation, which allows the model to speak to the observed union effect on the wage ratio between young and old workers. The two-sector setup with occupational choice endogenizes the supply of workers. This allows me to decompose the documented overall employment decline in the routine occupations into a downward shift in demand driven by technology adoption and an endogenous supply response driven by incoming workers entering and switching to non-routine occupations instead in order to avoid the automation impact. The model therefore captures the life-cycle paths of a group of workers that is affected by unionization but that is difficult to identify empirically: workers who would have entered routine occupations but, due to the combination of automation and unionization, enter non-routine occupations instead.

Firms trade off two opposing forces in their routine labor demand when deciding how to optimally adjust their workforce along the transition. First, adjustment costs incentivize adjustment through incoming workers as well as preemptively reducing the workforce in anticipation of automation adoption to avoid adjustment costs in the future. Second, routine workers of different ages are imperfect substitutes in production because they are finitely lived and accumulate occupation-

[5]See, for example, Millward et al. (1992) and Colonna (2008).

specific skills on the job which allows older workers to complete different tasks in production. Firms therefore prefer a balanced age composition of routine workers, which constrains the incentive to adjust through young, incoming workers only.

I calibrate the model to U.S. local labor market data, targeting in particular life-cycle wage profiles, the routine employment share, and the aggregate labor share in 1980 and 2010. I model a technological transition through an unexpected fall in the path of automation prices from 1980 onward that matches the price path of capital goods in the U.S. The level of adjustment costs measures the degree of unionization in the model and is calibrated to match the relative decline in the routine employment share between 1980 and 1990 in high and low-unionized MSAs. Lastly, I connect the model with the empirical findings by validating that it matches the untargeted evolution of overall routine employment and the evolution of the age composition of routine workers along the transition.

I first evaluate the impact of automation adoption on routine workers in a low-unionized labor market. Automation is most costly for incumbent routine workers who made their occupational choice without anticipating the upcoming transition. These workers are caught by surprise, facing the option to either stay in a declining sector or switch into non-routine occupations at the cost of losing their occupation-specific human capital. Especially routine workers who entered between 1970 and 1980 experience the full automation impact over their entire life-cycle, resulting in large permanent earnings losses. As a result, the welfare cost of automation to these workers reaches 10% of permanent consumption in 2000, measured as the permanent percent decrease in consumption they would be willing to accept to avoid automation and remain at the 1980 steady state. Workers entering the labor market during the automation transition take the current and future impact of automation into account when making their occupational choice. As routine jobs become less desirable, only workers with a sufficiently large labor productivity in routine tasks relative to non-routine tasks still enter the routine occupation. As a result, average labor productivity, and in turn average life-cycle earnings and consumption paths, of incoming routine cohorts rise. This endogenous response to automation limits its impact on incoming cohorts. Nevertheless, entering

routine workers would still pay up to 7% of permanent consumption to avoid automation.

I then study the automation impact in a high-unionized labor market to quantify to what extent unions reallocate the welfare cost of automation across generations. Unions protect incumbent routine workers by lowering their layoff risk and limiting their wage decline, which reduces the welfare cost of automation for incumbent cohorts by up to 4% of permanent consumption along the transition in the high relative to the low unionized labor market. However, the impact is shifted to incoming cohorts. As a result, the welfare cost of automation for incoming routine workers is up to 2% of permanent consumption larger in the high relative to the low unionized labor market, driven by falling routine entry wages. The difference in the welfare benefit for incumbent and the welfare cost for incoming cohorts reflects the ability of incoming workers to endogenously respond to automation by entering the non-routine occupation instead. Consistent with the empirical findings, high unionization causes a faster reallocation of employment from the routine to the non-routine occupation as firms in the high-unionized labor market preemptively reduce hiring in order to avoid future adjustment costs. The accelerated reallocation of labor means that non-routine wages fall relatively more in the high-unionized labor market early in the transition, resulting in a larger spillover of the automation impact from routine to non-routine occupations.

Lastly, motivated by the model findings, I empirically evaluate the political implications of the intergenerational conflict that unions magnify. An emerging political economy literature connects adverse economic outcomes to ideological realignment as well as a shift in political preferences and voting behavior (McCarty et al. (2016), Voorheis et al. (2015)). Autor et al. (2020) link trade-exposure to rising political support for strong-left and strong-right views as well as to pure rightward shifts across local labor markets in the U.S. My welfare analysis emphasizes that unionization has magnified the negative impact of automation on labor market experiences of less skilled cohorts of workers who entered routine and non-routine occupations after 1980. Cohorts of workers that have entered the labor market between 1980 and 2000 are in their 50s and 60s today, thus, the workers whose voting behavior has shifted (Pew Research Center (2014, 2017)). I test and find empirical support for the hypothesis that the union-induced employment decline among young routine manual

workers in the 1980s across local labor markets is associated with a shift in voting from Democrats to Republicans in the 2016 and 2020 presidential elections relative to previous elections. This suggests that while unions protected incumbent workers from the adverse automation impact, this also induces worsening labor market experiences for incoming workers and a shift in political preferences among these workers today.

Literature. This paper contributes first and foremost to the extensive empirical and quantitative literature studying the labor market impact of automation and its contribution to rising inequality (Acemoglu (2002) Goldin and Katz (2008)). Several papers provide empirical evidence on the effect of industrial robots adoption on worker outcomes and productivity (Graetz and Michaels (2018), Acemoglu and Restrepo (2020), Humlum (2021), Koch et al. (2021), Bessen et al. (2023)), documenting that robot adoption raises productivity, output and the wage bill of skilled workers while reducing the wages and employment share of less skilled workers. Similarly, across U.S. commuting zones, Acemoglu and Restrepo (2020) find negative effects on wages and employment that are more pronounced in routine manual and blue-collar occupations. A large empirical body of work specifically documents the decline of routine employment since 1980.[6] I contribute to this literature by studying the role of labor adjustment costs, specifically in the form of unionization. I show that labor adjustment costs shape the intergenerational distribution effects of automation as well as the timing of aggregate labor reallocation and put forward as the underlying mechanism the dynamic effects of adjustment costs on firms' decision how to optimally adjust their workforce over time. I thereby emphasize that, in the context of labor adjustment costs, automation is a dynamic choice because expectations about future workforce adjustments drive current labor demand. Consistent with Acemoglu and Restrepo (2020) who document negative spillover effects of robot adoption on nontradable sectors such as construction and services, my model suggests that as unionization induces faster labor reallocation to non-adopting occupations, it also accelerates the negative spillover effects on wages in these occupations.

A second body of work studies the intergenerational distribution effects of structural changes, such

[6]See, for example, Autor et al. (2003), Goos and Manning (2007), Autor (2010), Cortes et al. (2020).

as trade or automation, and the resulting aggregate transition dynamics (Guerreiro et al. (2022), Costinot and Werning (2023), Beraja and Zorzi (2022)). Similar to me, several papers emphasize the role of occupation-specific human capital (Adao et al. (2024), Dvorkin and Monge-Naranjo (2019), Guren et al. (2015), Traiberman (2019)). Through the lens of an estimated occupational choice model, Traiberman (2019) argues that the costs of switching occupations in response to trade shocks is largely driven by the resulting loss of occupation-specific human capital rather than other switching costs, such as retraining. Related to me, Adao et al. (2024) and Guren et al. (2015) focus on the role of different generations of workers during technological transitions within overlapping generation frameworks. Guren et al. (2015) focus on the dynamics of labor markets in response to trade shocks and study the role of sector-specific human capital in driving worker decisions to reallocate across sectors, and, thus, labor mobility. Adao et al. (2024) focus on transitions triggered by the arrival of new technologies and argue that transitions are slower when innovations require learning new skills as labor adjustment is then more strongly driven by entering cohorts rather than labor reallocation of existing, older workers. Focusing on automation, I also emphasize the importance of occupation-specific human capital of existing workers and the occupational choices by incoming cohorts in determining the incidence of wage and employment decline across generations of exposed workers and for the transitional dynamics. I contribute to this strand of the literature by developing a structural model of the intergenerational distribution effects of automation to quantify the welfare consequences of automation across generations. Moreover, I incorporate imperfect labor market competition to study how unions shift the distributional consequences of technological change.

Lastly, this paper also contributes to the extensive body of research trying to reconcile the rise and persistence of high European unemployment since the 1970s compared to the U.S. labor market. The literature has identified the interaction between shocks and labor market institutions as a key factor for the transatlantic gap in unemployment.[7] In their seminal work Ljungqvist and Sargent (1998, 2008) argue that labor market institutions in Europe, particularly policies of employment

[7]See, among others, Blanchard and Summers (1986), Lindbeck and Snower (1988), Ljungqvist and Sargent (1998, 2008), Haan et al. (2005), Baley et al. (2020, 2023).

protection that increase the cost of layoffs, reduced employment flows in the 1950s and 1960s when the economic environment was calm, thereby lowering frictional unemployment. However, economic turbulence starting in the 1970s limited the reemployment options for laid-off workers due to old human capital becoming obsolete and, as a result, labor market institutions in the form of generous unemployment benefits then increased unemployment by reducing the incentive of laid off workers to accept wage cuts in their new jobs. My paper is closely related to this broad literature, also emphasizing the interaction between labor market institutions, here in the form of labor adjustment costs, and economic change, here in the form of automation adoption. I contribute to this literature in two ways. First, I focus on the impact on employed workers and labor reallocation rather than unemployment. Second, I emphasize the expectation about future economic turbulence rather than current turbulence in driving current labor demand decision, thus, the dynamic effects of labor market institutions.

The remainder of the paper is organized as follows: Section 2 documents the empirical findings. Section 3 develops the model and discusses its elements. Section 4 takes the model to the data by outlining the calibration strategy and validating the model output. Section 5 presents the main quantitative analysis. Section 6 briefly presents evidence on the political implications of the model findings and Section 7 concludes.

1.2. Empirical Analysis

This section documents the two main empirical findings on the effect of unions on the wages and employment of workers exposed to labor-replacing technologies. I start this section by describing the data sources and outlining the empirical approach.

Data. I exploit variation in unionization across local labor markets in the US. In the main analysis, a local labor markets is defined as a metropolitan statistical area (MSA).[8] I use public use micro data from the 1980, 1990, 2010, and 2019 American Community Survey (ACS) to construct population, employment and wage income measures at the MSA level as well as at the industry and occupation level within MSAs at those four dates.[9] I further use the US Current Population Survey (CPS) to

[8] I provide additional evidence for robustness across states.
[9] See Ruggles et al. (2010).

compute unionization by MSA. I focus on workers in routine manual (RM) occupations and follow the recent literature in the classification of occupations based on their routine- and manual-task content.[10] These occupations focus on tasks that follow a well-defined set of instructions and, as a result, can more readily be performed by automation technologies. Routine employment, classified as such, has fallen significantly since 1980, and progress in labor-replacing automation technology has been identified as a main driver (Autor et al. (2003), Goos et al. (2014)). Throughout the empirical analysis, worker outcomes of interest as well as unionization are measured within routine manual occupations and refer to routine manual workers unless otherwise specified. Unionization is computed as the share of routine manual workers who are either a member of or covered by a labor union, averaged between 1995 and 2005. I use the average unionization rate over 10 years starting in 1995 to have better coverage across MSAs. I then validate that the results hold when using different unionization measures, and in particular earlier measures going back to 1986. Finally, I use exposure to robots estimates from Acemoglu and Restrepo (2020).

The empirical strategy is to study if routine manual workforces across MSAs, who differ in their rate of unionization but are similar otherwise, experience differential changes in the overall level and composition of their employment and wages from 1980 onward when automation technologies increasingly became available in all MSAs. Thus, I estimate to what extent differential decline in employment and wages among routine manual workers across MSAs can be explained by variation in unionization among routine manual workers, controlling for the ex-ante exposure to labor-replacing technologies within MSAs. First, in order to account for the ex-ante exposure to automation, that is, the expected amount of automation which is unrelated but potentially correlated with unionization, I construct a rich set of controls at 1980, prior to the transition. In particular, I control for the industry composition at the MSA level, the industry composition within routine manual occupations in each MSA, and the demographic composition of routine manual workers in the MSA. Lastly, I add the exposure to automation measure by Acemoglu and Restrepo (2020), which is a commuting zone level measure that combines the industry composition of a commuting zone with industry-

[10] The literature goes back to Berman et al. (1994), Levy and Murnane (1996), Autor et al. (1998). See Katz and Autor (1999) for a summary of the early literature and Cortes et al. (2020) for a classification.

level adoption of industrial robots between 1993 and 2014 at the national level. I aggregate the measure to the MSA level using 1980 population weights. The identification assumption is that the remaining variation in unionization among routine manual workers is exogenous to ex-ante exposure to technology adoption and changes in the age composition of workers conditional on adoption.

I use different measures of changes in employment and wages in routine manual occupations since the start of the transition as outcome variables. In particular, for the change in variable y after t years of the transition, I estimate the following model across MSAs i

$$\Delta y_{i,1980+t} = \beta_0 + \beta_1 \text{Unionization}_i + \gamma X_{i,1980} + u_{i,t}, \tag{1.1}$$

where $\Delta y_{\text{i},1980+t}$ is the realized change in y between in 1980 and $1980+t$ (e.g. the decline in routine manual employment). The set of controls is constructed in 1980, prior to the transition, except for the exposure measure from Acemoglu and Restrepo (2020), which is based on adoption data between 1993 and 2014. I then run this model for different outcome variables y and at different stages of the transition, $t \in [10, 30, 40]$, to understand the effect of unionization on the level as well as timing of changes.

1.2.1. The Aggregate Effect of Unionization

In order to understand the aggregate effect of unionization, that is, the effect on the MSA-level routine manual employment share, I look at the timing and extent of overall employment decline. In particular, I regress the decline in the routine manual employment share in a MSAs since 1980 on its routine manual unionization rate and the set of controls. I do so for the decline until 1990, 2010 and 2019.

Table 1.1: The dependent variable is the change in the routine manual employment share since 1980, the independent variable is routine manual unionization. The regression further includes the set of controls outlined above.

	Change in RM share		
	1990	2010	2019
	(1)	(2)	(3)
Unionization	-0.080^{***}	-0.040^{**}	-0.036^{*}
	(0.018)	(0.018)	(0.020)
Mean dependent	-0.062	-0.11	-0.11
Observations	147	147	147
R^2	0.712	0.629	0.554
Adjusted R^2	0.684	0.592	0.510
Note:	*p<0.1; **p<0.05; ***p<0.01		

Table 1.1 displays a negative effect of unionization on the change in employment, meaning employment falls more in high-unionized labor markets. Importantly, the effect is large between 1980 and 1990 and then falls off thereafter. In order to understand the size of the effect, I plot the change in the routine manual employment share when going from the 10th to the 90th percentile of unionization across MSAs. Thus, the graph below plots the estimated coefficient of the union effect, scaled by the difference in unionization between an MSA at the 90th and an MSA at the 10th percentile of unionization, which is a 29 percentage point difference in the unionization rate.

Figure 1.1: The graph shows the effect of going from the 10th to the 90th percentile of unionization on the RM employment share over time. The results hold for the 25th and the 75th percentile, see Appendix A.1.3.1 for details.

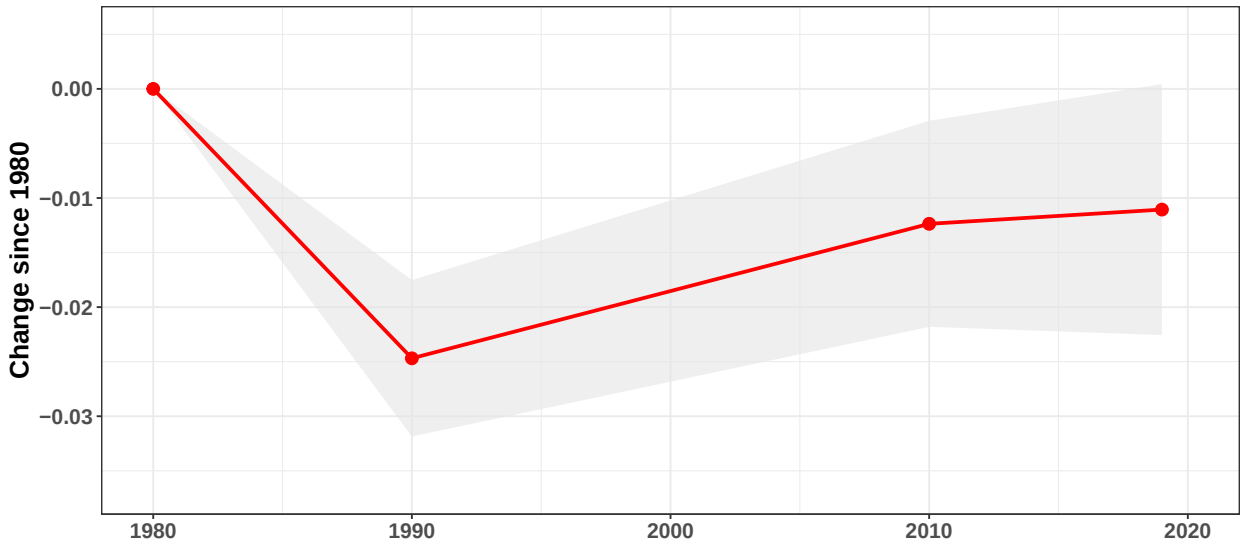
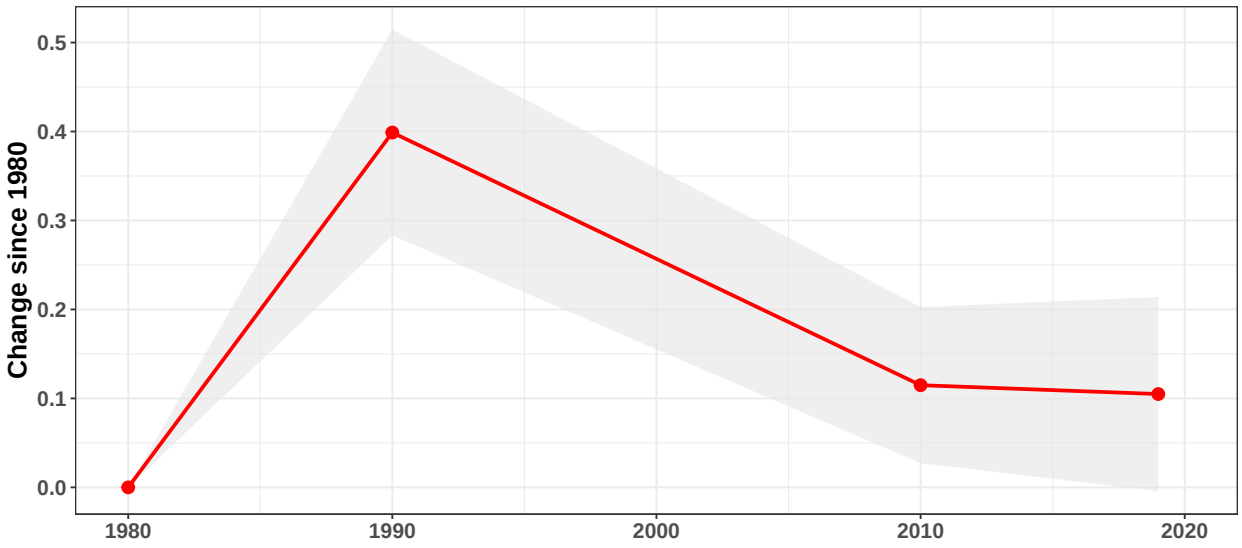


Figure 1.1 shows that unionization is associated with an accelerated decline in employment from 1980 onward across MSAs. Going from the 10th to the 90th percentile in the rate of unionization leads to an additional decline of 2.5 percentage points in the employment share of exposed occupations between 1980 and 1990. Throughout the transition, the decline in the employment share then catches up in less unionized labor markets. In 2019, the union effect has fallen to roughly 1 percentage and is insignificant. Figure 1.2 relates the union effect to the average decline in the routine manual employment share across all MSAs, dividing the above union effect by the average decline across all MSAs in the same time period.

Figure 1.2: The graphs show the effect of going from the 10th to the 90th percentile of unionization on the RM employment share over time, relative to the mean decline in the RM employment share across all MSAs.



The union effect is large when relating it to the average decline across MSAs. In particular, comparing the 10th to the 90th percentile of unionization implies an additional fall in the employment share of 40% of the average decline across MSAs during the first 10 years. The union effect then falls to roughly 10% until 2019 and becomes insignificant, that is, employment decline in less unionized labor markets catches up over time. Thus, unionization among exposed workers is associated with a substantial acceleration of employment decline early in the transition. By 2019, the gap in employment decline by unionization is insignificant and has mostly closed with less unionized labor markets exhibiting a modestly smaller fall in their routine manual employment share.

1.2.2. The Distributional Effect of Unionization

An extensive literature has documented the insider-outsider dynamics of employment regulation and organized labor, as model by Carruth and Oswald (1987) and the literature thereafter. In particular, employment protection is associated with greater job security for older, incumbent workers, and reduced the employment opportunities and wage prospects for younger workers

(Bassanini and Duval (2006), Botero et al. (2004)). This section documents to what extent the insider-outsider dynamics were prevalent during the decline of routine manual employment since the 1980s. Insider-outsider dynamics predict that unionization induces a downward shift in the demand for young, incoming workers. In turn, the downward shift in demand then translates into a fall in the price and quantity of young workers in high relative to low-unionized places. Guided by this intuition, I test whether unionization among routine manual workers is associated with a larger decline in wages and employment among young relative to older routine manual workers since 1980.

Throughout the analysis, I control for the decline in the routine manual employment share in order to measure how the composition of routine manual workers changed conditional on a decline in employment.

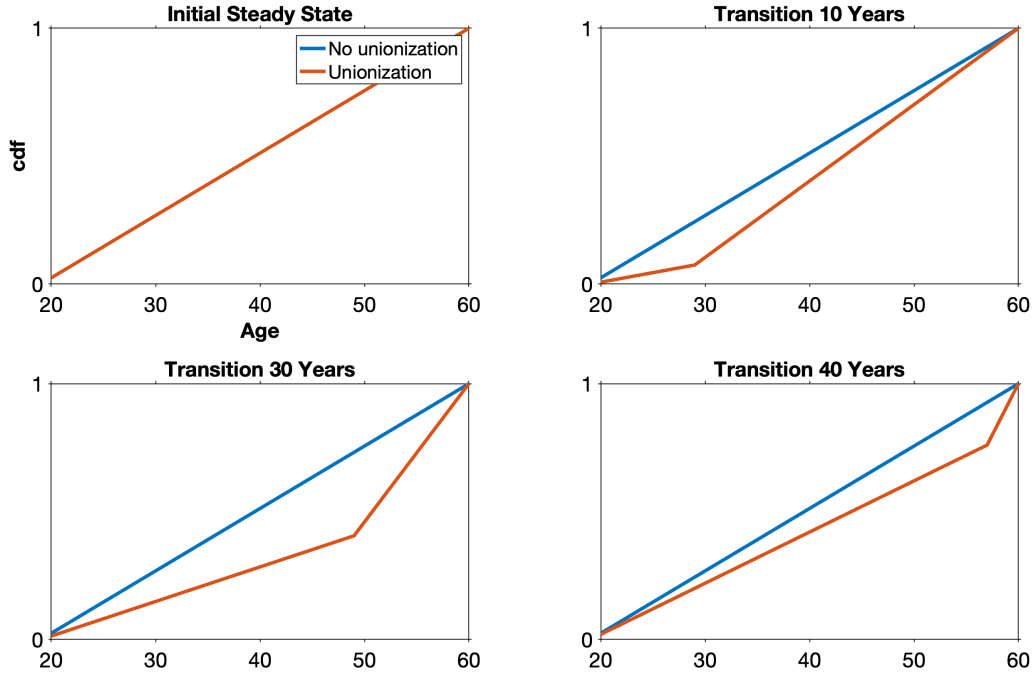
1.2.2.1 Employment Effect for Young and Old Workers

To quantify the impact of unionization on the relative employment of young and old workers, I measure the union effect on the age composition of routine manual workers along the transition. To build intuition and derive precise predictions to test in the data, I start by developing a simple model that isolates the effect of the hiring and layoff margin, and thereby illustrates how the age composition of workers evolves over time if the fall in employment is to a larger extent driven by reduced inflow (hiring) of young workers rather than increased outflow (layoffs) across the age distribution.

There are two labor markets, A and B, which are initially in steady state with an identical and uniform age distribution of homogeneous workers aged 20 to 60. Each year the 60 year old retire and are replaced by an inflow of 20 year old. In labor market A firms face zero labor adjustment costs, representing low or no unionization. By contrast, labor market B is unionized and firms face infinite labor adjustment costs. In 1980, an unexpected shock hits both labor markets which forces firms to shrink their workforce, firms in labor market A respond with uniform layoffs across the age distribution while firms in labor market B respond by lowering their hiring rate as layoffs are infinitely costly. Figure 1.3 shows the cdf of the age distribution of workers in both labor markets

along the transition from a simple simulation.

Figure 1.3: CDF of age distribution in high and low-unionized labor markets in simple model.

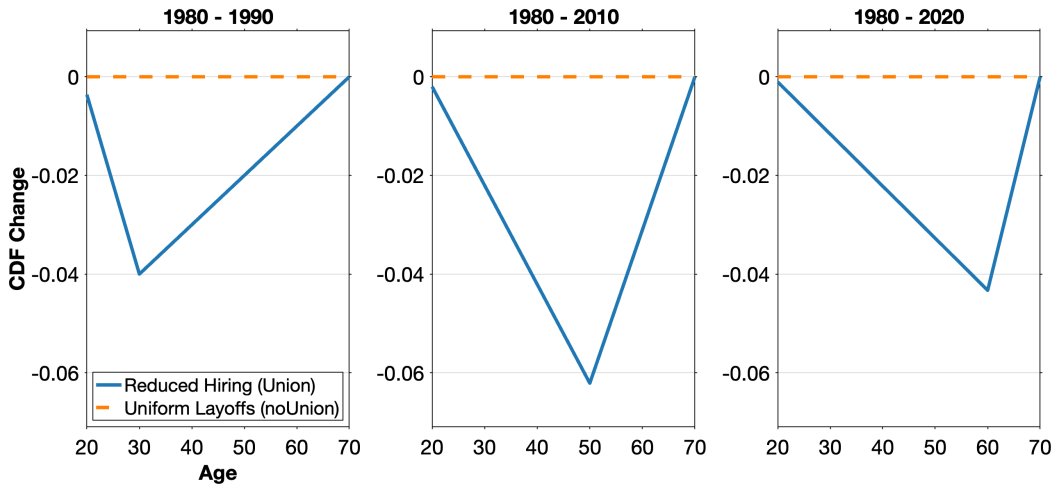


While the age composition in labor market A never changes, the reduction in hiring in labor market B leads to a fall in the share of young workers and a slow transition as the age composition adjusts. The workforce in labor market B ages relative to A as more old workers remain, which results in a downward shift of its CDF compared to labor market A. 10 years into the transition, the downward shift is largest at age 30, which is the first cohort that experienced reduced hiring. All cohorts between the age of 20 and 30 experienced reduced hiring while all cohorts above the age of 30 did not, and thus the share of workers below age 30 has fallen most relative to the steady state. The downward shift then evolves along the transition. In particular, the largest downward shift moves up the age ladder with the first cohort to experience reduced hiring as it ages over the course of the transition. 30 years into the transition the share of workers below the age of 50 has shifted down the most as all cohorts younger than age 50 have experienced reduced hiring. Thus, the simple

model makes two detailed predictions about the effect of labor adjustment costs that drive relative changes in hiring and layoffs to test in the data. First, the routine manual workforce in more unionized MSAs becomes relatively older during the transition, measured as a relative downward shift in the CDF across all ages. Second, the downward shift is largest for the cohorts who entered around 1980 and moves up the age ladder with that cohort over time.

To account for the fact that routine manual workforces in all MSAs experience a mixture of reduced inflow and increased outflow in the data, I estimate the union effect on the downward shift in the CDF relative to 1980. Figure 1.4 shows the downward shift in the age distribution in each labor market relative to their initial steady state levels in the thought experiment.

Figure 1.4: Change in CDF of age distributions relative to steady state in simple model.



To test the predictions for the age composition of routine manual workers, I estimate the gap between the orange and blue line with the following model:

$$\text{CDF(a)}_{\text{i},t} - \text{CDF(a)}_{\text{i},1980} = \Delta\text{CDF(a)}_{i,t} = \beta_0 + \beta_1^{\text{a},t} \cdot \text{Unionization}_{\text{i}} + \gamma X_{\text{i},t} + u_{\text{i},t}. \tag{1.2}$$

The coefficient $\beta_1^{\text{a},t}$ estimates the gap between the orange and blue line t years into the transition at age a. To account for differences in the initial age distributions, I control for the 1980 age composition among routine manual workers. Moreover, to isolate the insider-outsider dynamic

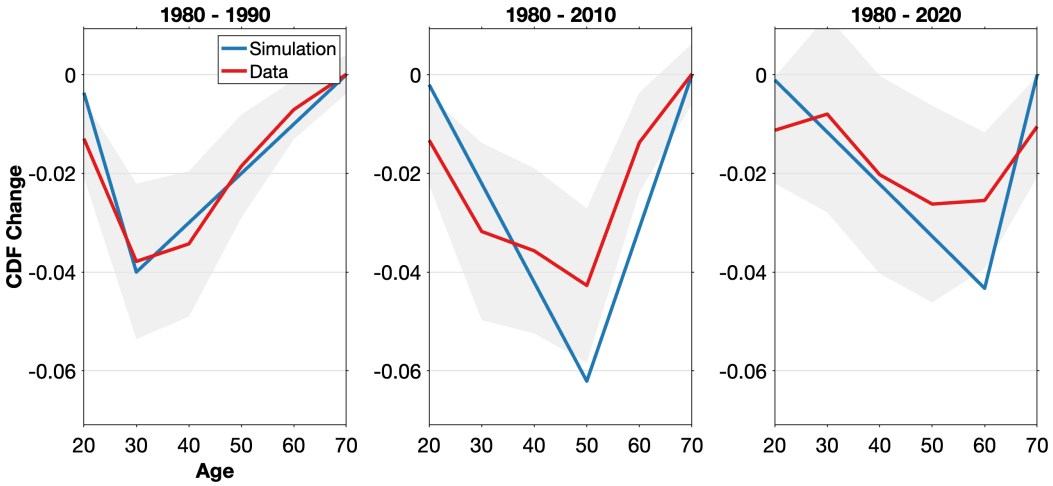
from the aggregate effect, I further control for the decline in the routine manual employment share between 1980 and t ($\Delta \text{RM}_{i,t}$).

Table 1.2: The table shows the effect of unionization on the change in the age distribution of routine manual workers between 1980 and different stages of the transition (1990, 2010, 2019). See A.1.3.2 for the full regression tables.

	Dependent variable: Change in CDF across Ages				
	Age 20	Age 30	Age 40	Age 50	Age 60
	(1)	(2)	(3)	(4)	(5)
CDF Change 1980-1990	-0.043^{***}	-0.126^{***}	-0.114^{***}	-0.062^{***}	-0.023^{**}
	(0.012)	(0.027)	(0.026)	(0.020)	(0.011)
CDF Change 1980-2010	-0.044^{***}	-0.106^{***}	-0.119^{***}	-0.142^{***}	-0.046^{***}
	(0.014)	(0.035)	(0.028)	(0.030)	(0.017)
CDF Change 1980-2019	-0.037^{**}	-0.026	-0.067^{**}	-0.087^{***}	-0.084^{***}
	(0.017)	(0.029)	(0.033)	(0.031)	(0.024)
Note:	*p<0.1; **p<0.05; ***p<0.01.				

Table 1.2 displays the coefficient $\beta_1^{\text{a},t}$ measuring the union effect across the age distribution (columns) at different times in the transition (rows). First, the union effect is significant and negative throughout. Consistent with the first prediction, unionization is associated with a downward shift in the CDF of workers relative to 1980 across all ages. That means, more unionized workforces have become older relative to less unionized workforces along the transition as the share of young workers has fallen and more older workers have remained in unionized workforces. Second, between 1980 and 1990 the downward shift is largest at young ages and peaks at age 30. Over time, the downward shift moves up the age ladder with the cohorts who entered around 1980, consistent with the second prediction. Figure 1.5 constructs the graphs from the simple model from the regression estimates to directly compare the results with the prediction and to understand the magnitude of the effects. In particular, it again plots the union effect when comparing the average MSA at the 10th with the average MSA at the 90th percentile of routine manual unionization.

Figure 1.5: The plot shows the shift in the age distribution (CDF) relative to 1980 when going from the average MSA at the 10th percentile of routine manual unionization to the average MSA at the 90th percentile of routine manual unionization. The difference in unionization is 29 percentage points. The results hold for the 25th and the 75th percentile, see Appendix A.1.3.3 for details.



The plot closely resembles the prediction from the model. To understand the magnitude, note that the share of workers below the age of 30 falls by roughly 4 percentage points in the high-unionized MSA relative to the low-unionized MSA during the first 10 years of the transition. This translates into an additional 11% decline relative to the 1980 share of workers below the age of 30.

To summarize, conditional on reducing employment, unionzation is associated with a larger reduction in employment of young workers and a smaller reduction in employment of old workers, resulting in more unionized workforces becoming older relative to less unionized workforces throughout the transition. This is consistent with labor adjustment costs on incumbent workers imposed by unionization, which incentivize firms to adjust relatively more through young and incoming workers from 1980 onward. The initial decline in the employment share of young workers in more unionized workforces has then translated into persistent changes in the age composition over time as the middle and right panel show.

1.2.2.2 Wage Effect for Young and Old Workers

A downward shift in the demand for young workers driven by unionization should further lead to a fall in the price of young workers, that is, their wage. To quantify the differential effect of unions on wages of young, incoming and older, incumbent workers, I look at the changes in the wage ratio between young to old routine manual workers.[11] Table 1.3 displays the results of regressing the change in the wage ratio between 1980 and 1990 on unionzation as well as the set of controls. The wage ratio is measured as the average wage of workers below the age of 30 divided by the average wage of workers over the age of 30 in the first two columns, and divided by the average wage of workers over the age of 50 in the last two columns. Columns 2 and 4 additionally control for the overall decline of the routine manual employment share between 1980 and 1990.

Table 1.3: Table shows the effect of unionization on the change in the wage ratio between young and older routine manual workers between 1980 and 1990. The first two columns define the wage ratio as the average wage of workers below the age of 30 divided by the average wage of workers over the age of 30. In the last two columns, the wage ratio is measured as the average wage of workers below the age of 30 divided by the average wage of workers over the age of 50.

	Change in Wage Ratio 1980-1990			
	$\Delta \frac{\text{Wage age} \leq 30}{\text{Wage age} > 30}$		$\Delta \frac{\text{Wage age} \leq 30}{\text{Wage age} > 50}$	
	(1)	(2)	(3)	(4)
Unionization	-0.184^{***}	-0.175^{**}	-0.307^{***}	-0.289^{***}
	(0.069)	(0.069)	(0.096)	(0.096)
Change RM-share 1980s		0.250		0.525
		(0.290)		(0.374)
Mean dependent	0.032		0.024	
Observations	200	200	200	200
R^2	0.282	0.285	0.281	0.289
Adjusted R^2	0.244	0.243	0.243	0.247
Note:	*p<0.1; **p<0.05; ***p<0.01.			

Unionization is negatively correlated with a change in the wage ratio, that is, in more unionized

[11] Note, since I display a ratio, there is no need to deflate the wages by prices.

workforces the wages of young workers have declined more relative to wages of older workers compared to less unionized workforces. Looking at the first two columns, a 1 percentage point increase in unionzation is associated with a 0.18 percentage point decline in the wage ratio between workers below and above age 30. The effect rises to a roughly 0.3 percentage point decline for the wage ratio between workers below age 30 and above age 50, shown in the last two columns. To put this into perspective, going from the 10th to the 90th percentile of unionization is then associated with a 6 and 9 percentage point decline in the wage ratio between 1980 and 1990 in the first and last two columns, respectively. This effect is quantitatively large as the average unconditional change in the wage ratio across local labor markets is even slightly positive with 0.032 and 0.024 percentage point increases, respectively. Thus, while the wage ratio between young and older routine manual workers has been stable on average across MSAs, wages of young workers have declined significantly relative to wages of older workers in unionized local labor markets. The effect grows larger when conditioning the group of older workers to higher ages, shown by the last two columns relative to the first two columns, which is consistent with union protection rising in tenure and age.

To the extent that the effects measured above are driven by a union induced fall in the demand for less protected young workers during the transition, the relative decline in wages of young workers should be a temporary effect during the transition rather than a persistent change in the wage structure. Table 1.4 looks at the union effect on changes in the wage ratio between 1980 and 2010. Over a longer horizon of the transition, between 1980 and 2010, the effect vanishes and unionization is not associated with an additional decline in the wage ratio between young and old workers.

Table 1.4: Table shows the effect of unionization on the change in the wage ratio between young and older routine manual workers between 1980 and 2010. The first two columns define the wage ratio as the average wage of workers below the age of 30 divided by the average wage of workers over the age of 30. In the last two columns, the wage ratio is measured as the average wage of workers below the age of 30 divided by the average wage of workers over the age of 50.

	Change in Wage Ratio 1980-2010			
	$\Delta \frac{\text{Wage age}\leq 30}{\text{Wage age}>30}$		$\Delta \frac{\text{Wage age}\leq 30}{\text{Wage age}>50}$	
	(1)	(2)	(3)	(4)
Unionization	−0.037	−0.029	0.030	0.035
	(0.096)	(0.094)	(0.117)	(0.116)
Change RM-share 1980-2010		0.837*		0.437
		(0.434)		(0.550)
Mean dependent	0.023		-0.062	
Observations	200	200	200	200
R^2	0.125	0.147	0.167	0.170
Adjusted R^2	0.069	0.088	0.114	0.112
Note:	*p<0.1; **p<0.05; ***p<0.01.			

To summarize, unionization is associated with a larger fall in employment and wages of young workers during the first decades of the transition. These effects are temporary and vanish by 2019, consistent with a fall in demand for young workers during the initial adjustment phase of the transition. Importantly, while the employment share of young workers in more unionized routine manual labor markets recovers by 2019, the initial fall in their employment moves up the age ladder over time and thereby has a persistent effect on the age composition of routine manual workers.

1.3. The Model

Motivated by the empirical findings, I develop a quantitative dynamic equilibrium model that interprets the documented distributional and aggregate effect of unions jointly through the lens of union-imposed labor adjustment costs interacting with gradual and endogenous technology adoption by firms over time. After validating that the model can replicate the different transitions observed in high and low unionized labor markets, I use it as a measurement device to quantify

the welfare cost of automation for routine workers and the intergenerational transfer that unions give rise to during technological transitions. I first outline the model and provide a more detailed discussion of the model choices and properties in section 1.3.7.

1.3.1. Overview

Time is discrete and one period corresponds to 10 years. The model is a small open economy without aggregate uncertainty, combining three core elements. First, firms produce the final good by combining output from non-routine and routine occupations while endogenously and gradually adopting automation in routine production as capital prices fall. Second, overlapping generations of workers make an occupational choice between routine and non-routine occupations based on their expected life-cycle wage paths in each occupation. Third, a monopoly union represents incumbent routine workers by posting the wage schedule for routine workers of different skills, and thus ages, each period, taking labor demand of firms into account. The level of labor adjustment costs parameterizes the rate of unionization as the union derives its ability to impose wage premia from labor adjustment costs which limit how much firms can reduce employment in response.

1.3.2. Job levels

In the routine occupation, firms, the union, and workers interact through job levels. In practice, job levels describe the specific task requirements of each job. An extensive literature on the internal labor markets (ILMs) and the production process of firms has documented the importance of job levels in the design of production processes.[12] In particular, job levels are a key input in production, and progression across job level accounts for the majority of life-cycle wage growth of workers (Bayer and Kuhn (2023), Pierce (1999)). Moreover, unions directly bargain for wages at different job levels (Bayer and Kuhn (2023)). Based on these findings, routine production in this economy is organized around job levels, and firms decide how many workers to employ at each job level. Young workers entering the routine occupation are hired at the lowest job level and progress in job levels by accumulating human capital on the job. Lastly, the union sets the job level wage profile in the routine occupations.

[12] See, for instance, Doeringer and Piore (1985), Baker et al. (1994), Pierce (1999), Strub et al. (2008) Bayer and Kuhn (2023). Baker and Holmstrom (1995) provide an overview of the early literature.

1.3.3. Production

Technology. There is a continuum of perfectly competitive firms that produce the final consumption good by combining the output y_t^i from two occupations i, the non-routine and the routine occupations, with a CES production technology G according to:

$$y_t = G(y_t^R, y_t^N) = \left[\phi(y_t^R)^\nu + (1-\phi)(y_t^N)^\nu\right]^{\frac{\theta}{\nu}}, \tag{1.3}$$

where ϕ is the share of automatable (routine) occupations and $(1-\phi)$ is the share of non-automatable (non-routine) occupations in the economy. $\nu < 1$ is the elasticity of substitution between routine and non-routine occupations. To accommodate decreasing returns to scale from convex adjustment costs while abstracting from firm heterogeneity, I assume that firms need to use land as another input in production which is in fixed and limited supply L, as in Huo and Ríos-Rull (2020). Without loss of generality, I assume there is a total of one unit of land $L = 1$ and there is a firm operating each unit of land. $\theta < 1$ measures the returns to scale and the land is then priced by the value of the representative firm.

The non-routine occupations use homogenous labor input N_t to operate a constant returns to scale technology given by:

$$y_t^N = N_t. \tag{1.4}$$

Routine Production. The routine occupations use automation α_t, and workers at $J = 5$ different job levels $l_t = (l_{t,1}, ..., l_{t,J})$ to produce routine output with a CES production technology F given by:

$$y_t^R = F_t(l_{t,1}, ..., l_{t,J}, \alpha_t) = \left[\sum_{j=1}^{J} \eta_j l_{(t,j)}^\varphi + \eta_\alpha \alpha_t^\varphi\right]^{\frac{1}{\varphi}}, \tag{1.5}$$

where $\eta = (\eta_1, ..., \eta_J, \eta_\alpha)$ governs the share of automation and job level input, and $\varphi < 1$ is the elasticity of substitution between routine inputs.[13]

Workers accumulate the necessary skills to produce the tasks required at higher job levels on the job. In particular, I assume the technology is such that it takes workers one period (10 years) to learn the skills to work the next job level. That is, a worker on job level j in period t can be promoted to job level $j+1$ in period $t+1$. I restrict attention to employment contracts between firms and routine entrants that commit the firm to compensate the rising human capital path of workers or otherwise terminate the contract at firing cost c_f. Thus, when hiring a routine worker, firms commit to progressing workers one job level per period consistent with their accumulated skill or firing them otherwise. As a result, age becomes a sufficient statistic for job level which makes the firm and worker problem tractable.

Optimization. Taking the path for the non-routine wage and the routine wages across job levels, $(w_t^N, \{w_{t,j}^R\}_j)_t$, as well as the price of automation $(p_t)_t$ as given, the firm chooses period t automation α_t and labor demands $(\{l_{t,j}\}_j, N_t)$ to maximize the discounted sum of future profits:

$$W_t(\{l_{t-1,j}\}_{j=1}^J) = \max_{\alpha_t, \{l_{t,j}\}_j, N_t} G(y_t^R, y_t^N) - \sum_{j=1}^J w_{t,j}^R l_{t,j} - p_t \alpha_t - w_t^N N_t - \sum_{j=1}^J c_f(f_{t,j}) + \frac{1}{1+r_t} W_{t+1}(\{l_{t,j}\}_j), \tag{1.6}$$

$$\text{s.t.} \qquad f_{t,j} = l_{(t-1,j-1)} - l_{t,j} \quad \forall j \geq 2,$$
$$c_f(f_{(t,j)}) = c \cdot f_{(t,j)}^2,$$

where $f_{t,j} = l_{t-1,j-1} - l_{t,j}$ denotes fired workers at job level j. Firing costs are specified the same way across job levels and parameterized by c. Thus, labor adjustment costs in this model are governed by c. I assume firing is a lottery, thus, firms decide how many workers to fire at each job level but which workers at job level j are fired is random.

[13] Bayer and Kuhn (2023) document five possible job levels in a German Employment Survey (BIBB/BAuA) and 15 job levels for the United States using the National Compensation Survey (NCS), consistent with Pierce (1999).

1.3.4. Households

Agents and Preferences. The economy is populated by overlapping generations of households. Each period a measure one of young households is born who live for 5 periods, from age 20 to 70. Thus, in every period there is a total of 5 generations alive. Young workers choose which occupation to work in and spend resources on consumption and saving while supplying labor inelastically and accumulating human capital on the job. Workers born in period t maximize expected lifetime utility U_t given by:

$$U_t = \sum_{a=1}^{5} \beta^{a-1} E\left[u(c_{t+a-1,a})\right], \tag{1.7}$$

where the period utility function $u(c)$ is at least twice continuously differentiable with $u'(c) > 0$ and $u''(c) < 0$, and satisfies the lower Inada condition, thus $\lim_{c\to 0} u'(c) = \infty$.

Human capital accumulation process. Workers are born and enter the labor market with initial routine labor productivity z^R and non-routine labor productivity z_1^N. Initial non-routine labor productivity z_1^N is ex-ante identical across workers while routine labor productivity z^R differs across workers and is drawn from distribution f_z. Human capital is occupation-specific and deterministically accumulates on the job in both occupations.

Households working in the non-routine occupation accumulate human capital each period in the form of labor productivity. They move up a discrete labor productivity ladder $(z_1^N, z_2^N, z_3^N, z_4^N, z_5^N)$, which is calibrated to match average life-cycle wage paths of non-routine workers in the data.

Workers in the routine occupation accumulate human capital on the job through job level progression which drives their life-cycle wage growth. In particular, routine workers who are not laid off move up one job level per period as specified in their employment contract. A worker's routine labor productivity z^R is a permanent type that applies to all job levels. As a result, human capital in the routine occupation follows a step function. Job level progression captures steps over the life-cycle which are common across workers and give rise to wage dispersion across age. z_R captures the overall level of the step function which differs across workers and gives rise to wage dispersion within age groups. Routine workers have perfect information about the endogenous probability of

being laid off in the routine occupations.

Occupational Choice. At labor market entry, workers only differ in their permanent routine labor productivity type z^R which determines their initial occupational choice. They take into account the expected life-cycle path of earnings in each occupation. There is no aggregate uncertainty, thus, workers have full information about the future path of wages in each occupation $(\{w^R_{t,j}\}_j, w^N_t)$. They face individual uncertainty in the form of firing risk when working in the routine occupations. The probability of being fired at each job level is endogenously chosen by firms but is fully known by workers. In each consecutive period, workers choose whether to switch or stay in their current occupation. Routine workers who are laid off switch to the non-routine occupations and stay there for the remainder of their life. I assume they cannot reenter the routine occupations after being laid off.

Assets. Financial markets are incomplete, in particular routine workers cannot trade contingent assets against the risk of being laid off. Households have access to risk-free bonds at world interest rate R which is exogenous and constant. In the baseline model, the land, and thus firms, is owned by risk-neutral capitalists who receive the firm dividends. In practice, equity participation is limited, especially for low and medium skilled workers who usually are less educated (Mankiw and Zeldes (1991)), and 90% of firm equity is held by the 10% wealthiest households in the U.S. (Survey of Consumer Finances, (2022)). Since the model is designed to capture the impact of automation on the subset of less skilled workers who are most exposed to displacement by automation technologies, I take as baseline an economy in which these workers do not hold equity and are therefore impacted through wages but not through profits. In appendix A.2.1, I show results for the case when workers hold fixed equity shares in the firms.

The Worker Problem. At the beginning of period t, the state of a worker is her age a, wealth b, labor productivities (z^R, z^N), and previous occupation, s. There is no incentive for a worker to initially enter the non-routine occupation and then switch to the routine occupation later. Thus, the problem of a worker previously employed in the non-routine occupation, $s = 1$, can be simplified and solved by imposing the occupational choice to stay in the non-routine occupation. I verify in equilibrium that non-routine workers do not want to switch. The problem then consists only of the

consumption-savings decision for the remainder of her life given by:

$$V_t^{hh}(k, z^R, z_i^N, a, s=1) = \max_{c,k'} u(c) + \beta V_{t+1}^{hh}(k', z^R, z_{i+1}^N, a+1, s'=1), \tag{1.8}$$
$$\text{s.t.} \quad c + k' = w_t^N z_i^N + (1 + r_t)k,$$
$$k' \geq 0.$$

The problem of a worker previously employed in the routine occupation, $s = 0$, is more complicated as it involves the discrete choice about whether to stay a routine worker or switch into the non-routine occupation. It can be formulated as a two-stage problem. In the first stage, the household decides on the occupation s', in the second stage the household makes a consumption-savings decision conditional on the realization of the occupational choice. Note that since routine workers face an endogenous, possibly positive probability of being fired, a worker may decide to stay in the routine occupation but is fired and nevertheless forced to switch occupations. Conditional on working in occupation s' after stage 1, the stage 2 problem of a worker previously employed in the routine occupation is then given by:

$$v_t^{hh}(k, z^R, z^N, a, s=0|s') = \max_{c,k'} u(c) + \beta V_{t+1}^{hh}(k', z^R, z^N, a+1, s'), \tag{1.9}$$
$$\text{s.t.} \quad c + k' = s' w_t^N z^N + (1 - s') w_{t,j(a)}^R z^R + (1 + r_t)k,$$

where $w_{t,j(a)}^R$ is the routine wage at job level $j(a) = \frac{a}{10} - 1$ which maps the age of workers $a = (20, 30, 40, 50, 60)$ into their job level $j = (1, 2, 3, 4, 5)$.

For routine workers who either decide to switch into the non-routine occupation or who are fired, the above stage 2 problem is the same as the beginning of period problem of workers who were previously employed in the non-routine occupation:

$$v_t^{hh}(k, z^R, z^N, a, s=0|s'=1) = V_t^{hh}(k, z^R, z^N, a, s=1). \tag{1.10}$$

Given the value in stage 2, one can solve for the occupational choice in stage 1 of a routine worker. I assume workers face choice specific taste shocks to smooth the discrete occupation choice, $\sigma_s \epsilon_t(s)$. The taste shocks are additively separable and follow an extreme value distribution, as in McFadden (1973) and the literature thereafter. Due to firing, the realized occupation s' can differ from the chosen occupation $\tilde{s}'$, which solves the stage 2 problem given by:

$$\begin{aligned} V_t^{hh}(k, z^R, z^N, a, s=0) = \max_{\tilde{s}'} \Bigg\{ & \tilde{s}' \Big(V_t^{hh}(k', z^{R\prime}, z^{N\prime}, a+1, s'=1) + \sigma_s \epsilon_t(\tilde{s}'=1) \Big) \\ & + (1-\tilde{s}') \Big(\mu_{t,j(a)} V_t^{hh}(k', z^{R\prime}, z^{N\prime}, a+1, s'=1) + \sigma_s \epsilon_t(\tilde{s}'=1) \\ & + (1-\mu_{t,j(a)}) V_t^{hh}(k', z^{R\prime}, z^{N\prime}, a+1, s'=0) + \sigma_s \epsilon_t(\tilde{s}'=0) \Big) \Bigg\}, \end{aligned} \tag{1.11}$$

where $\mu_{t,j(a)}$ denotes the probability of being fired from the routine occupation as a worker at job level $j(a)$. The probability of deciding to stay a routine worker is given by the discrete choice policy function, $\mathcal{P}_t(\tilde{s}'|k, z^R, z^N, a)$, which is equal to the standard logit choice probability with an extreme value distributed taste shock:

$$\mathcal{P}_t(\tilde{s}'|k, z^R, z^N, a) = \frac{\exp\left(V_t^{hh}(k', z^{R\prime}, z^{N\prime}, a+1, \tilde{s}')/\sigma_s\right)}{\exp\left(V_t^{hh}(k', z^{R\prime}, z^{N\prime}, a+1, =0)/\sigma_s\right) + \exp\left(V_t^{hh}(k', z^{R\prime}, z^{N\prime}, a+1, 1)/\sigma_s\right)}. \tag{1.12}$$

Based on the discrete choice policy function, the probabilities for the realization of the occupational choice, $P_t(s'=1|k, z^R, z^N, a)$, that accounts for endogenous firing is then given by:

$$P_t(s'=1|k, z^R, z^N, a) = \mathcal{P}_t(\tilde{s}'=1|k, z^R, z^N, a) + \mu_{t,j(a)} \mathcal{P}_t(\tilde{s}'=0|k, z^R, z^N, a), \tag{1.13}$$

$$P_t(s'=0|k, z^R, z^N, a) = (1-\mu_{t,j(a)}) \mathcal{P}_t(\tilde{s}'=0|k, z^R, z^N, a). \tag{1.14}$$

1.3.5. The Union

All incumbent workers in the routine occupations are represented by a labor union. The union acts within a standard monopoly union framework by setting wages as a monopolist while firms choose

labor demand in response.[14] However, I extend the basic monopoly union model by allowing the union to set the full job level wage profil to account for the fact that workers at different job levels are imperfect substitutes in production and therefore can have different wages. The union then chooses wage growth across job levels by setting the full job level wage profile in the current period, and seeks to maximize the total wage bill paid to its current members:

$$\Theta_t = \max_{\{w^R_{t,j}\}_j} \sum_{j=2}^{J} w^R_{t,j} l_{t,j}. \tag{1.15}$$

Thus, the union is constrained by the employment response of firms $\{l_{t,j}\}_j$ which is endogenous to the wage schedule, $\{w^R_{t,j}\}_j$, posted by the union.

The rate of unionization in the model is measured by the level of firing costs c. The union's ability to extract rents from firms by raising wage premia is determined by the elasticity of labor demand of firms. If the elasticity is high, firms respond to wage premia by reducing their labor input which drives down the wage bill. Firing costs reduce the demand elasticity and thereby increase the ability of the union to increase the wage bill through wage premia, thus, providing a measure of the degree of unionization in the model. In practice, unions obtain higher firing costs for their members, consistent with the union model here.[15]

1.3.6. Competitive Equilibrium

I focus on perfect foresight equilibria in which there is no aggregate uncertainty.

Definition 1 (Competitive Equilibrium). Given a path for automation prices p_t and interest rates r_t, and an initial worker distribution Φ_0, a competitive equilibrium consists of paths for non-routine wages w^N_t, routine wages $(w^R_{t,1}, ..., w^R_{t,J})$, firm policies $(l_{t,1}, ..., f_{t,J}, \alpha_t, N_t)$, worker policies $V_t, c_t, k_{t+1}, \mathcal{P}_t$, and the worker distribution Φ_t that satisfy for all $t \geq 0$:

1. Given the paths for prices $\{r_t, w^N_t, (w^R_{t,1}, ..., w^R_{t,J})\}_{t\geq 0}$, and the firm implied firing probabilities $\{(\mu_{t,1}, ..., \mu_{t,J})\}_{t\geq 0}$, $V_{t\geq 0}$ solves the optimization problem of workers and $\{c_t, k_{t+1}, \mathcal{P}_t\}_{t\geq 0}$ are

[14] The basic monopoly union framework goes back to Fellner (1949) and Cartter (1959).
[15] See, for example, Parsons (2005a,b,c), Millward et al. (1992) and Colonna (2008).

the corresponding decision rules.

2. Given the paths for prices $\{r_t, w_t^N, (w_{t,1}^R, ..., w_{t,J}^R)\}_{t\geq 0}$, $W_{t\geq 0}$ solves the optimization problem of the firm and $\{l_{t,1}, ..., f_{t,J}, \alpha_t, N_t\}_{t\geq 0}$ are the corresponding policies.

3. Given the paths for prices $\{r_t, w_t^N, (w_{t,1}^R, ..., w_{t,J}^R)\}_{t\geq 0}$, the period t routine wage schedule $\{w_{t,1}^R, ..., w_{t,J}^R\}_t$ solves the period t union problem.

4. The non-routine wage clears the labor market for non-routine workers:

$$N_t = \int z^N \mathrm{d}\Phi_t(k, z^R, z^N, a, s = 1) + \int z^N \mathrm{d}\Phi_t(k, z^R, z^N, a, s = 0) P_t(s' = 1 | k, z^R, z^N, a, s = 0). \tag{1.16}$$

5. The labor market for incoming routine workers clears:

$$l_{t,1} = \int z^R \mathrm{d}\Phi_t(k, z^R, z^N, a = 1, s) P_t(s' = 0 | k, z^R, z^N, a = 1, s). \tag{1.17}$$

6. The law of motion of the worker distribution is induced by the optimal decisions of the firm and workers.

1.3.7. Discussion

This section discusses the two key model elements, the job-level based routine production process and the union model.

Job Levels. Routine production in this model is organized around job levels. In practice, job levels categorize jobs by explicitly describing specific task requirements of jobs along the dimensions of responsibilities, complexity, and autonomy. This builds on an extensive literature that studies internal labor markets (ILMs) of firms and career dynamics. It emphasizes the role of the organizational structure of firms, and, in particular, the importance of job levels in the design of the production process.[16] One of the main insights going back to Doeringer and Piore (1985) and

[16] See the seminal works of Doeringer and Piore (1985) and Baker et al. (1994), as well as the literature thereafter.

confirmed by the subsequent literature is that "in many jobs in the economy, wages are not attached to workers, but to jobs." (Doeringer and Piore (1985, p. 77)). Based on that idea, the literature documents two findings with respect to the determinants of wages and wage growth that make job levels a suitable modeling choice in this context. First, life-cycle wage growth is largely driven by job level progression over time (Baker et al. (1994), Dohmen et al. (2004), Bayer and Kuhn (2023)). Second, unions bargain for wages and benefits at the level of job levels (Bayer and Kuhn (2023)).

In the model, routine workers accumulate skill on the job and, as a result, move up in job levels. Thus, this yields a standard process of human capital accumulation as wages rise in response to skill accumulation. Consistent with the empirical evidence on how unions operate, the union sets wages at the level of job levels in the model, taking into account labor demand. As a result, the job level model yields life-cycle wage growth that reflects a standard human capital accumulation process as well as an endogenous and time-varying union wage premium.

From the firm's perspective, the job level model allows for workers with different experience levels to be imperfect substitutes in production because they perform different tasks, which gives rise to an endogenous and time-varying wage ratio between younger and older routine workers that can be mapped to the data. This aligns with the task-based approach (Autor et al. (2003)), which builds on the idea that wages are determined by the tasks that a worker performs in their job. As a result, the firm optimally produces with a range of workers of different experience levels and therefore of different age. I restrict the analysis to employment contracts with full utilization of human capital, meaning firms commit to progressing workers to the next job level each period as their experience accumulates. This makes the model tractable as age becomes a sufficient statistic for job levels, and it allows me to abstract from the managerial decision about which workers to promote and when to promote. In practice, unions similarly bargain for wage growth in employment contracts, and, thus, for contractual commitments by the firm to compensate rising human capital over time or terminate the contract otherwise.

The Union. The model of the labor union here follows the literature initiated by Dunlop (1944) whose starting point is the microeconomic theory of firms. The labor union is modeled as an eco-

nomic entity that maximizes an economic objective, such as the wage bill, while facing constraints, in particular the labor demand of the firm. The literature thereafter largely uses two frameworks. First, monopoly union models in which the union acts as a monopolist and imposes its wage policy while the firm chooses employment in response. Second, since the 1980s game-theoretic bargaining frameworks were developed. These frameworks were developed with the intend to properly model the sources of bargaining power, such as strikes, and the bargaining process. In both cases, the union generally imposes a wage premium, resulting in an inefficient outcome in which the wage is above and employment is below their market clearing levels. While in the bargaining framework variation in the bargaining power offers a direct model analog to the empirically observed variation in the rate of unionization, the difficulty of introducing bargaining here stems from the fact that it requires specifying the value of the disagreement outcome. Therefore, I abstract from the bargaining process and model the union within a monopoly union framework. Importantly, the solution to bargaining frameworks in which the firm and union bargain over wages coincides with the monopoly union model when the union has all the bargaining power. I then use the level of firing costs as a measure of unionization instead of an explicit bargaining power. Empirically, unions obtain higher firing costs for members (see, for instance, Parsons (2005a,b,c), Millward et al. (1992), Colonna (2008)), through bargaining for higher severance pay and the ability to impose strike costs. In the model, the union maximizes the wage bill of its members by extracting rents from firms. Since the union acts as a monopolist, its ability to extract rents is driven by the elasticity of labor demand of firms. Firing costs reduce the demand elasticity and, thus, increase the ability of the union to expand the wage bill through wage premia. Thus, firing costs drive the strength of the union and thereby provide an intuitive measure of the rate of unionization in the model.

1.4. Quantitative Evaluation

In this section, I outline the calibration strategy before evaluating the quantitative behavior of the model. I then connect the model back to the empirical findings by validating that it replicates the untargeted distributional and aggregate union effects along the transition.

1.4.1. Calibration

I calibrate the initial steady state of the model to MSA-level data in the U.S. in 1980. I then explore the response of the economy to an unexpected fall in the path of automation prices from 1980 to 2010 that matches the decline in capital prices observed in the U.S., as measured by Hubmer (2023). In particular, agents in the economy learn in 1980 about the complete future path of automation prices, thus, there is no aggregate uncertainty. The timing is motivated by the fact that existing measures of capital prices show a more rapid decline from the 1980s onward (Hubmer (2023)), and the adoption of industrial robots has picked up from 1990 onward (Acemoglu and Restrepo (2020)). This is also consistent with the observed fall in the routine employment share and the manufacturing labor share from 1980 onward (Hubmer (2023), Cortes et al. (2020)).

I take the low-unionized labor market as an economy that is characterized by low firing costs, and calibrate that economy to the average MSA at the 10th percentile of unionization. The high-unionized labor market corresponds to a MSA at the 90th percentile of unionization with high firing costs. I calibrate the common parameters in the low-unionized labor market, which is the baseline economy. A subset of parameters is calibrated exogenously, either following direct empirical observation or the existing literature. The remaining parameters are estimated in the model using the method of simulated moments. Since the objective is to use the model as a measurement device to quantify the impact of automation and unionization on the consumption paths of different workers, the calibration aims in particular at matching three sets of targets: First, the 1980 and 2010 routine manual employment share to capture the amount of workers that are exposed to displacement by automation technologies as well as the amount of employment loss they experience along the transition. Second, the 1980 and 2010 aggregate labor share to capture the fact that while automation raises productivity and thereby output, the share of output that accrues to labor has declined over time. Third, life-cycle wage profiles of workers in both occupations.

Data. I estimate the targeted data moments using the same data as in the empirical section. Thus, I construct MSA-level estimates by combining public use micro data from the American

Community Survey (ACS), and the Current Population Survey (CPS).[17] I take the remaining targets from the existing literature and indicate when I do so.

Share paramaters in the production technology. The share parameters in the production function are calibrated to match the employment and labor shares. In particular, I calibrate the share parameter ϕ for routine output to match an initial routine manual employment share in 1980 of 27%. The share parameter of automation in the routine technology, η_α, is calibrated to match an initial aggregate labor share in 1980 of 64%, which I take from the U.S. Bureau of Labor Statistics (BLS). Lastly the share parameters of job level inputs in the routine technology, $(\eta_1, ..., \eta_5)$, are calibrated to match life-cycle wage profiles in 1980, the estimation of life-cycle wage profiles is outlined below.

Substitution elasticities in the production technology. The substitution elasticities are calibrated to match the change in the routine employment and aggregate labor share between 1980 and 2010. I calibrate the substitution elasticity across the two occupations, ν, to match the routine manual employment share in 2010 of 16%. The substitution elasticity across routine inputs, φ, is calibrated to match an aggregate labor share in 2010 of 56% (BLS).

Firing costs. The firing cost schedule is pinned down by one parameter c. Recall that the low-unionized economy is characterized by a low level of firing costs, c_l, which I calibrate to match the change in the average age of routine manual workers in MSAs at the 10th percentile of unionization between 1980 and 1990. The level of firing costs in the high-unionized economy, c_h, is then calibrated to match the documented union effect on the decline in the routine manual employment share between 1980 and 1990. That is, I target that the routine employment share in the high-unionized economy declines by 2.5 percentage points more between 1980 and 1990 than in the low-unionized economy.

Human capital accumulation. Using repeated cross-sectional data from the CPS, I estimate life-cycle wage profiles by decomposing earnings growth into cohort, experience and time effects for cohorts born between 1940 and 1980, following Heckman et al. (1998), and more recently Lagakos et al. (2018) and Fang and Qiu (2023). The estimated experience effects capture the com-

[17] See Ruggles et al. 2010

ponent of life-cycle wage growth that is driven by human capital accumulation. I then estimate experience effects separately for routine manual and non-routine occupations to calibrate life-cycle wage paths in both sectors in the model.[18] Workers in the non-routine occupation all enter with the same labor productivity z_1^N, and accumulate labor productivity every period on the job. I calibrate the life-cycle path of non-routine labor productivity, $z^N = (z_1^N, z_2^N, z_3^N, z_4^N, z_5^N)$, to match the estimated experience effect, and normalize mean labor productivity, $\bar{z}^N = 1$.

Preferences. The period utility function of households is given by Constant Relative Risk Aversion (CRRA) utility:

$$u(c_t) = \frac{c^{1-\sigma}}{1-\sigma}$$

where $\sigma = 2$ to get an intertemporal elasticity of substitution of 0.5. I set the discount factor to $\beta = 0.75$; recall that one model period corresponds to 10 years, and thus, the annualized discount factors of households is $\beta_{\text{annualized}} = 0.97$.

Remaining parameters. The world interest rate is set to 3% annually, based on estimates of the natural rate of interest for the U.S. from Davis et al. (2024). I abstract from the fact that real rates started to decline particularly from 2000 onward and keep the interest rate constant over the transition. Note that savings play a minor role in this model as households have rising life-cycle wage paths, do not face retirement, and face permanent income risk in the form of a small layoff probability in the routine occupation. The initial automation price $p_{1980} = 0.12$ is fixed in a first-stage calibration. The automation share η_α is then calibrated conditional on $p_{1980} = 0.12$ as the two are not separately identified.

[18] See appendix A.1.1 for details on the estimation.

Table 1.5: Externally calibrated parameters.

Parameter	Description	Value	Target
	Preferences		
$1/\sigma$	IES	0.5	Standard
β	Discount factor	0.75	$\beta_{\text{annualized}} = 0.97$
σ_s	Taste shocks	0.05	Small - smooth occ choice
	Human capital		
z^N	Non-routine labor productivity		Life-cycle wage profile
$\bar{z}^N$	Mean labor productivity	1	Normalization
	Small open economy		
r	Rate of return	0.34	3% annual Davis et al. (2024)
p_{1980}	Automation price 1980	0.12	Normalization
g_p	Growth rate of price	-0.06	Hubmer (2023)

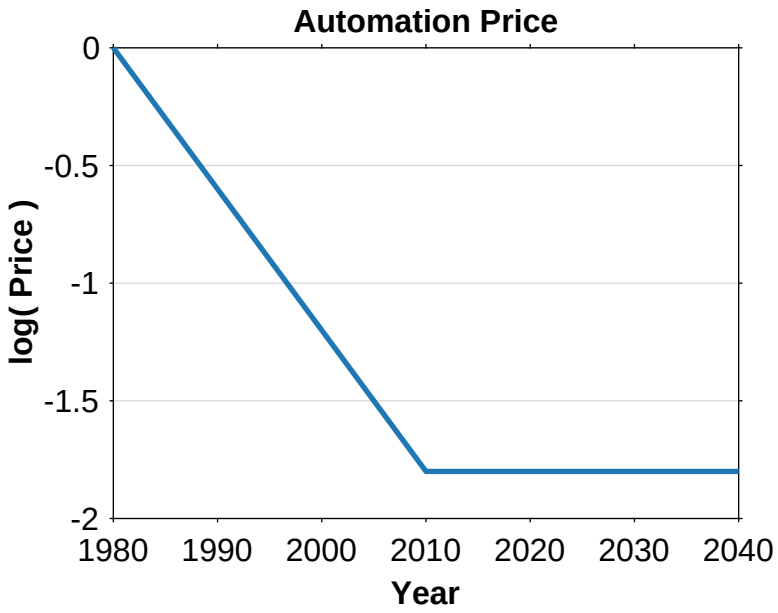
Table 1.6: Internally calibrated parameters.

Parameter	Description	Value	Target
	Production and technology		
f_z	Routine labor productivity	$\mathcal{U}(0.2, 1.8)$	Routine Wage Dispersion
ϕ	Share of automatable occupations	0.75	1980 RM employment share
θ	Returns to scale	0.8	Land-output ratio
η_l	Job level shares		Life-cycle wage profile
η_α	Automation share	0.3	1980 labor share
ν	Substitution elasticity: sectors	0.75	2010 RM employment share
φ	Substitution elasticity: routine inputs	0.85	2010 labor share
	Union		
(c_l, c_h)	Firing costs	$(0.01, 0.05)$	Agg. union effect 1980-1990

1.4.2. Model Mechanism

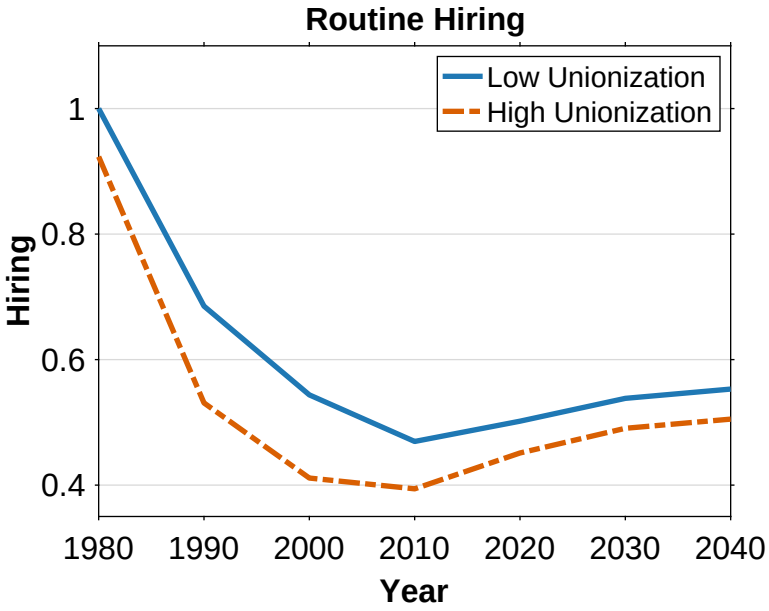
Figure 1.6 shows the fall in capital prices which induces firms to gradually adopt automation over time and thereby triggers the transitional dynamics.

Figure 1.6: Price of automation along the transition matches the decline in measured capital prices in the U.S. between 1980 and 2010 from Hubmer (2023).



In 1980, the economy is in steady state. In 1990, agents in the economy wake up and learn about the new path of automation prices. Thus, they learn that the automation price has already fallen in 1990 and will further decline until 2010. The price decline matches the price decline of capital goods in the U.S. since 1980. The fall in automation prices triggers a fall in routine employment in the high as well as low-unionized economy as routine workers are imperfectly substitutable with automation. However, higher firing costs in the high-unionized economy increase the cost of layoffs and incentive firms to adjust to a larger extent through the hiring margin. As a result, demand for young workers entering the routine occupation falls relatively more in the high-unionized economy. Figure 1.7 shows hiring of young workers in the routine occupations relative to the 1980 steady state in both economies.

Figure 1.7: Hiring falls more in the high-unionized labor market as firing cost generate insider-outsider dynamics.

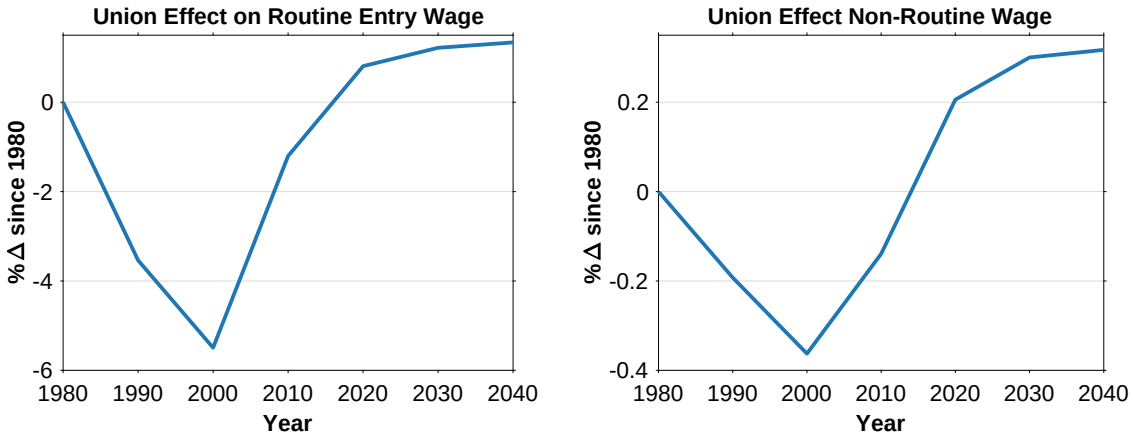


In the initial steady state, hiring of routine workers is 8% lower in the high-unionized than in the low-unionized economy since wage premia imposed by the union increase wages but reduce the level of routine employment across all age groups. After automation becomes available, routine hiring falls in both labor markets, however, it falls substantially more in the high-unionized labor market early in the transition. The greater fall in hiring in the high-unionized labor market in 1990 is driven by larger adjustment through incoming workers in response to current automation adoption as well as by a further preemptive reduction in hiring in anticipation of future adoption to avoid adjustment costs along the transition.

Figure 1.8 shows the union effect on wages and paints a similar picture. Early in the transition, unionization further depresses entry wages in the routine occupation by lowering demand for incoming workers. In particular, routine entry wages decline 3.5% more until 1990 and 5% more until 2000 in the high-unionized relative to the low-unionized labor market. The union effect spills over into the non-routine occupation, resulting an additional 0.3% decline in non-routine wages until 2000 in the high-unionized labor market. The spillover is driven by the accelerated routine

employment decline in the high-unionized labor market, which in turn accelerates the reallocation of workers to the non-routine occupations and reduces wages there.

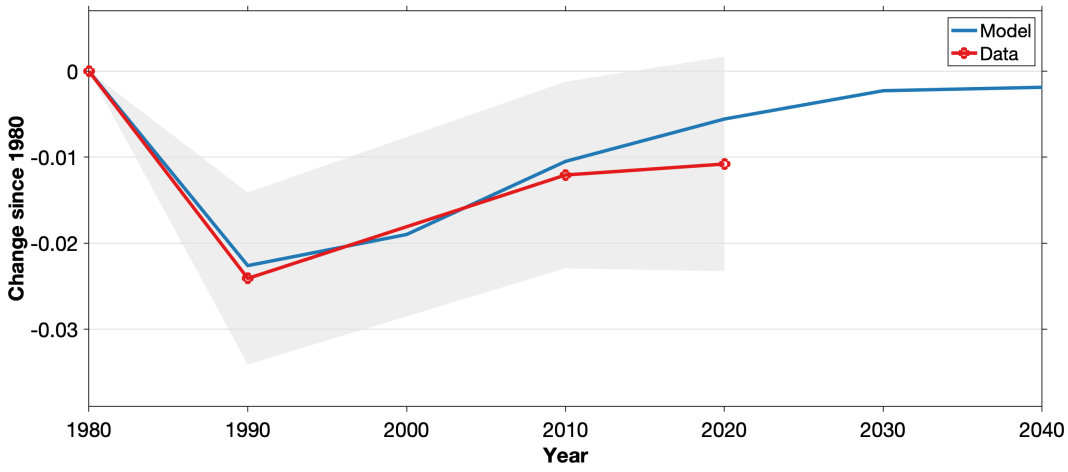
Figure 1.8: The routine entry and non-routine wage temporarily fall in the high-unionized relative to the low-unionized economy during the transition.



1.4.3. Model Validation

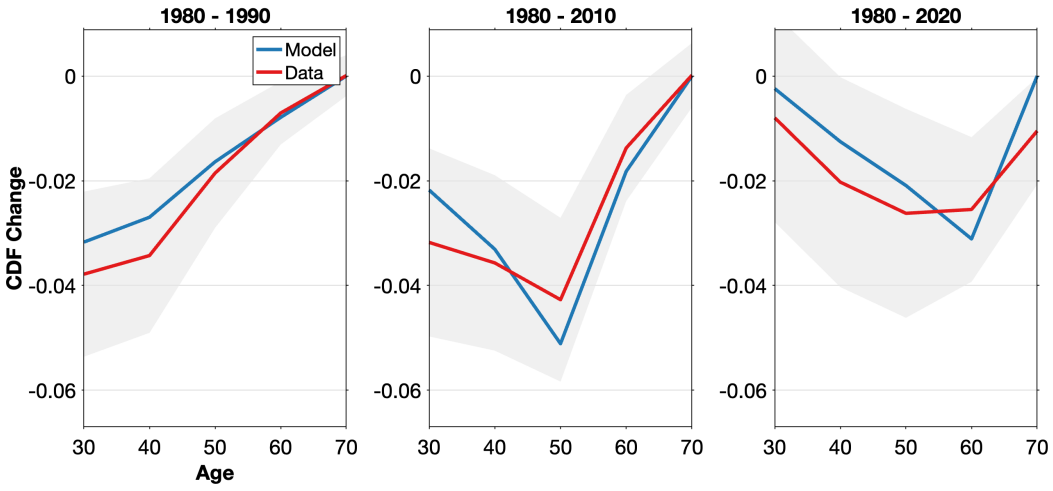
The evolution of aggregate routine employment and the evolution of the age distribution of routine workers along the transition are not targeted by the calibration. To validate the model and connect it with the empirical findings, I test whether it matches the evolution of routine employment and the age composition, and thus, whether it matches the aggregate and distributional union effect documented in the data.

Figure 1.9: Effect of unionization on routine employment over time. The red line shows data estimates for going from the 10th to the 90th percentile of unionization. The blue line shows the model output, comparing the low and high-unionized economy.



How does unionization shape the timing and extent of the overall routine employment decline in the model? Figure 1.9 displays the union effect on the routine employment share as documented in the data as well as in the model. In particular, for the model it shows the difference between the change in the routine employment share in the low and high-unionized economy. The model captures the accelerated routine employment decline in the high relative to the low-unionized economy between 1980 and 1990, driven by a preemptive reduction in routine employment in anticipation of future adoption to avoid firing costs along the transition. Recall, the difference in firing costs between the two economies is calibrated to match the 1990 data point and, thus, the close match between data and model in 1990 is no surprise. Without being targeted, the model matches well that after 1990 the routine employment decline in the low-unionized economy catches up as automation adoption increases and routine workers are being displaced in the low-unionized economy. By 2020, most of the gap has closed as routine employment in the high-unionized economy has declined by roughly 0.5 percentage points more, which is slightly larger than in the data. I now turn to the underlying distributional effect to see if the model can match how unions shape the distribution of transitional costs between young and old.

Figure 1.10: Effect of unionization on the age composition of routine workers over time. The red line shows data estimates for going from the 10th to the 90th percentile of unionization. The blue line shows the model output, comparing the low and high-unionized economy.



The model matches well the untargeted downward shift in the cdf of the age distribution of routine workers as displayed by Figure 1.10. Consistent with the data, unionization results in a downward shift in the cdf that is driven by a fall in the share of young workers. By 1990, the share of incoming workers below the age of 30 declines by 3.2 percentage points more in the high-unionized economy. This initial reduction in new hires then moves up the age ladder over time as the 1980 cohort ages. As a result, unionization induces an aging of the routine workforce, and these differences between the age composition of routine workers in the low and high-unionized economy persist throughout the transition.

1.5. The Welfare Cost of Automation

1.5.1. Measurement of the Welfare Cost of Automation

To measure the individual-specific welfare impact of automation, I calculate the permanent percent decrease in consumption a worker would be willing to accept to return to the 1980 steady state and thereby avoid automation, keeping her individual states fixed. I then compute this consumption equivalent variation at different times during the transition for different cohorts of workers to understand the evolution of the cross-sectional automation impact over time. Let $x_t(s, k, z^R, z^N, a)$

be the required compensation in consumption to be indifferent to automation for a worker with individual state (s, k, z^R, z^N, a) in period t. It is given by

$$x_t(s, k, z^R, z^N, a) = \left(\frac{V_{1980}(s, k, z^R, z^N, a)}{V_t(s, k, z^R, z^N, a)} \right)^{\frac{1}{1-\sigma}}.$$

My primary interest is in understanding to what extent welfare costs differ across routine workers of different cohorts and how unionization impacts these welfare costs. In particular, I focus on the difference between workers who entered the routine occupation before the automation shock hit and are caught by surprise, and workers who entered during the transition and therefore anticipate the current and future impact of automation when making their occupational choice.

1.5.2. The Welfare Cost of Automation for Routine Workers

Before studying how unionization affects the welfare cost of automation, I start by looking at the welfare cost of automation for routine workers in the low-unionized labor market to understand how automation shapes life-cycle wage paths of routine workers from 1980 onward.

Figure 1.11: Welfare cost of automation for routine workers in 1990 in the low-unionized economy.

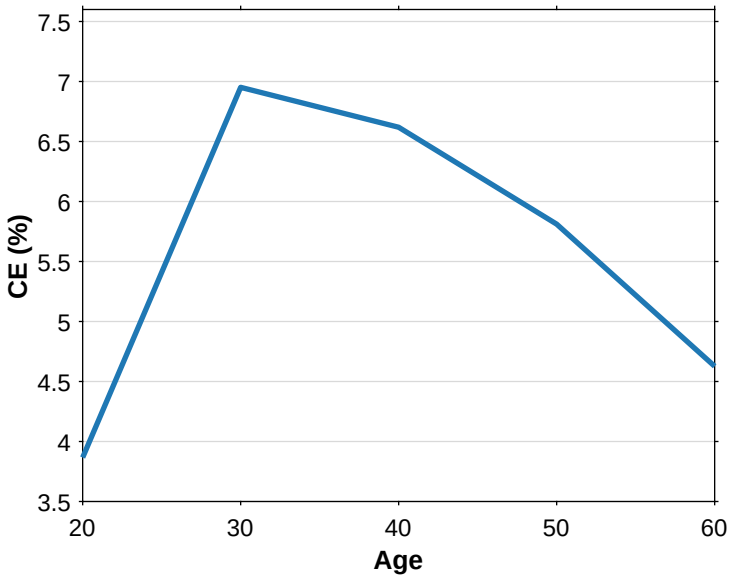


Figure 1.11 displays the welfare impact of automation for routine workers of different ages, averaged across all routine workers in that age group, in the low unionized labor market in 1990. The welfare costs are postitive for all age groups, meaning automation in 1990 is costly for all existing routine workers. For these workers, automation is costly for two reasons: displacement risk in the form of layoff risk, and permanent earnings losses as current and future routine wages fall. The welfare costs are large and range from 4% of permanent consumption for the incoming workers to 7% for workers aged 30.

What drives the inverted u-shape of the welfare costs by age? As the shock hits in 1990, incoming workers fully anticipate the negative impact of current and future automation on their entire life-cycle wage path and expected layoff risk in the routine occupation. In turn, a routine career becomes less desirable and the required routine labor productivity z^R that justifies entering the routine occupation rises. The average labor productivity of the incoming routine cohort in 1990 is therefore higher than for the older, incumbent cohorts. More productive workers have a higher life-cycle wage and consumption path which limits the welfare impact of automation for them. By contrast, incumbent routine workers are less productive on average as they made their occupational choice prior to the automation shock. Their consumption paths on average are lower which increases the welfare impact of automation. The costs are particularly driven by the subset of workers who would not have entered the routine occupation if they had anticipated the automation transition but are now stuck as switching occupations comes at the cost of losing their accumulated occupation-specific human capital. Among incumbent cohorts, the welfare costs are highest for the youngest workers aged 30 as they still have 40 years of routine work ahead of them and will experience the full automation impact over time. The horizon of older, incumbent routine workers is shorter which limits the welfare impact of automation for them to its short and medium-term effects, resulting in falling welfare costs by age.

Figure 1.12: Welfare cost of automation for routine workers along the transition in the low-unionized labor market.

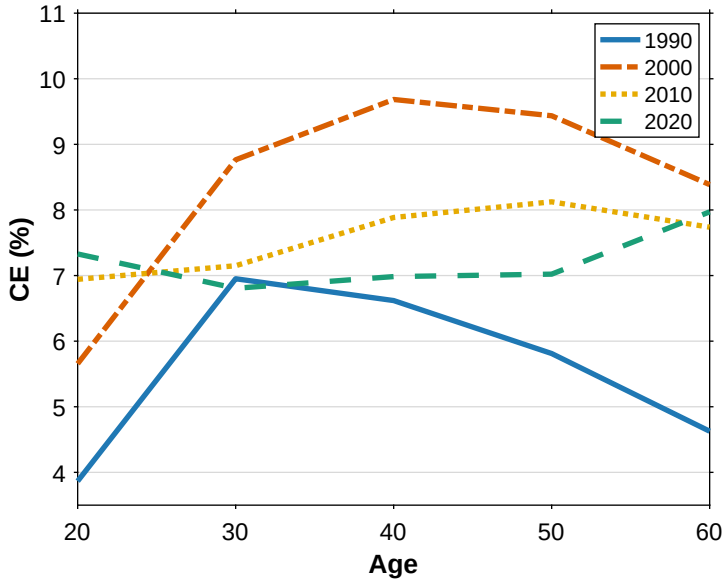


Figure 1.12 displays the welfare cost to routine workers along the transition. As capital prices keep falling, firms increase automation adoption in 2000 which raises the level of welfare cost relative to 1990. Note that the routine workers aged 40 in 2000 are, abstracting from layoffs and occupational switches, the same workers who were 30 years old in 1990. Along the transition, the welfare cost remains highest for the cohort of workers that entered the routine occupations in 1980, right before the automation shock. These workers experience large permanent earnings losses over their entire life-cycle and would be willing to give up almost 10% of permanent consumption in 2000 to avoid automation. Automation prices fall until 2010 and cohorts entering from 2010 onward enter either at the end or after the transition. As a result, the welfare costs in 2010 and 2020 are driven by the long-term impact of automation on routine workers which is similar across cohorts, leading to a flattening of the curve.

1.5.3. The Union Induced Transfer of Automation Costs

How does unionization shape the welfare cost of automation for routine workers along the transition? To answer this question, I compare the welfare cost of automation as computed above in the low

and high unionized labor market.

Figure 1.13: Union effect on the welfare cost of automation in 1990.

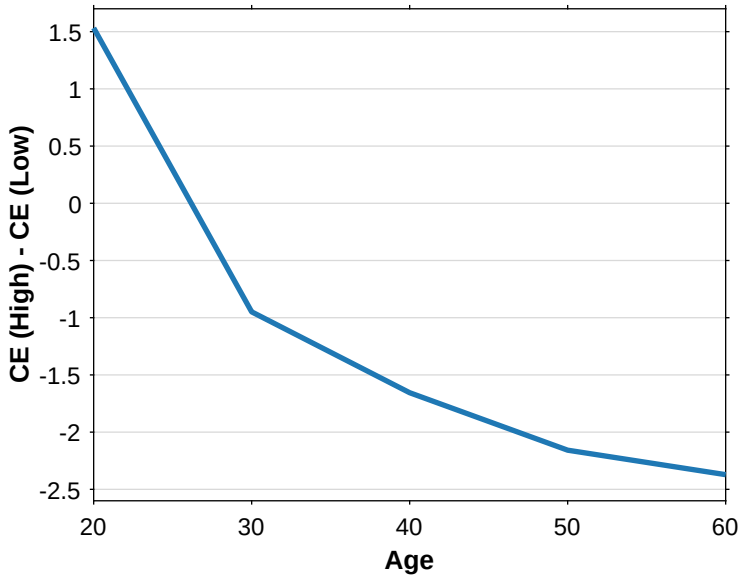


Figure 1.13 shows the difference in the welfare impact of automation between the low and high-unionized economy in 1990. Unionization shifts the welfare costs from incumbent workers to young, incoming workers who enter the routine occupation. Driven by the negative union effect on routine entry wages shown in Figure 1.8, the cost of automation for incoming routine workers is 1.5% of permanent consumption larger in the high-unionized relative to the low-unionized economy. By contrast, the union protects incumbent routine workers by stabilizing their current wages and limiting their layoff risk, reducing the cost of automation to them by up to 2.5% of permanent consumption.

Figure 1.14: Union effect on the welfare cost of automation along the transition.

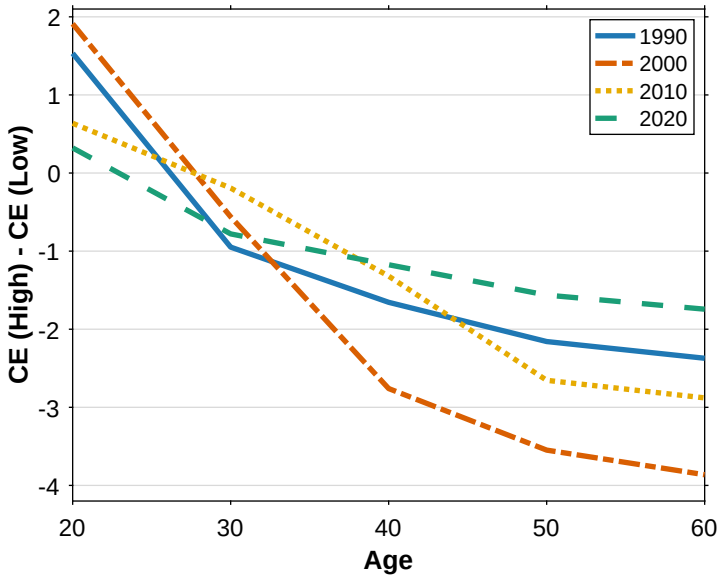


Figure 1.14 shows the union effect along the transition. As automation adoption increases in 2000, the productivity of routine workers falls which puts downward pressure on wages and increases the incentive of firms to lay off workers. As a result, the union protection becomes more valuable for incumbent routine workers. Especially for 60 year old routine workers layoffs would result in large earnings losses, meaning unionization reduces the welfare cost of automation for them by almost 4% of permanent consumption. While the positive union effect on older, incumbent workers increases substantially between 1990 and 2000, the negative impact on incoming workers rises only slightly, reaching close to 2% of permanent consumption. This reflects the fact that incoming workers in 1990 and 2000 endogenously respond to the impact of automation on their expected life-cycle earnings path in the routine occupation by not entering the routine occupation unless their routine labor productivity is sufficiently high. From 2010 onward, capital prices stop falling and the economy transitions to its new long-run steady state. As a result, the union effect begins to flatten and decline in magnitude.

The welfare analysis emphasizes two things: First, among exposed routine workers who are substitutable with technology, automation is particularly costly for existing workers who made their

occupational choice prior to the transition. These workers are caught by surprise and are stuck in a declining occupation, facing increased layoff risk and reduced earnings while switching occupations comes at the cost of losing their occupation-specific human capital. While automation also comes at substantial cost for workers who still enter the exposed occupations during the transition by reducing their expected life-cycle earnings path, these workers are on average more productive as they incorporate the current and future consequences of automation into their occupational choice which limits the welfare impact. Second, unionization shifts the welfare cost from incumbent routine cohorts to incoming workers. Workers who still enter the routine occupation during the transition experience declining entry wages due to unionization, and less productive young workers who would have entered the routine occupation in the past instead enter the non-routine occupation in order to avoid the automation impact.

1.6. Political Implications of the Conflict

An emerging political economy literature connects adverse economic shocks and outcomes to ideological realignment, which induces shifts in political preferences and economic policy. In particular, ideological polarization by race and education have widened among voters, most notably seen in a shift of less-educated Whites to the GOP (Pew Research Center (2014, 2017)). Mian et al. (2014) document a temporary increase in polarization in congressional voting outcomes following financial crises. Several studies document that the widening ideological polarization in Congress correlates with rising U.S. income inequality (McCarty et al. (2016), Voorheis et al. (2015)). Autor et al. (2020) find support for an ideological realignment in trade-exposed local labor markets in the form of rising support for strong-left and strong-right views, as well as pure rightward shifts. However, the causal relation between economic outcomes and shifts in voting behavior remains unclear.

How can the findings of this paper speak to and inform the narrative that puts economic factors at the center of political polarization? The model emphasizes that while unionization protects incumbent routine workers in response to an automation shock, it does so by shifting the cost of automation to young cohorts entering the labor market. In particular, unionization thereby causes a greater decline in routine entry wages for the subset of workers that still enter the routine

occupation, as well as a larger reallocation of young workers to non-routine jobs, such as service sector jobs. As a result, unionization has intensified the deterioration of labor market experiences of less skilled incoming cohorts in routine and non-routine occupations since 1980. Cohorts of workers that have entered the labor market between 1980 and 2000 are in their 50s and 60s today, and precisely the workers whose voting behavior has shifted (Pew Research Center (2014, 2017)). In order to test the importance of economic hardship as a driver of the shift in voting behavior, I test the hypothesis that union-induced employment decline of young routine manual workers between 1980 and 1990 across local labor markets is associated with a larger shift of voting to Republicans in the 2016 and 2020 presidential election.

I use data on county-level returns for presidential elections from 2000 to 2020 from the MIT Election Data and Science Lab (2017) and aggregate that data to the MSA level. The data shows only Democratic and Republican voter shares at the overall MSA level, not for the subset of routine manual workers. The dependent variable is the change in the Republican voter share in a MSA in the 2020 elections relative to the four previous elections in 2004, 2008, 2012 and 2016. I then regress the change in the voter share on the change in the share of routine manual workers that is below the age of 40 between 1980 and 1990, the unionization rate among routine manual workers, the interaction between unionization and the age shift, as well as controls. The interaction term is the coefficient of interest. As before, a greater fall in the share of young workers below the age of 40, that is, an aging of the workforce, indicates less employment prospects for young routine manual workers. The interaction with unionization then measures the additional fall in the routine manual employment share of young workers that is driven by unionization which is the variation of interest.

Table 1.7: The table shows the results from regressing changes in the republican voter share over time on the change in the share of young routine manual workers between 1980 and 1990, routine manual unionization and their interaction at the MSA level.

	Dependent variable: Change in Republican Voter Share			
	2020-2004	2020-2008	2020-2012	2020-2016
	(1)	(2)	(3)	(4)
1980s CDF Change Age 40	0.129	0.370***	0.348***	0.077
	(0.131)	(0.127)	(0.120)	(0.095)
Unionization	0.038	0.034	0.082**	−0.021
	(0.051)	(0.042)	(0.041)	(0.026)
Interaction	−1.241	−1.620***	−1.504***	−0.440
	(0.611)	(0.587)	(0.535)	(0.378)
Mean dependent	-0.025	0.026	0.0071	0.0059
Observations	167	167	167	167
R^2	0.495	0.540	0.491	0.285
Adjusted R^2	0.433	0.484	0.429	0.198
Note:	*p<0.1; **p<0.05; ***p<0.01.			

Table 1.7 shows the regression results. The interaction terms are negative for all four models. Across MSAs, high unionization combined with a greater fall in the share of young workers in routine manual occupations between 1980 and 1990 is associated with an increase in the Republican voter share in the 2020 presidential election relative to previous elections. The effect is particular pronounced relative to the 2004, 2008 and 2012 election. Concretely, comparing a MSA at the 10th to a MSA at the 90th percentile in both independent variables, routine manual unionization and the fall in the share of young routine manual workers during the 1980s, is associated with a 3.9 percentage point and 3.6 percentage point additional increase in the Republican voter share in the 2020 relative the 2008 and 2012 presidential elections, respectively. The effect is weaker and not significant relative to the 2016 election, consistent with the 2016 presidential election already being polarized.

1.7. Conclusion

This paper argues that labor adjustment costs shape how the adverse labor market impact of labor-replacing technological change, such as automation, is distributed across different generations of workers as well as the timing of aggregate labor reallocation. To support the argument, I first document that unions have shifted the incidence of wage and employment declines among routine manual workers from older, incumbent to young, incoming cohorts since 1980. Moreover, unions have accelerated the decline in overall employment within routine manual occupations, resulting in a greater fall in employment in high-unionized labor markets between 1980 and 2000, and a subsequent catch-up of employment decline in less unionized labor markets.

I build a quantitative dynamic equilibrium model of endogenous technological change and unionization which demonstrates that the combination of gradual automation adoption over time and labor adjustment costs imposed by unions can jointly rationalize the two empirical observations. Labor adjustment costs incentivize firms to replace their workforce through reduced hiring rather than through costly layoffs in response to automation adoption. Moreover, when firms anticipate further automation in the future, they shrink their current workforce preemptively in order to avoid adjustment costs along the transition. This anticipatory adjustment channel is strong in the model and gives rise to an accelerated overall employment decline in routine occupations in high-unionized labor markets.

I use the model to quantify the effect of automation and unionization on the life-cycle consumption paths of routine workers across cohorts. The automation impact is most pronounced for incumbent routine workers who made their occupational choice prior to the transition, the welfare cost of automation to these workers reaches 10% of permanent consumption in 2000 in a low-unionized labor market. Workers entering the labor market during the transition endogenously adjust their occupational choice, and particularly less productive workers select into non-routine occupations in order to avoid the automation impact. Nevertheless, entering routine workers would still pay up to 7% of permanent consumption to avoid automation. In a high-unionized labor market, unions protect incumbent routine workers by lowering their layoff risk and wage decline which reduces

the welfare cost of automation to these workers along the transition by up to 4% of permanent consumption relative to low-unionized labor markets. However, the cost is shifted to incoming cohorts. Routine entry wages and hiring falls relatively more in high-unionized labor markets, increasing the welfare cost of automation to incoming routine workers by up to 2% of permanent consumption along the transition. The difference in the welfare benefit to incumbent workers and welfare cost to incoming workers reflects the endogenous response of incoming workers to the automation shock. While older, incumbent workers are stuck in a declining occupation, having made their occupational choice not anticipating automation, incoming workers take the automation transition into account when making their occupational choice.

Finally, I provide suggestive evidence that the union induced shift of the adverse automation impact to young, incoming workers in the 1980s and 1990s has, through its persistent effects on the labor market experiences of these workers, implications for their voting behavior today.

CHAPTER 2

THE MEDICAL EXPANSION, LIFE-EXPECTANCY AND ENDOGENOUS DIRECTED TECHNICAL CHANGE

2.1. Introduction

In 1820 the remaining life expectancy of a twenty year old person living in the United States was approximately 40 years, per capita income was $\$2,674$ (in constant 2011 US dollars, according to the Maddison data base) and virtually none of that income was spent on health goods and services, abstracting from expenditures on basic food and hygiene[19]. In 2019, the year prior to the COVID-19 pandemic, remaining life expectancy at age twenty stood above 60 years, income per capita rose to $\$55,355$ (again in constant 2011 US dollars), and close to 20% of that income was spent on goods produced by a modern, high-tech health sector. In this paper, we build a quantitative theory to explain these observations, and use the theory to investigate what drove this transition from a life that was poor and short and without much medical care to one with high incomes, long lives and high-tech, expensive health care.

To do so we develop a two-sector overlapping generations model with endogenous and directed technical change in which income growth, life expectancy, technological progress in the health and the final goods sector, as well as the size of the health sector and the quality and price of the goods it produces are jointly determined in general equilibrium. The model interprets the facts as three phases of a dynamic equilibrium in which households are initially poor and the price of health goods is prohibitively high so that demand for health goods is zero, life is short and life expectancy stagnant. As income grows, fueled by technological progress, households start consuming basic health goods, life expectancy starts to rise, and directed technological progress eventually, with a delay of ca. 100 years, leads to the emergence and expansion of a modern health sector that commands a significant and growing share of labor demand, production and household expenditures.

[19] Unless otherwise noted, we use *cohort life expectancy* as the relevant measure of life expectancy.

We calibrate the model parameters (which includes the initial conditions for the state of technology in both sectors) to observations on income per capita and life expectancy in 1820 as well as to the timing of increase in life expectancy and the emergence of the modern health sector, and then use the model as a quantitative laboratory to answer two applied questions. After having verified that the model accounts well for the time series in per capita income, life expectancy and the size of the modern health sector, we ask what share of the increase in life expectancy from 1940 onward can be attributed to the modern health sector. We find that share to be approximately 30%, suggesting that the increased expenditures on basic health goods (e.g., better food and hygiene) played a major role in the expansion of life expectancy even in the 20th century, and even in the presence of a modern health sector that approaches an expenditure share of 20% of total GDP.

Second, we use the model to decompose the increase in the relative price of health goods and services between 1940 and 2020 into two components: a term driven by the income-growth-induced rising household demand for health goods relative to final goods; and a term due to productivity growth in the modern health sector relative to the final goods sector, driven by endogenous technological progress. We show that between 1940 and 1980, both components contribute roughly half of the overall increase in the relative health price, but after 1980, technological progress in the modern health sector accelerates and becomes the dominant force.

Related Literature. Our model has three key building blocks, a two-period overlapping generations structure with production akin to Diamond (1965), endogenous investments into health and longevity by private households, as in Grossman (1972) and the endogenous evolution of technological change in the Schumpeterian growth tradition (see, e.g., Aghion and Howitt (1992) or Aghion and Howitt (1998)) whose speed differs across sectors in the economy, akin to Acemoglu and Guerrieri (2008). It seeks to describe the path of economic and health stagnation, take-off during a transition period and, eventually, balanced growth as one dynamic equilibrium, as in the general literature on unified growth theory (see, e.g., Galor (2011) or Hansen and Prescott (2002)).

In trying to explain long run trends in life expectancy and connect it to technological progress our

paper builds on the work by Cervellati and Sunde (2005) who develop a model of the take-off of life expectancy by modeling the feedback loop between income growth, human capital formation, increases in life-expectancy and the size of the population. In contrast to them, we seek to provide a unified theory not only of the take-off in life expectancy in the 19th century, but also the emergence of the modern health sector. That purpose is shared with Hejkal et al. (2022) but their focus is on explaining cross-country differences (and similarities) in the reduction of mortality as well as the evolution of the world population.

More broadly, in terms of model-building this paper contributes to the literature on health spending, R&D in the health sector and endogenous growth. Borger et al. (2008) develop a model with endogenous technology adaptation in the health sector to predict future health spending shares and conclude that health spending will slow down. Ehrlich and Yin (2013) construct an endogenous growth model where human capital is the engine of growth. Both these elements are encompassed in the work of Kuhn and Prettner (2016) who model a final goods sector with an intermediate R&D sector and a labor-intensive health sector. They argue that an expansion of this sector may reduce growth by shifting resources away from R&D spending. Like us, Frankovic and Kuhn (2023) develop an overlapping generations model with endogenous health and two production sectors to evaluate the quantitative impact of the introduction of health insurance through Medicare and Medicaid on health spending trends, macroeconomic performance, and trends in life expectancy since 1960 when Medicaid was introduced. In their model, the growth rate of the final goods sector is exogenous and endogenous in the health sector (as in the work by Böhm et al. (2021) who model the evolution of individual health through an accumulation of health deficits). In contrast, we model endogenous growth symmetrically in both sectors, but permit it to be unbalanced.[20]

Finally, a vibrant positive literature studies potential reasons behind the increase in the health expenditure share in U.S. postwar data and a normative literature explores what share of economic activity should be dedicated to health goods and services. The normative perspective includes the work by Hall and Jones (2007), Jones (2004) and Jones (2016). Hall and Jones (2007) model

[20] An important mechanism in these models is typically a market size effect which triggers innovation spending. Empirical support for this mechanism in the health sector is provided by Acemoglu and Linn (2004).

health spending as a superior good with an income elasticity larger than one. As a consequence, income growth leads to an expansion of the health expenditure share, as in our paper. We extend this framework to a two-sector model with a symmetric treatment of endogenous growth in both sectors, and where unbalanced growth emerges as an equilibrium outcome. In contrast to their paper our main purpose is positive, seeking to understand the expansion of the modern medical sector. Jones (2004) develops an endogenous growth model with R&D to explain the increasing health expenditure share. Our model shares the same broad narrative, but provides an explicit treatment of production in a two-sector model so that income growth spurs quality improvements and thus technological progress in both sectors. Jones (2016) also considers growth in two sectors by studying the optimal rate of consumption growth relative to growth of life-saving technologies.

On the positive side, Anderson et al. (2003) argue that the increase of the health spending share is primarily due to the relative price increase of medical goods,[21] which could either be due to an increase in market power of the supply side relative to the demand side in the market for health goods, as Anderson et al. (2003) suggest,[22] or perhaps due to imperfectly measured quality improvements as our model implies.[23] The quality improvements in turn are the result of costly technological progress. The Congressional Budget Office (2008) presents a review of the literature arguing that technological progress contributes to 40-60 percent of the growth in real per capita health care spending, with demographic change towards an older population being another important contributor. Zhao (2014) argues that an introduction and expansion of social security added to increase of health expenditures. See also Fonseca et al. (2021) in this regard.[24]

[21] Fonseca et al. (2023) estimate US health prices to be 33% higher than those of European countries, which explains 60% of differences in health expenditures.

[22] As suggested by Retzlaff-Roberts et al. (2004), the U.S. health care system is plagued by larger inefficiencies than the systems in other industrialized countries. This is a central theme in a growing literature attributing increases in health spending in the US and cross-country differences across the OECD to health prices and inefficiencies in health care markets, see, e.g. Cooper et al. (2019), Horenstein and Santos (2019), and Feldman and Pretnar (2023). Our monopolistically competitive pricing structure in the health sector may reflect such inefficiencies. Higher prices, and the associated higher average asset returns could also be a reflection of compensation for medical innovation risk, see e.g., Koijen et al. (2016).

[23] In this regard, our paper relates to the literature on a lack of quality adjustments of health goods, e.g., Graboyes (1994), Lawver (2011) and Berndt et al., who find that health price indices profoundly underestimate the quality improvements documented in, e.g., Cutler and McClellan (2001).

[24] A parallel literature studies the sources of level differences in health spending shares across countries and attributes the higher share in the U.S. to the fact that the U.S. is the leading country for the invention of new costly health products, see Chandra and Skinner (2012).

The remainder of this paper proceeds as follows. Section 2.2 presents stylized facts on life-expectancy, aggregate health spending and prices of health goods used to motivate, calibrate and evaluate the model. Section 2.3 lays out the model and defines equilibrium. Section 2.4 contains a theoretical characterization of parts of the equilibrium and Section 2.5 presents the calibration of the model. Section 2.6 contains the quantitative evaluation of the model, both along dimensions targeted in the calibration and validates the model along dimensions not targeted in the calibration. Section 2.7 concludes the paper. Details of the construction of the empirical facts, the computational procedure, the calibration of the model as well as on technical-theoretical derivations are contained in Appendices B.1, and B.2, respectively.

2.2. Stylized Facts

In this section we document the main facts that motivate our model, starting with the timing of the increase in life expectancy in Subsection 2.2.1, then briefly reviewing the time path of income per capita, starting from the industrial revolution in Subsection 2.2.2, and finally documenting the emergence and evolution of the Modern Health Sector in Subsection 2.2.3.

2.2.1. Life Expectancy

When documenting life expectancy for the last two centuries we face two crucial choices. First, in the early period a significant increase of life expectancy is due to a decline in child mortality, with later improvements mainly accruing to increased life expectancy conditional on survival to adulthood. Since our model focuses on the the second part, the improved longevity of adults, so will our discussion of the data. Second, life expectancy at a given point in time can be measured using purely cross-sectional survival rates or employ cohort survival rates. The first, cross-sectional concept only requires age-specific survival rates at a given point in time, but assumes that a current 20 year old adult will have the same survival rate at age 50 (that is, 30 years into the future) as a current 50 year old individual, thereby ignoring potential technological improvements in the health sector. Cohort life expectancy is more data-demanding since it requires future age-specific survival rates of the cohort under consideration. Since it fully captures the impact of medical innovations,

a key aspect of our model, we prefer this concept for the purpose of this paper. In conclusion, although we present various time series of life expectancy here, our main focus in the quantitative analysis will be on remaining cohort life expectancy at age 20.

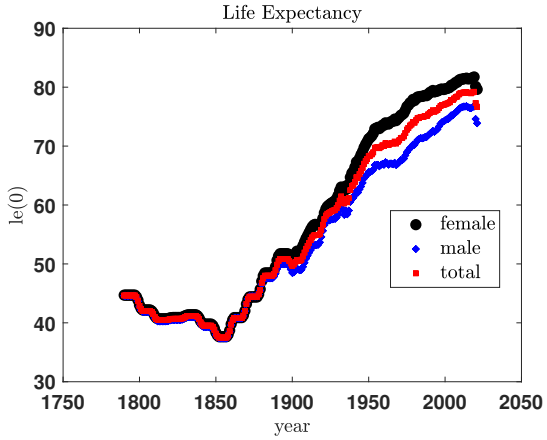
Figure 2.1 shows in panel (a) life expectancy at birth according to the cross-sectional concept.[25] Moving from there to panel (b) of the figure we take three transformations of the data. First, we apply a Hodrick-Prescott filter on the age and time specific mortality rates to extract the age-specific trend components, second, we compute cohort life expectancy instead of cross-sectional life-expectancy and, third, we look at remaining (cohort) life expectancy at age 20 of a person in a given year. From this picture we make the following observations. First, before about 1840, remaining cohort life-expectancy in the US was basically flat and the average—taken for the years 1790 to 1840—was about 39.83 years. Second, since then life expectancy has been increasing so that now—i.e., in year 2020—remaining cohort life expectancy at age 20 stands at 66.22 years. Third, due to death related to pregnancy and child birth, the remaining cohort life expectancy of women was lower than that of men until 1864, when the familiar positive life expectancy gap between female and male life expectancy starts to emerge.[26] In panels (c) and (d) we focus on average life expectancy in the population—averaged across men and women—and we zoom in on the subperiods before and after 1900 and add bootstrapped confidence intervals. Acknowledging the uncertainty about the very precise timing of the take-off in adult life expectancy—which we illustrate here by displaying the bootstrapped 95% confidence intervals of the cohort life expectancy—, based on panel (c) we date this take-off at 1840 for the remainder of this paper. Panel (d) indicates diminishing gains to cohort life expectancy at age 20 in the past decades. In contrast to these diminishing gains, the path of cohort life expectancy at age 60—which we display in panel (e)—indicates increasing gains. This visualizes that in the past century, gains in life expectancy increasingly took place at older ages, which points to the importance of a modern health sector for shaping life expectancy.

[25] A detailed description of data sources and our methods used to calculate life expectancy—according to both concepts—is contained in Appendix B.1.

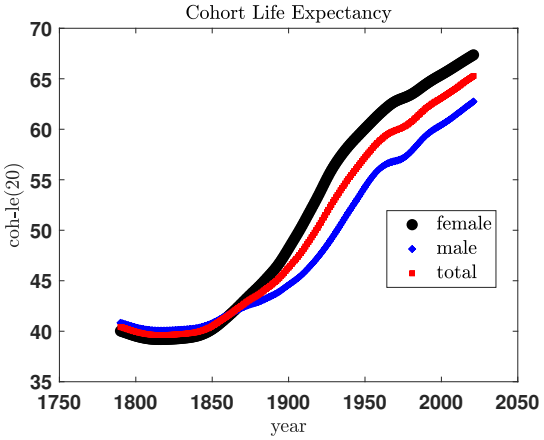
[26] On the contrary, the gap in remaining cohort life expectancy at age 40 between women and men is positive for all years.

Figure 2.1: Life-Expectancy in the US

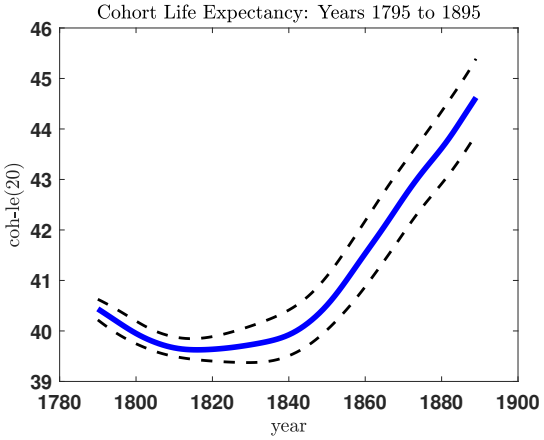
(a) Cross-Sectional Life Expectancy at Birth



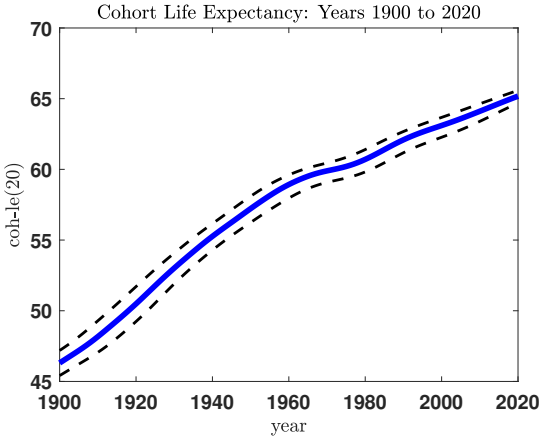
(b) Remaining Cohort Life Expectancy at Age 20



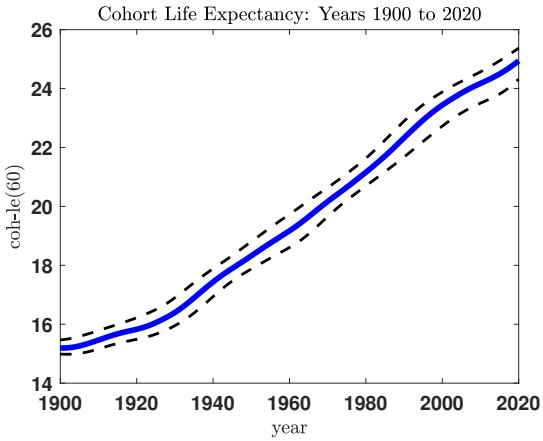
(c) Remaining Cohort Life Expectancy at Age 20: until 1900



(d) Remaining Cohort Life Expectancy at Age 20: after 1900



(e) Remaining Cohort Life Expectancy at Age 60: after 1900

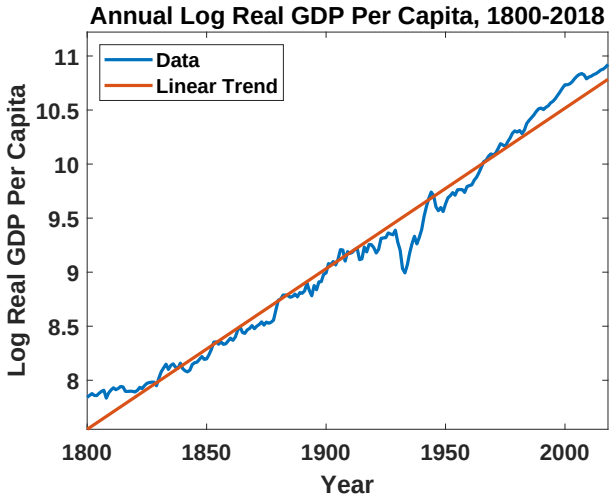


Notes: Cross-sectional life-expectancy in panel (a), cohort life expectancy in panels (b)–(e). 95% Bootstrapped confidence intervals are shown as dashed lines in panels (c)–(e). *Sources*: Hacker (2010), Human Life-Table Database, Human Mortality Database, own computations, see Appendix B.1 for details.

2.2.2. Income per Capita over the Last Two Centuries

Our theory ascribes crucial importance to the increase in income per capita in generating the takeoff, first in life expectancy and then in the emergence of a modern health sector, driven by rising demand for health goods from the household sector.[27] We therefore briefly review the main facts concerning income per capita (growth) in the U.S. over the long run (that is, abstracting from short-run business cycles).

Figure 2.2: Income per Capita in the U.S., 1800-2018



Source: Maddison Project Data Base 2020.

Accordingly, Figure 2.2, which plots the natural logarithm of real income per capita for the U.S. for the last two centuries, as documented in the Maddison Project Data Base, displays the well-documented fact that income per capita started to rise significantly around 1820 and since then has grown at a roughly constant (in fact slightly accelerating) pace of approximately 2% per year.

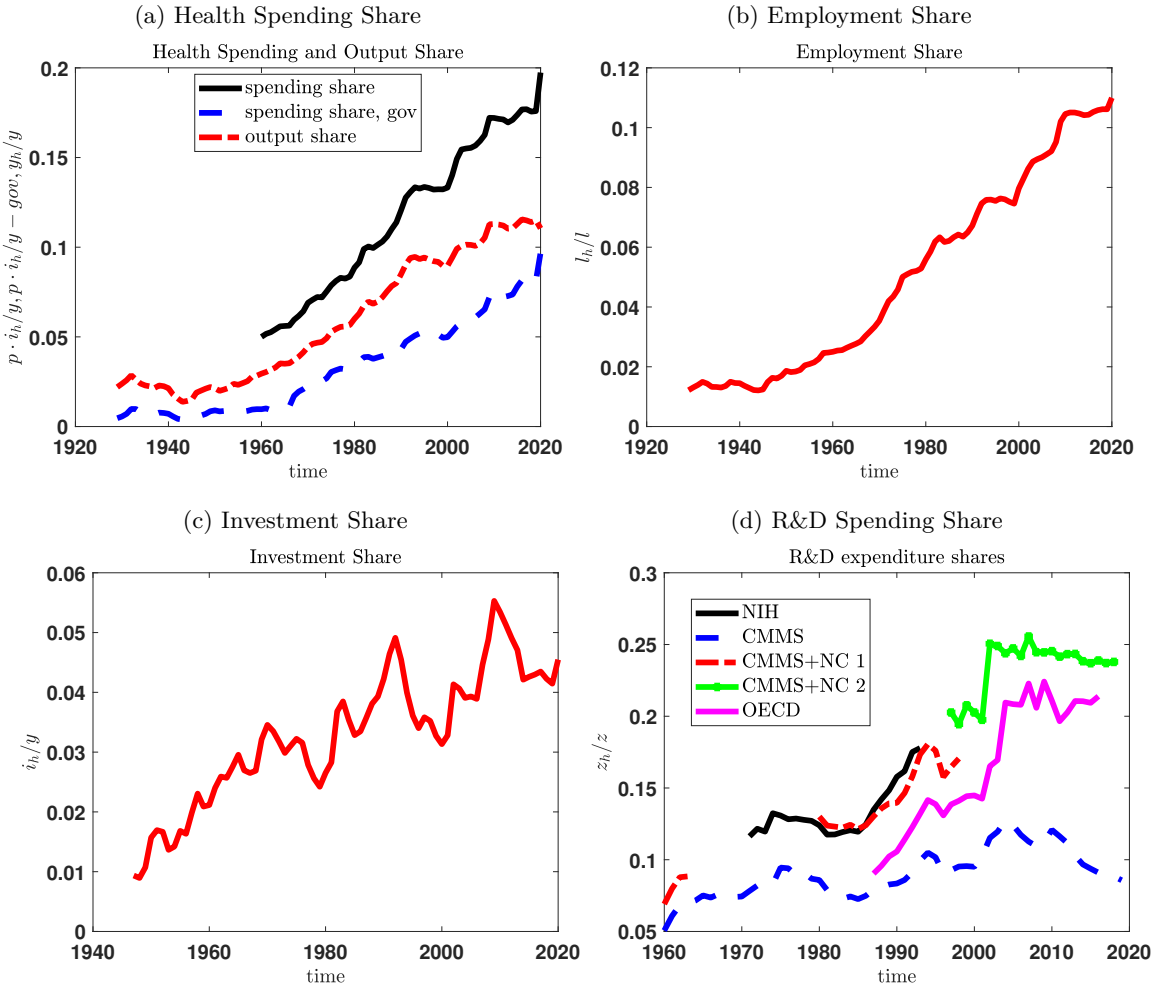
2.2.3. The Emergence and Evolution of the Modern Health Sector

The third set of facts that motivate our analysis is the emergence and growth of a modern health sector in the 20th century. Parker (2019) dates the start of the era of modern medicine at ca. 1920 in his book, pointing to landmark breakthroughs such as the discovery of Penicillin in 1928 and the

[27] It is important to stress, though, that in our model income growth is not exogenously assumed, but emerges endogenously as part of the dynamic equilibrium transition path from an initial period of stagnation towards a balanced growth path with constant, positive growth rates of incomes and production in both sectors and constant expenditure shares.

start of its mass production towards the end of WWII, the discovery and analysis of blood types and blood transfusions and the associated understanding of the causes of diabetes mellitus, as well as the emergence of cancer research and cures.

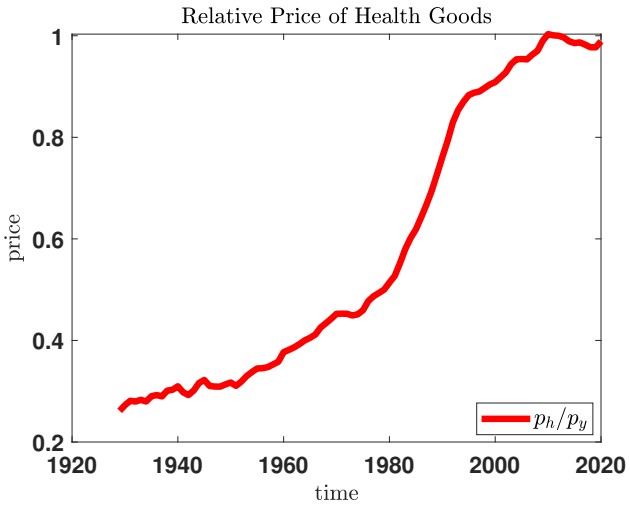
Figure 2.3: Health Shares (Spending, Employment, Investment, R&D) in the U.S



Sources: Panel (a): The government spending share (blue dashed line) is taken from www.usgovernmentspending.com. The total health expenditure share (black solid line) is taken from the National Health Expenditures data of the Centers for Medicare & Medicaid Services (CMS). The output share (red dashed line) are household expenditures on health care taken from the Bureau of Economic Analysis (BEA), Table 1.5.5. All series are nominal and related to nominal GDP taken from the BEA data. Panel (b): The employment share is computed as full-time equivalent employees in the health sector relative to the total number of US full-time equivalent employees using data from the BEA, Tables 6.4 and 6.5. Panel (c): The investment share is computed as real investments in the health sector relative to total real investments in the US, with data taken from BEA, Tables 3.7 and 3.8. Panel (d): The real R&D expenditure share data are likewise defined as the ratio of real R&D expenditures in the health sector relative to total real R&D expenditures and compiled from various sources, the CMS, the National Science Foundation (NFS), and the OECD Stan R&D expenditure data.

In Figure 2.3 we plot shares of various measures of output, investment and employment devoted to

Figure 2.4: The Relative Price of Health Goods in the U.S.



Sources: The relative price of health is computed as the ratio of the price indices of household expenditures on health services to the GDP price index taken from BEA, Table 1.6.4.

the health sector.[28] The range of the time series is dictated by data availability, but all measures point to the same broad observation: the share of economic activity contributed by the health sector was close to zero prior to World War II and since then has steadily been increasing over time, to more than 10% of total output (panel (a) and employment (panel (b)) and close to 20% of overall household spending and R&D investment. It is this emergence and continued expansion of the modern health sector we seek to explain as the endogenous equilibrium outcome driven by income growth on the demand side, and endogenous (temporally) unbalanced technological progress on the supply side of the model.

Finally, in Figure 2.4 we plot the relative price of health goods, measured as the ratio of the price index of household expenditures on health services to the GDP price deflator. Health goods have become more expensive over time; it is an open question what share (potentially larger than 100%) of this increase can be attributed to improvement in the quality of modern health care. Our model will permit the interpretation of this observation since it will deliver a time series for the price of one unit of health care, and a series for the price of one unit of quality-adjusted health goods.

In the remainder of the paper we now construct a two-sector overlapping generations model with

[28] Since we display shares, there is no need to deflate the measures by prices

endogenous and directed technical change in which income growth, life expectancy, technological progress in the health and the final goods sector, as well as the size of the health sector and the quality and price of the goods it produces are jointly determined in general equilibrium.

2.3. The Model

We model a small open economy with overlapping generations, where in every period t a unit measure of identical young individuals is born. The number of old individuals is denoted by n_t^o and is endogenous in the model, as described below. The total population is denoted by $n_t = 1 + n_t^o$. When young, households work, earn income, spend resources on consumption and health, and save for consumption in the second period of their lives. The other actors in the economy are firms in three sectors of the economy, final goods firms in the consumption good sector and the health sector that operate under perfect competition, and R&D sector firms that seek to invent intermediate goods of higher quality, and if successful, become the monopolistically competitive suppliers of intermediate goods of a certain variety that they sell to the final goods producers at a markup sufficient to recover the R&D costs needed to generate the new inventions. We now describe the household sector and then the several production sectors of the economy, before defining a competitive equilibrium.

2.3.1. Households

Households derive utility from consumption in young age c_t^y and old age c_{t+1}^o. They survive from the first to the second period of their life with probability ψ that depends on their investment i_t into health goods when young. The utility of being dead is set to zero, and therefore expected lifetime utility is given by

$$(1-\beta)u(c_t^y) + \beta\psi\left(i_t\right)u(c_{t+1}^o) \tag{2.1}$$

where the period utility function $u(c)$ is at least twice continuously differentiable with $u'(c) > 0$ and $u''(c) < 0$, and satisfies the lower Inada condition, thus $\lim_{c\to 0} u'(c) = \infty$.[29] The sur-

[29] Since households survive the first period of life for sure, and since we assume that they only value consumption in the second period, the level of utility from being alive in the first period is immaterial.

vival function $\psi(i)$ is increasing in health investment i, that is, $\psi'(i) > 0$. We further assume that $\lim_{i\to\infty} \psi(i) = 1$ and $0 < \psi(0) < 1$ and $\psi'(0) < \infty$ that is, households survive with positive probability into the second period even absent any health investment, and the marginal benefit of health investment is finite at $i = 0$.

Health investment i_t is the composite of two health goods, health good purchases i_{ft} from the final goods sector (standing in for expenditures on basic health goods such as hygiene and nutritious food) and purchases i_{ht} of goods from a separate health production sector (such as modern hospital services and treatments as well as drugs). Thus,

$$i_t = f(i_{ft}, i_{ht}) \tag{2.2}$$

Young households supply labor to both sectors of the economy, with l_{ft} denoting labor supply to the final goods sector and l_{ht} denoting the corresponding supply to the health sector. To model, in a reduced form, the frictions associated with butchers (workers in the consumption sector) becoming surgeons (workers in the health sector) we assume the following effective constraint on the labor supplied by the unit mass of households of the form:

$$1 = g(l_{ft}, l_{ht}). \tag{2.3}$$

A constraint of the form of (2.3) is sometimes used in multi-sector models to model, in a simple manner, imperfect labor mobility across sectors.[30] It also implies that wages can potentially differ across the two sectors.

We choose the final consumption good as the numeraire and normalize p_{ft} to 1. We then denote by $p_t = p_{ht}$ the relative price of goods produced by the modern health sector, and by $w_{jt}, j \in \{f, h\}$, wages in the two sectors of production. We envision the representative young household being composed of a large number of members of size 1, so that total labor income of the household is given by $w_{ft}l_{ft} + w_{ht}l_{ht}$. Furthermore, households receive transfers T_t implied by accidental

[30] E.g., Giagheddu and Papetti (2020) calibrate function $g(.)$ referencing evidence by Cardi and Restout (2015).

bequests from the share of the older generation $1 - \psi(i_{t-1})$ that do not survive until old age. Young households take these transfers as exogenous. The maximization of the utility function (2.1) is then subject to the constraints:

$$c_t^y + i_{ft} + p_t i_{ht} + s_t = w_{ft} l_{ft} + w_{ht} l_{ht} + T_t \tag{2.4a}$$

$$i_t = f(i_{ft}, i_{ht}) \tag{2.4b}$$

$$1 = g(l_{ft}, l_{ht}) \tag{2.4c}$$

$$c_{t+1}^o = R s_t. \tag{2.4d}$$

We assume that the depreciation rate on capital is 1, so that the gross return on saving s_t is given by the world interest factor $R = 1+r$ which we assume to be exogenous and constant. Since optimal saving is always strictly positive, potential borrowing constraints never bind and the period budget constraints can be consolidated to the lifetime budget constraint

$$c_t^y + i_{ft} + p_t i_{ht} + \frac{c_{t+1}^o}{R} = w_{ft} l_{ft} + w_{ht} l_{ht} + T_t \equiv x_t. \tag{2.5}$$

where x_t is cash-on-hand of the household. In equilibrium, transfers to the generation born in period t due to accidental bequests from generation $t-1$ are given by:

$$T_t = R s_{t-1}(1 - \psi(i_{t-1})). \tag{2.6}$$

Thus transfers are positive if and only if $\psi(i) < 1$ and households die with positive probability between young and old ages.

2.3.2. Firms, Production and R&D

2.3.2.1 Final Goods Producers

Let $j \in \{f, h\}$ stand for the final and the health sector of the economy, respectively, and p_{jt} for the price of the output of each of the two sectors. In each sector a representative firm uses a continuum of intermediate inputs indexed by i and labor to produce sectoral output y_{jt} according

to the production function

$$y_{jt} = \left(\int_0^1 q_{jit}^{1-\alpha_j} y_{jit}^{\alpha_j} di \right) l_{jt}^{1-\alpha_j}, \tag{2.7}$$

where $0 < \alpha_j < 1$ and y_{jit} is the quantity of intermediate input i used to produce the output good in sector j at date t and l_{jt} is the number of workers employed in sector j. The entity q_{jit} denotes the quality of intermediate input i at date t in sector j. Growth in this model results from innovations that increase the quality q_{jit} of intermediate inputs. Since the final good producer is competitive and takes factor input prices as given, she hires labor and intermediate inputs to equate marginal productivities to these input prices, taking as given their qualities q_{jit}. Let the wage rate in sector j be given by w_{jt} and the price of one unit of intermediate good i in sector j is p_{jit}. The first order conditions are

$$p_{jt} (1 - \alpha_j) \left(\int_0^1 q_{jit}^{1-\alpha_j} y_{jit}^{\alpha_j} di \right) l_{jt}^{-\alpha_j} = w_{jt} \tag{2.8}$$

for labor demand and

$$p_{jt} \alpha_j q_{jit}^{1-\alpha_j} y_{jit}^{\alpha_j - 1} l_{jt}^{1-\alpha_j} = p_{jit} \tag{2.9}$$

for the demand for intermediate goods, given their quality q_{jit}.

2.3.2.2 Intermediate Goods Producers

Each intermediate good producer i is a monopolist that takes the demand function (2.9) as given and uses capital (which depreciates immediately after use) to produce the intermediate good according to:

$$y_{jit} = k_{jit}. \tag{2.10}$$

The gross rental rate of capital is given by R, so that each intermediate goods monopolist producer maximizes profits, taking as given the demand function of the final goods producer,

$$\pi_{jit} = \max_{y_{jit}} \left\{ \left[p_{jt} \alpha_j q_{jit}^{1-\alpha_j} y_{jit}^{\alpha_j - 1} l_{jt}^{1-\alpha_j} \right] y_{jit} - R y_{jit} \right\},$$

with first order condition

$$y_{jit} = \left(\frac{p_{jt} \alpha_j^2}{R} \right)^{\frac{1}{1-\alpha_j}} q_{jit} l_{jt} \tag{2.11}$$

and profits

$$\pi_{jit} = \frac{1-\alpha_j}{\alpha_j} R y_{jit} > 0. \tag{2.12}$$

The monopolistic price follows from using (2.11) in (2.9) as

$$p_{jit} = \frac{1}{\alpha_j} R > R, \tag{2.13}$$

hence featuring the standard markup over marginal costs, R. It is the same across all intermediate input producers i. Finally, observe from (2.11) that $\frac{y_{jit}}{q_{jit}}$ is constant across varieties i. Likewise the ratio of profits to quality $\frac{\pi_{jit}}{q_{jit}}$ is constant across varieties i, which we state for further reference using (2.11) in (2.12) as

$$\frac{\pi_{jit}}{q_{jit}} = \frac{1-\alpha_j}{\alpha_j} \left(\frac{p_{jt}\alpha_j^2}{R^{\alpha_j}} \right)^{\frac{1}{1-\alpha_j}} l_{jt}. \tag{2.14}$$

2.3.2.3 Aggregation of Production Sector

Because the ratios of variety-specific intermediate outputs to quality y_{jit}/q_{jit} and profits to output (or quality) π_{jit}/y_{jit} (π_{jit}/q_{jit}) are constant across varieties i we get immediate aggregation results for each sector. For each production sector j we can determine aggregate capital input as

$$k_{jt} = \int_0^1 k_{jit} di = \int_0^1 y_{jit} di = \left(\frac{p_{jt}\alpha_j^2}{R} \right)^{\frac{1}{1-\alpha_j}} q_{jt} l_{jt} \tag{2.15}$$

where

$$q_{jt} = \int_0^1 q_{jit} di \tag{2.16}$$

is an aggregate quality index of intermediate inputs in sector j. Furthermore, exploiting (2.11) and (2.15) in (2.7) yields as aggregate production function for sector j

$$y_{jt} = k_{jt}^{\alpha_j} \left(q_{jt} l_{jt} \right)^{1-\alpha_j}. \tag{2.17}$$

Using equations (2.8) and (2.15) delivers as factor prices for labor inputs and capital inputs:

$$w_{jt} = (1 - \alpha_j) \frac{p_{jt} y_{jt}}{l_{jt}} \tag{2.18a}$$

$$R = \alpha_j^2 \frac{p_{jt} y_{jt}}{k_{jt}}. \tag{2.18b}$$

Finally we can use (2.12) and (2.15) to determine aggregate profits in each sector j as

$$\pi_{jt} = \alpha_j (1 - \alpha_j) p_{jt} y_{jt} \tag{2.19}$$

and thus in each sector j output exhausts factor input payments plus profits:

$$p_{jt} y_{jt} = \pi_{jt} + R k_{jt} + w_{jt} l_{jt} \tag{2.20}$$

To summarize the aggregation result, in each of the two sectors output is produced with a Cobb-Douglas production function with capital and labor inputs in which the level of technology is given by q_{jt}. However, final goods producers cannot rent capital directly, but have to go through monopolistically competitive intermediaries. As a consequence owners of the capital (which will be the old households in equilibrium) command only a fraction α^2 of the value of output, with a fraction $\alpha(1 - \alpha)$ accruing to the monopolist intermediaries.

2.3.2.4 Research and Development

An R&D developer that specializes in intermediate good i spends resources of the final consumption good z_{jit} on R&D to achieve innovation. If successful in innovation, the quality of the intermediate good increases from q_{jit-1} to

$$q_{jit} = \lambda_j q_{jit-1} \tag{2.21}$$

where $\lambda_j > 1$ is a parameter. The successful innovator immediately becomes the monopolist, and for one period enjoys monopoly profits π_{jit} associated with technology level $q_{jit} = \lambda_j q_{jit-1}$. In a product line i in which innovation is not successful a randomly chosen entrepreneur becomes the monopolist and produces at quality $q_{jit} = q_{jit-1}$ with associated profits.

We assume that the probability of innovating is related to the quality reached when successfully innovating given by λq_{jit-1} as well as the size of the corresponding final production sector given by the employment share l_{jt} so that

$$\phi_j(z_{jit}; l_{jt}, q_{jit-1}) = \min\left[\varphi_j \left(\frac{z_{jit}}{\lambda_j q_{jit-1}}\right)^{\gamma_j} \cdot l_{jt}^{-1}, 1\right], \tag{2.22}$$

with $\gamma_j \in (0,1)$ and $\varphi_j > 0$. First, an increase in the scale of the final production sector measured as the employment share l_{jt} dilutes the effects of research outlays, z_{jit}. Conditional on having a new product, successful innovation requires supplying the intermediate good to the respective final production sector. The negative dependence on l_{jt} captures that a larger final production sector benefits the incumbent monopolist (e.g. due to existing supply chain networks, contracts and relationships with hospitals/doctors, etc.), thus lowering the probability of successful innovation. Second, the inverse relationship between the success probability and current quality q_{jit-1} reflects the fact that it becomes increasingly harder to innovate if already a level of quality is reached for variety i. Note that the probability of innovating is bounded between 0 and 1. As a result, there is an upper bound on R&D spending, z_{jit}, which achieves an innovation probability of 1 and beyond which additional spending is unproductive. The upper bound is more likely to become binding in the early stages of a final production sector in which l_{jt} is small.

The R&D entrepreneur then spends resources z_{jit} and, if successful, collects profits π_{jit}. Hence the problem is

$$\max_{z_{jit}} \left\{\pi_{jit}\phi_j\left(z_{jit}; l_{jt}, q_{jit-1}\right) - z_{jit}\right\} \tag{2.23}$$

For interior solutions the first order condition is

$$\frac{\pi_{jit}}{\lambda_j q_{jit-1} l_{jt}} \varphi_j \gamma_j \left(\frac{z_{jit}}{\lambda_j q_{jit-1}}\right)^{\gamma_j - 1} = 1, \tag{2.24}$$

which yields as solution a ratio of R&D spending to potential period t technology $\frac{z_{jit}}{\lambda_j q_{jit-1}}$

$$\frac{z_{jit}}{\lambda_j q_{jit-1}} = \left[\varphi_j \gamma_j \frac{\pi_{jit}}{\lambda_j q_{jit-1} l_{jt}}\right]^{\frac{1}{1-\gamma_j}}. \tag{2.25}$$

In the interior solution the innovation probability is then

$$\varphi_j \left(\frac{z_{jit}}{\lambda_j q_{jit-1}}\right)^{\gamma_j} l_{jt}^{-1} = \varphi_j \left[\varphi_j \gamma_j \frac{\pi_{jit}}{\lambda_j q_{jit-1} l_{jt}}\right]^{\frac{\gamma_j}{1-\gamma_j}} l_{jt}^{-1}. \tag{2.26}$$

The condition for the upper bound on the innovation probability to bind is then

$$1 \geq \varphi_j \left[\varphi_j \gamma_j \frac{\pi_{jit}}{\lambda_j q_{jit-1} l_{jt}}\right]^{\frac{\gamma_j}{1-\gamma_j}} l_{jt}^{-1} \tag{2.27}$$

Noticing that in case of success $q_{jit} = \lambda_j q_{jit-1}$, we can now use the profit equation (2.14) in the above to find effective R&D spending in the binding and non-binding case

$$\frac{z_{jit}}{\lambda_j q_{jit-1}} = \begin{cases} \left[\frac{l_{jt}}{\varphi_j}\right]^{\frac{1}{\gamma_j}} & \text{if } \phi_j\left(z_{jit}; l_{jt}, q_{jit-1}\right) = 1 \\ \left[\frac{1-\alpha_j}{\alpha_j} \varphi_j \gamma_j \left(\frac{p_{jt}\alpha_j^2}{R^{\alpha_j}}\right)^{\frac{1}{1-\alpha_j}}\right]^{\frac{1}{1-\gamma_j}} & \text{if } \phi_j\left(z_{jit}; l_{jt}, q_{jit-1}\right) \in (0,1). \end{cases} \tag{2.28}$$

Notice that effective R&D spending and the probability of innovation are the same across all varieties i. Using the above back in (2.22) we observe that in the interior solution the share of varieties innovating is (due to the law of large numbers)

$$\mu_{jt} = \int \varphi_j \left(\frac{z_{jit}}{\lambda_j q_{jit-1}}\right)^{\gamma_j} l_{jt}^{-1} di = \varphi_j^{\frac{1}{1-\gamma_j}} \left[\gamma_j \frac{1-\alpha_j}{\alpha_j} \left(\frac{p_{jt}\alpha_j^2}{R^{\alpha_j}}\right)^{\frac{1}{1-\alpha_j}}\right]^{\frac{\gamma_j}{1-\gamma_j}} l_{jt}^{-1} \tag{2.29}$$

and is therefore independent of the distribution of qualities across varieties i. For future reference,

also observe that resources spend by entrepreneur i are

$$z_{jit} = \left[\frac{1-\alpha_j}{\alpha_j}\varphi_j\gamma_j\left(\frac{p_{jt}\alpha_j^2}{R^{\alpha_j}}\right)^{\frac{1}{1-\alpha_j}}\right]^{\frac{1}{1-\gamma_j}}\lambda_j q_{jit-1} \tag{2.30}$$

so that total resources devoted to R&D in sector j are equal to

$$z_{jt} = \int z_{jit}di = \left[\frac{1-\alpha_j}{\alpha_j}\varphi_j\gamma_j\left(\frac{p_{jt}\alpha_j^2}{R^{\alpha_j}}\right)^{\frac{1}{1-\alpha_j}}\right]^{\frac{1}{1-\gamma_j}}\lambda_j q_{jt-1}, \tag{2.31}$$

which are also independent of the distribution of qualities across varieties in sector j.

2.3.3. Definition of Equilibrium

In this section we define a competitive equilibrium for our economy. We immediately proceed to defining the equilibrium for the aggregate economy, thereby already exploiting the aggregation results developed in sections 2.3.2.3 and 2.3.2.4. Recall that we assume a small open economy (SOE) facing the exogenous and constant interest rate factor R.

Definition 1. *Given an initial population, $1, n_t^o$, and initial conditions s_0, q_{f0}, q_{h0} and given exogenous return R, a competitive equilibrium is a sequence of household allocations $c_1^o, \{s_t, i_{ht}, i_{ft}, c_t^y, c_{t+1}^o\}_{t=1}^{\infty}$, a sequence of capital and labor inputs of goods producers $\{k_{jt}, l_{jt}\}_{t=1}^{\infty}$, foreign asset holdings $\{f_t\}_{t=1}^{\infty}$, a sequence of R&D expenditures, profits and consumption of R&D developers $\{z_{jt}, \pi_{jt}, c_{jt}\}_{t=1}^{\infty}$, a sequence of aggregate capital and technology $\{k_t, q_{ft}, q_{ht}\}_{t=1}^{\infty}$, prices $\{p_t, w_{ft}, w_{ht}\}_{t=1}^{\infty}$ and transfers $\{T_t\}_{t=1}^{\infty}$ and a law of motion of the old population n_t^o such that*

1. *Household maximization: for each $t \geq 1$, given prices and transfers $w_{ft}, w_{ht}, p_t, R, T_t$, the allocations $i_{ht}, i_{ft}, s_t, c_t^y, c_{t+1}^o$ maximize (2.1) subject to (2.4).*

2. *Transfers T_t satisfy equation (2.6).*

3. *Factor prices satisfy equations (2.18a) and (2.18b).*

4. *Optimal R&D spending z_{jt} in each sector is given by (2.31) and consumption of R&D en-*

trepreneurs is determined as $c_{jt} = \pi_{jt} - z_{jt}$.

5. *The equilibrium innovation intensity* μ_{jt} *is given by equation (2.29) and technology in each sector evolves according to*

$$q_{jt} = (1 - \mu_{jt})q_{jt-1} + \mu_{jt}\lambda q_{jt-1} \tag{2.32}$$

6. *Markets clear: for all* $t \geq 1$

   (a) *Labor Market*

$$1 = g(l_{ft}, l_{ht}). \tag{2.33}$$

   (b) *Capital Market*

$$\sum_j k_{jt} = k_t. \tag{2.34}$$

   (c) *Final Goods Market*

$$s_t + c_t^y + c_t^o n_t^o + i_{ft} + \sum_j [c_{jt} + z_{jt}] = k_{ft}^{\alpha} \left(q_{ft} l_{ft}\right)^{1-\alpha} + Rf_t. \tag{2.35}$$

   (d) *Health Goods Market*

$$i_{ht} = k_{ht}^{\alpha} \left(q_{ht} l_{ht}\right)^{1-\alpha}. \tag{2.36}$$

   (e) *International Capital Market*

$$f_t = s_{t-1} - k_t.$$

   (f) *The population evolves according to*

$$n_t^o = \psi(i_{t-1}). \tag{2.37}$$

2.4. Theoretical Characterization of Equilibrium

In this section we characterize the optimal solution to the household problem. Given prices and transfers, young households choose labor allocation l_{ft}, l_{ht}, health investment allocation i_{ft}, i_{ht} and thus total health investment i_t and the survival probability $\psi(i_t)$, and consumption in both periods of their life c_t^y, c_{t+1}^o, where they maximize (2.1) subject to (2.4). We think of i_{ft} as goods such as expenditures on hygiene and fresh water that are beneficial for longevity but not measured as part of health expenditures. We would like to generate the following properties of the household problem, and will make appropriate functional form assumptions to ensure that they are true in equilibrium.

Proposition 1. *Suppose that the sequence of prices and cash at hand* $\{p_t, x_t\}$ *satisfy*

$$\frac{x_{t+1}}{p_{t+1}} > \frac{x_t}{p_t}. \tag{2.38}$$

That is, there is real income growth along the transition. Then there exist threshold time periods $0 < T_1 < T_2 < \infty$ *such that*

1. *For all* $t < T_1$ *we have* $i_t = i_{ft} = i_{ht} = 0$ *and* $\psi(i_t) = \psi(0)$. *We call this phase 1.*

2. *For all* $t \in [T_1, T_2)$ *we have* $i_t = i_{ft} > 0$ *and* $i_{ht} = 0$ *as well as* $\psi(i_t) > \psi(0)$. *Life expectancy is increasing due to better basic hygiene and food intake, but the modern health sector remains inoperative. This is phase 2.*

3. *For all* $t \geq T_2$ *we have* $i_{ft} > 0$ *and* $i_{ht} > 0$ *as well as* $\psi(i_t) > \psi(0)$. *Life expectancy is further increasing fueled by increasing expenditures in the modern health sector. This is phase 3.*

4. *For* $t \to \infty$, *the economy converges to a balanced growth path with constant expenditure shares in cash-on-hand* x_t. *This is the balanced growth path.*

2.4.1. The Division of Labor Across the Two Sectors

Note that the labor allocation problem is straight-forward and separated from the health spending consumption problem. Thus, we can solve the household model sequentially, first by solving for the labor allocation that maximizes income and second, taking transfers T_t as given, allocating cash-on-hand x_t optimally to consumption, saving and health investment which boils down to a two-dimensional maximization problem.

We specify the function $g(l_{ft}, l_{ht})$ assuming a constant elasticity of ϵ as

$$1 = g(l_{ft}, l_{ht}) = \left(\sum_j l_{jt}^{1+\frac{1}{\epsilon}} \right)^{\frac{1}{1+\frac{1}{\epsilon}}}$$

Notice that $\epsilon = \infty$ this implies perfect labor mobility across the two sectors, and $\epsilon = 0$ it implies no mobility at all.[31] For interpretational purposes we think of a cohort representative household who optimally decides to allocate its labor across the two sectors. In the case of perfect substitution ($\epsilon = \infty$) we have a corner solution whenever wages are not equalized given by

$$l_{it} = \begin{cases} 1 & \text{if } \; w_{it} > w_{jt} \\ \in [0,1] & \text{if } \; w_{it} = w_{jt} \\ 0 & \text{if } \; w_{it} < w_{jt} \end{cases}$$

Thus, we get an interior solution in which both sectors are operative $l_{jt} \in (0,1), j \in \{f,h\}$ iff $w_{ht} = w_{ft}$. In the interior solution the labor supply to each sector is not determined. In the general case of imperfect substitution ($\epsilon < \infty$), which we take as the benchmark, we directly obtain from (2.3) the transformation function

$$l_{ht} = \left(1 - l_{ft}^{1+\frac{1}{\epsilon}}\right)^{\frac{1}{1+\frac{1}{\epsilon}}} . \tag{2.39}$$

[31] Standard estimates of ϵ range between $\epsilon = 0.6$ and $\epsilon = 1.8$ Giagheddu and Papetti (2020).

The first-order condition w.r.t l_{ft} gives

$$w_{ht} = w_{ft}\left(1 - l_{ft}^{1+\frac{1}{\epsilon}}\right)^{\frac{\frac{1}{\epsilon}}{1+\frac{1}{\epsilon}}} l_{ft}^{-\frac{1}{\epsilon}},$$

which simplifies to

$$\frac{l_{ht}}{l_{ft}} = \left(\frac{w_{ht}}{w_{ft}}\right)^{\epsilon} \tag{2.40}$$

and thus determines the relative labor allocation across the two sectors as a function of relative wages.

2.4.2. Quasi-Linear Health Investment Function and the Division of Health Investment

Health expenditures on final goods i_{ft} and in the health production sector i_{ht} are aggregated into effective health investment i_t according to the quasi-linear specification

$$i_t = i_{ht} + (\nu + i_{ft})^{\zeta}. \tag{2.41}$$

for $\nu > 0, \zeta \in (0,1)$. Health investment then enters the survival function satisfying the CDF of a type 2 Pareto distribution given by

$$\psi(i_t) = 1 - (1 + i_t)^{-\xi}. \tag{2.42}$$

for $\xi > 0$. Note that $\psi(\cdot)$ is strictly increasing in ξ, and is strictly increasing in i_t with $\psi(i_{ft} = i_{ht} = 0) = 1 - [1+\nu]^{-\xi} > 0$ and $\lim_{i_t \to \infty} \psi(i_t) = 1$. In addition to giving positive survival for zero health investments, the non-homotheticity parameter $\nu > 0$ prevents the lower Inada condition to hold for final goods investment i_{ft} so that $\lim_{i_{ft} \to 0} \frac{\partial \psi(i_t)}{\partial i_t} \frac{\partial i_t}{\partial i_{ft}} < \infty$ which is a necessary condition for the existence of phase 1 in which $i_{ft} = i_{ft} = 0$. The concavity parameter $\zeta \in (0,1)$ in turn implies that health investment is linear in i_{ht} and concave in i_{ft}. Thus, we have the standard quasi-linear property that agents initially—i.e., once they possess sufficient resources and health spending becomes positive—only buy the good with decreasing marginal benefit i_{ft} and then switch to the

linear good i_{ht} forever as soon as marginal benefits over marginal costs of the two are equalized. This property gives rise to the possibility of the model to generate phase 2, where $i_{ft} > 0, i_{ht} = 0$, and subsequently in time phase 3, where $i_{ft}, i_{ht} > 0$.

We first solve for the optimal split between final goods and health goods for a given amount of health expenditures e_t. Then we solve for the optimal amount of health expenditures e_t. See appendix B.2.1 for the full derivation. We start by solving

$$\begin{aligned} i_t &= i_t(p_t, e_t) = \max_{i_{ft}, i_{ht}} f(i_{ft}, i_{ht}) \\ s.t.\ p_t i_{ht} + i_{ft} &= e_t \\ i_{ft}, i_{ht} &\geq 0 \\ f(i_{ft}, i_{ht}) &= i_{ht} + (\nu + i_{ft})^{\zeta} \end{aligned}$$

Corner Solution with $i_{ht} = 0$, $i_{ft} = e_t$: Health investment i_t is given by

$$i_t = f(i_{ht}, i_{ft}) = i_{ht} + (\nu + i_{ft})^{\zeta} = (\nu + e_t)^{\zeta}.$$

Interior Solution: In the interior solution the first-order conditions hold with equality, which yields

$$\begin{aligned} i_{ft} &= \tilde{\lambda}_t - \nu \\ i_{ht} &= \frac{e_t - (\tilde{\lambda}_t - \nu)}{p_t} \end{aligned}$$

where $\tilde{\lambda}_t \equiv (\zeta p_t)^{\frac{1}{1-\zeta}}$.

Existence of Phase 2: The corner solution with $i_{ht} = 0$ and $i_{ft} = e_t$ characterized above (phase 2) exists if and only if

$$\begin{aligned} \tilde{\lambda}_t &\geq \nu \\ (\zeta p_t)^{\frac{1}{1-\zeta}} &\geq \nu \end{aligned}$$

Existence of phase 2 requires the non-homotheticity factor ν to be sufficiently small relative to health sector price p_t and ζ.

Characterizing the Phases: For a given level of health expenditures e_t we can fully characterize the phases now. Assuming $\tilde{\lambda}_t > \nu$ for existence of phase 2, the phases are then characterized by

$$\text{Phase} = \begin{cases} 1, & \text{if} \quad e_t = 0 \\ 2, & \text{if} \quad e_t \in (0, \tilde{\lambda}_t - \nu] \\ 3, & \text{if} \quad e_t > \tilde{\lambda}_t - \nu. \end{cases}$$

2.4.3. Level of Health Expenditures

Given the optimal division of health investment we now optimize over the allocation of cash-on-hand x_t into savings s_t and health expenditures e_t. That is, the household now solves

$$\max_{0 \leq c_t^y, e_t \leq x_t} (1-\beta) u(c_{t,y}) + \beta \psi \left(i_t(p_t, e_t) \right) u(R \left[x_t - e_t - c_{t,y} \right].$$

Define the share of young consumption in cash-on-hand and the share of health expenditures in old-age spending, respectively, as

$$\begin{aligned} \vartheta_{t,c} &= \frac{c_t^y}{x_t} \in [0,1], \\ \vartheta_{t,e} &= \frac{e_t}{e_t + s_t} = \frac{e_t}{(1-\vartheta_{t,c}) x_t} = \frac{p_t i_{ht} + i_{ft}}{(1-\vartheta_{t,c}) x_t} \in [0,1]. \end{aligned}$$

Then, the maximization problem can be rewritten in terms of those two spending shares

$$\max_{0 \leq \vartheta_{t,c}, \vartheta_{t,e} \leq 1} (1-\beta) u(\vartheta_{t,c} x_t) + \beta \psi \left(i_t(p_t, (1-\vartheta_{t,c}) \vartheta_{t,e} x_t) \right) u(R x_t \left(1 - \vartheta_{t,c}\right) \left(1 - \vartheta_{t,e}\right)).$$

2.4.3.1 Balanced Growth Consistent Functional Form Assumptions

To further characterize the optimal level of health investment, we need to make a decision on the functional form of the per period utility function $u(\cdot)$ in addition to the assumed functional form

assumption of the survival rate in equation (2.42). We choose a functional form that is consistent with the existence of a balanced growth path by following Hall and Jones (2007) and others. We accordingly assume that the utility function takes the form

$$u(c) = \frac{c^{1-\sigma}}{1-\sigma} + b,$$

where $\sigma \geq 0$ and $b \geq 0$ are parameters. Parameter b measures the value of life.

Corner Solution: The corner solution with $\vartheta_{t,e} = 0$ corresponds to phase 1 without any health expenditure. We solve for the cash-on-hand level at which the first kickoff happens, that is, at which health expenditure become positive. For $\sigma = 2$ and $\xi = 1$, we obtain an analytical solution which is given by

$$\begin{aligned} x_{\text{kickoff1}} &= \frac{1}{bR}\left(\frac{1}{2}\left(1+\sqrt{1+4bRA_1}\right)\right) \\ &\equiv \frac{1}{bR}\Delta(b,R,A_1) \end{aligned} \tag{2.43}$$

where

$$A_1 \equiv \frac{1}{\xi}\frac{\left(1+\nu^{\zeta}\right)\left(\left(1+\nu^{\zeta}\right)^{\xi}-1\right)}{\zeta\nu^{\zeta-1}} \begin{cases} = 0, & \text{if } \nu = 0 \\ > 0, & \text{if } \nu > 0. \end{cases}$$

Notice that $\nu > 0$ ensures $A_1 > 0$ which in turn ensures $\Delta(b,R,A_1) > 1$. Further, $x_{\text{lowerbound}} = \frac{1}{bR}$ is the lower bound on cash such that there is no suicide. Then the interval

$$x \in [x_{\text{lowerbound}}, x_{\text{kickoff1}}] \tag{2.44}$$

is non-empty for $\nu > 0$ and characterizes the cash-on-hand region for phase 1 without suicide.

Interior Solution during the Transition: The interior solution with positive health expenditures, $\vartheta_{t,e} > 0$, corresponds to phases 2 and 3. We cannot solve for the shares analytically in the interior solution, instead we have a system of two equations from the first-order conditions given

by

$$(1-\beta)u_y' = \beta\psi u_o' R(1-\vartheta_{t,e}) - \beta\psi' \frac{\partial i_t}{\partial \vartheta_{t,c}} \frac{u_o}{x}$$
$$\frac{\psi}{\psi'} x_t R(1-\vartheta_{t,c}) = \frac{u_o}{u_o'} \frac{\partial i_t}{\partial \vartheta_{t,e}}$$

Interior Solution on the BGP: In the interior solution of the BGP, where both the final goods and the health goods sector are active, we can find the optimal health expenditure share by taking the limit case: $x \to \infty$. Plugging in the functional forms, the first-order condition for the health expenditure share becomes

$$\frac{1}{\xi}(1+i_t)\left((1+i_t)^{\xi}-1\right) x_t R(1-\vartheta_c^*) = \left(\frac{1}{1-\sigma} + b(c_{t+1}^o)^{\sigma-1}\right) c_{t+1}^o \frac{\partial i_t}{\partial \vartheta_t}. \tag{2.45}$$

For a BGP with the properties $p_t = p^*$ is constant, and, as x_t converges to infinity, $i_t \to i_{ht} \to \frac{e_t}{p_t} = \frac{(1-\vartheta_c^*)\vartheta_e^* x_t}{p^*}$ so that i_t and c_{t+1}^o are both constant shares of cash-on-hand x_t to exist we therefore require $\xi = \sigma - 1$. Under this parametric restriction, solving for the limit case where $\psi(i_t) \to 1$, we can find the health expenditure share on the BGP

$$\vartheta_e^* = \left(1 + \left[\frac{(p^* R)^{1-\sigma}}{b\xi}\right]^{\frac{1}{\sigma}}\right)^{-1}.$$

Plugging ϑ_e^* back into the Euler equation yields the BGP share of young consumption in cash-on-hand

$$\vartheta_c^* = \left[1 + \left(\frac{\beta}{1-\beta}\psi\left[R(1-\vartheta_e^*)\right]^{1-\sigma}\right)^{\frac{1}{\sigma}}\right]^{-1}.$$

2.4.4. Transitional Dynamics

In equilibrium, the health price p_t and relative labor allocation $\frac{l_{ht}}{l_{ft}}$ adjust to clear the health goods market and labor market. We analytically derive the demand for labor by final good producers and the supply of labor by households in terms of the wage ratio. We then combine the two which

yields the labor market clearing condition and characterizes the relationship between the health price and the quality ratio across the two sectors along the transition.

Labor demand: Combining the first-order condition for labor from final good producers with the intermediate good producer solution yields for wages in each sector j

$$w_{jt} = p_{jt}^{\frac{1}{1-\alpha_j}} (1-\alpha_j) \left(\frac{\alpha_j^2}{R_t} \right)^{\frac{\alpha_j}{1-\alpha_j}} q_{jt}$$

This delivers the following relationship for the wage ratio

$$\begin{aligned} \frac{w_{ht}}{w_{ft}} &= p_{ht}^{\frac{1}{1-\alpha_j}} \frac{1-\alpha_h}{1-\alpha_f} \alpha_h^{\frac{2\alpha_h}{1-\alpha_h}} \alpha_f^{\frac{2\alpha_f}{\alpha_f - 1}} \frac{q_{ht}}{q_{ft}} \\ &\equiv p_{ht}^{\frac{1}{1-\alpha_j}} \frac{q_{ht}}{q_{ft}} C(\alpha_f, \alpha_h) \end{aligned}$$

where $C(\alpha_f, \alpha_h) \equiv \frac{1-\alpha_h}{1-\alpha_f} \alpha_h^{\frac{2\alpha_h}{1-\alpha_h}} \alpha_f^{\frac{2\alpha_f}{\alpha_f - 1}}$.

Labor supply: The first-order condition for households' labor supply is given by

$$\frac{w_{ht}}{w_{ft}} = \left(\frac{l_{ht}}{l_{ft}} \right)^{\frac{1}{\epsilon}}$$

Equilibrium: In equilibrium labor markets have to clear. Setting the demand and supply condition for the wage ratio equal to each other yields the following equilibrium condition:

$$\left(\frac{l_{ht}}{l_{ft}} \right)^{\frac{1}{\epsilon}} = p_{ht}^{\frac{1}{1-\alpha_j}} \frac{q_{ht}}{q_{ft}} C(\alpha_f, \alpha_h). \tag{2.46}$$

Note that qualities are determined endogenously in the R&D sector and depend on previous qualities q_{jt-1} and the relative health price p_{ht}. Thus, the only period t endogenous variable on the right hand side is the health price p_{ht}.

Let us provide some intuition for the role of the equilibrium relationship (2.46). It disciplines the relationship between the health price and the allocation of labor across the two sectors along the

transition and thereby pins down the price of health p_t. With perfect labor mobility, $\epsilon = \infty$, the left hand side is constant and equal to 1 as wages have to be equalized across sectors for households to optimally supply labor to both sectors. As a result, the left hand side is not growing along the transition which limits how much the quality ratio $\frac{q_{ht}}{q_{ft}}$ can grow. With imperfect labor mobility, $\epsilon < \infty$, the quality ratio $\frac{q_{ht}}{q_{ft}}$ has to grow faster along the transition in order to generate growth in the wage ratio $\frac{w_{ht}}{w_{ft}}$ which is necessary to incentivize labor reallocation from households.

2.5. Calibration

In this section we exposit the calibration of the model. We first discuss our choice of initial conditions, and then the calibration of the remaining parameters.

We interpret each model period as 40 years. We assume that economic life starts at the age of 20, when adults make the health investment decisions which determine the probability to survive to the next period. Thus, the first period covers the biological age span 20-60. The second period is accordingly 60-100. Life-expectancy at biological age 20 in the model is therefore $20+(1+\psi(i))\cdot 40$ years.

In our main experiment, we treat the years prior to 1940 as years prior to the onset of modern medical times. The data on investments share, employment share and output share in the health sector suggest that the modern medical time period starts in about 1940, cf. Figures 2.1 and 2.3. A number of additional salient facts support this interpretation. First, the widespread use of penicillin to treat infections started in the second world war and can thus be dated to about 1943. Second, while it is hard to obtain historical data on health spending shares, the data reported on https://www.usgovernmentspending.com/healthcare_spending support the interpretation of a start of modern medical times in the 1950s.

Also notice that the US medicare system was only introduced in 1965. Again, recall from Figures 2.1 and 2.3 that employment, investment and output shares in health start increasing prior to that year. This supports our interpretation of the data that growth of the health sector essentially is the consequence of research and development efforts, which in turn are triggered by economic

developments (the treatment of soldiers in the second world war might be an additional trigger; our model has nothing to contribute to this observation), rather than interpreting the introduction of Medicare as the trigger as in Frankovic and Kuhn (2023), who develop their theory on the basis of a conjecture in Weisbrod (1991).

Given the assumed frequency of our model of 40 year periods, the years we look at immediately before the opening up of the modern health sector and thereafter are years 1940, 1980 and 2020. According to this interpretation of the data and the mechanics of our model, prior to 1940 any increase of life-expectancy through the lens of our model is attributed to health spending on the final consumption good, such as hygiene health and nutrition. Regarding the initial stage, we notice form Panel (b) of Figure 2.1 that remaining cohort life-expectancy at age 20 in the US was basically flat until cohorts that are of age 20 in year 1820 and starts to rise between 1820 and 1860. Thus, years 1780 and 1820 are the initial period before the kickoff in health spending where society is poor. Already in 1860 spending on the hygiene health good leads to increasing life-expectancy. Accordingly, years 1860 and 1900 correspond to stage 2. Table 2.1 summarizes the stages and the corresponding years.

Table 2.1: Stages and Calendar Years

1780	1820	1860	1900	1940	1980	2020	...
Stage 1		Stage 2		Stage 3			

Notes: Stage 1 is the initial stage where the economy is poor and $i_{ft} = i_{ht} = 0$. In stage 2 all investment in health takes place through spending on the hygiene health good, $i_{ft} > 0, i_{ht} = 0$. In stage 3 the modern health sector is also operative so that $i_{ft} > 0, i_{ht} > 0$.

We calibrate a subset of parameters exogenously either by reference to other studies or by simply fixing their values (first-stage parameters). Others are calibrated to match selected moments in the data. We then intend to endogenously calibrate the following parameters $b, \zeta, \epsilon, \lambda_f, \lambda_h$.

2.5.1. External Calibration

We set $\sigma = 2$ to generate an intertemporal elasticity of substitution of 0.5. To ensure the existence of a BGP, we impose $\xi = \sigma - 1 = 1$, see 2.4.3.1 for the derivation of the BGP. We set the weight on second period utility relative to first period utility, β, to match an annualized discount factor

of 0.92. We choose an annual real interest rate in the small open economy of 1%. The capital intensity in the final good sector is $\alpha_f = 1/3$, consistent with estimates for the U.S. According to Frankovic et al. (2017), based on Acemoglu and Guerrieri (2008), an appropriate estimate for the capital intensity in the health sector is $\alpha_h = 1/5$.

Table 2.2: External Calibration

Parameter	Description	Value	Target
	Small open economy		
R-1	Rate of return	1.5	1 % annual return
	Households		
$1/\sigma$	IES	0.5	Standard
β	Discount factor	0.078	$\beta_{\text{annualized}} = 0.94$
ξ	Curvature survival function	$\sigma - 1 = 1$	Ensures BGP existence
	Firms		
α_f	Capital intensity final	0.33	Standard
α_h	Capital intensity health	0.2	FKW (2017), AG (2008)

2.5.2. Internal Calibration

We aim at expressing in any period t, x_t in terms of the state variables of the problem in order to derive expressions for the conditions for an inoperative health sector. The state variables in any period t are q_{ft-1}, q_{ht-1}, T_t.

Initial Conditions. We choose the initial conditions and a subset of parameters such that the dynamic equilibrium has the desired properties in the initial phases (phases 1 and 2). We now derive the initial conditions and state the desired properties as well as the calibration strategy to satisfy those properties.

Conditional on the modern health sector being inactive in period 1 (which we verify below), we have $l_{f1} = 1$ which allows us to solve for the production side in period 1 analytically. Given R and some q_{f0}, which remains to be determined, we can use (2.29) and (2.32) to compute the quality in

period 1 as:

$$q_{f1} = (1 + (\lambda_f - 1)\mu_{f0})\, q_{f0} = \Upsilon_t(R) q_{f0-1} \tag{2.47}$$

$$= \left(1 + (\lambda_f - 1)\varphi_f^{\frac{1}{1-\gamma_f}} \left[\frac{1-\alpha_j}{\alpha_j} \gamma_f \left(\frac{\alpha_f^2}{R^{\alpha_f}} \right)^{\frac{1}{1-\alpha}} \right]^{\frac{\gamma_f}{1-\gamma_f}} \right) q_{f0} = \Upsilon_t(R) q_{f0}. \tag{2.48}$$

Further note that for $l_{f1} = 1$ we obtain from (2.15) that the capital stock employed in production is

$$k_{f1} = \left(\frac{\alpha_f^2}{R} \right)^{\frac{1}{1-\alpha_f}} q_{f1}. \tag{2.49}$$

1. **Initial assets**: We assume that the small open economy has a zero net foreign asset position in period 1 so that $k_{f1} = s_0$.

   Now use (2.47) in (2.49) and set $k_{f1} = s_0$ to get

$$s_0(q_{f0}, R) = \left(\frac{\alpha_f^2}{R} \right)^{\frac{1}{1-\alpha_f}} \Upsilon_1(R) q_{f0}$$

$$= \left(\frac{\alpha_f^2}{R} \right)^{\frac{1}{1-\alpha_f}} \left(1 + (\lambda_f - 1)\varphi_f^{\frac{1}{1-\gamma_f}} \left[\frac{1-\alpha_j}{\alpha_j} \gamma_f \left(\frac{\alpha_f^2}{R^{\alpha_f}} \right)^{\frac{1}{1-\alpha}} \right]^{\frac{\gamma_f}{1-\gamma_f}} \right) q_{f0} \tag{2.50}$$

2. **Initial income and first kickoff**: We calibrate the initial quality in the final good sector $q_{f,0}$ and the value of life b jointly targeting two moments. First, an initial income level of the economy such that households in period 1 are indifferent between suicide and survival. Second, the timing of the first kickoff in 1860, meaning the period in which households start consuming basic health goods, which corresponds to period 2 in the model. Note, this implies that there are no investments in basic and modern health goods, $i_0 = i_{f0} = i_{h0}$ in the initial period. As a consequence, $n_0^o = \psi(0)$.

   Given the value of life b, we can solve analytically for the initial quality in the final good

sector that ensures that households in period $t = 1$ are indifferent between survival and suicide. Observe that for $l_{ft} = 1$ wages in the final goods sector and transfers are given by

$$w_{ft} = (1 - \alpha_f) k_{ft}^{\alpha_f} q_{ft}^{1-\alpha_f} \tag{2.51}$$

$$T_t = R s_{t-1} (1 - n_t^o), \tag{2.52}$$

Now, use (2.40) for $l_{ft} = 1$ to obtain

$$w_{ft} = R \frac{1 - \alpha_f}{\alpha_f^2} k_{ft}$$

and thus cash-on-hand $x_t = w_{ft} + T_t$ is given by

$$x_t = R \frac{1 - \alpha_f}{\alpha_f^2} k_{ft} + R s_{t-1} (1 - n_t^o)$$

and thus we can express the period 1 cash-on-hand in terms of initial savings using that $k_{f1} = s_0$ as

$$x_1 = R \left(\frac{1 - \alpha_f}{\alpha_f^2} + (1 - n_1^o) \right) s_0(q_{f0}, R)$$

where $s_0(q_{f0}, R)$ is given by (2.50). Finally, recall that for the lower bound (to ensure that households do not commit suicide) we need

$$\begin{aligned} & x_1 \geq \frac{[b(\sigma - 1)]^{\frac{1}{1-\sigma}}}{R} \\ \Leftrightarrow \qquad & R \left(\frac{1 - \alpha_f}{\alpha_f^2} + (1 - n_1^o) \right) s_0(q_{f0}, R) \geq \frac{[b(\sigma - 1)]^{\frac{1}{1-\sigma}}}{R} \end{aligned}$$

which simplifies to $x_1 \geq \frac{1}{bR}$ for $\sigma = 2$. Using (2.50) then gives

$$R\left(\frac{1-\alpha_f}{\alpha_f^2} + (1-n_1^o)\right) \cdot \left(\frac{\alpha_f^2}{R}\right)^{\frac{1}{1-\alpha_f}} \left(1 + (\lambda_f - 1)\varphi_f^{\frac{1}{1-\gamma_f}} \left[\frac{1-\alpha_j}{\alpha_j}\gamma_f \left(\frac{\alpha_f^2}{R^{\alpha_f}}\right)^{\frac{1}{1-\alpha}}\right]^{\frac{\gamma_f}{1-\gamma_f}}\right) q_{f0} \geq \frac{[b(\sigma-1)]^{\frac{1}{1-\sigma}}}{R}$$

and thus we parametrize (breaking indifference as survival) the initial quality in the final goods sector as

$$q_{f0} = \frac{[b(\sigma-1)]^{\frac{1}{1-\sigma}}}{R} \cdot \left[R\left(\frac{1-\alpha_f}{\alpha_f^2} + (1-n_1^o)\right)\left(\frac{\alpha_f^2}{R}\right)^{\frac{1}{1-\alpha_f}} \left(1 + (\lambda_f - 1)\varphi_f^{\frac{1}{1-\gamma_f}} \left[\frac{1-\alpha_j}{\alpha_j}\gamma_f \left(\frac{\alpha_f^2}{R^{\alpha_f}}\right)^{\frac{1}{1-\alpha}}\right]^{\frac{\gamma_f}{1-\gamma_f}}\right)\right]^{-1} \tag{2.53}$$

3. **Kickoff of modern health sector**: We calibrate the initial quality in the health sector $q_{h,0}$ to match the kickoff timing of the modern health sector in 1940. We assume that during phases 1 and 2 while the modern health sector is inactive, health quality q_{ht} grows exogenously at the same rate as quality in the final goods sector, q_{ft}.

4. **Survival function and initial life expectancy**: We calibrate ν in the survival function to match initial life expectancy during phase 1. Observe that ν^ζ pins down the initial life expectancy (before the kick-off). Recall that

$$i_t = i_{ht} + (\nu + i_{ft})^\zeta$$

and

$$\psi(i_t) = 1 - (1 + i_t)^{-\xi}$$

so that with zero health investments in both goods and with the calibration of $\xi = 1$ we get

$$\psi(i_t = \nu^{\zeta}) = 1 - \left(1 + \nu^{\zeta}\right)^{-1} \tag{2.54}$$

Remaining life-expectancy at economic birth is given by

$$y = (1 + \psi) \cdot 40. \tag{2.55}$$

Using (2.54) in (2.55) we get as value for ν under the maintained parametric restriction $\xi = 1$ and for given ζ that

$$\nu = \left[\left(\frac{40}{80 - y} - 1\right)\right]^{\frac{1}{\zeta}}. \tag{2.56}$$

To calibrate y in the above we take our estimate of cohort life expectancy in year 1790 (the beginning of our sample), of $y = 40.43$ years.

Calibration of Remaining Parameters. We calibrate the curvature of basic health spending in the health investment function, ζ, to minimize the distance between life expectancy in the model and in the data from 1860 onward. We further calibrate the step size for innovations in the final good sector, λ_f, targeting average GDP per capita growth between 1820 and 2020, and the step size for modern health, λ_h, targeting the average growth rate of the output share of the modern health sector between 1940 and 2020. Lastly, we fix the scaling and curvature parameters of the innovation probability, γ and ϕ, to 0.5 in the first stage and calibrate the remaining parameters conditional on that.

2.6. Results

In this section we contrast the positive predictions of the model concerning the time paths of income, life expectancy, the share of economic activity devoted to the health sector and the relative price of health with the empirical facts documented in Section 2.2 of the paper. We will also use the model as a measurement tool to quantify a) the contribution of the modern health sector to the overall increase in life expectancy and b) how much in the increase in the observed relative price

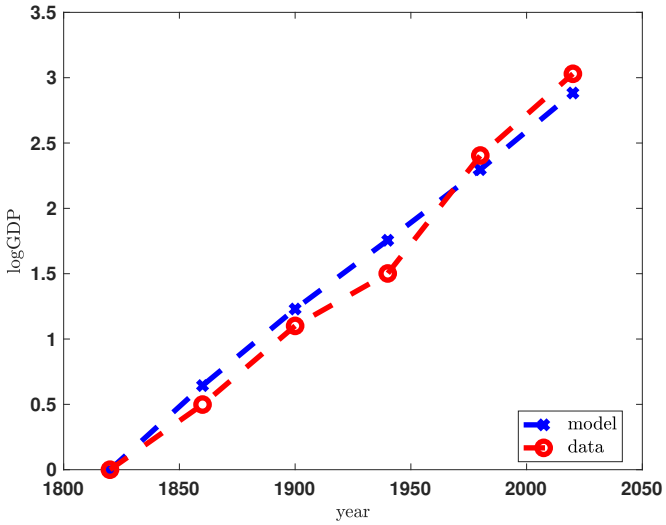
Table 2.3: Internal Calibration

Parameter	Description	Value	Target
	Households		
b	Value of life	129	LE20 in 1860
ζ	Investment curvature	0.6	LE20 after 1860
ϵ	Labor mobility	2	Growth: Q-adj price
	Firms and R&D		
λ_f, λ_h	Growth factor	120, 100	GDP per capita and health GDP
γ_f, γ_h	Innovation probability: curvature	0.5, 0.5	First stage
ϕ_f, ϕ_h	Innovation probability: scaling	0.5, 0.5	First stage
	Initial Conditions		
$q_{f,0}$	Initial quality: final	0.71	Initial income
$q_{h,0}$	Initial quality: health	0.01	Kickoff 1940

of health services in the data should be attributed, from the perspective of the model, to changes in the relative quality q_{ht}/q_{ft}.

In Figure 2.5 we display the time series of income per capita (real GDP per capita) in the model and in the data, both plotted in (natural) log-scale. Per capita income growth in the model is endogenous and driven by innovation and the associated growth in quality (q_{ft}, q_{ht}), initially only in the final goods sector (since the health sector is inoperative), and after 1940 also driven by innovation and thus quality/productivity growth in the modern health sector. We observe that the model matches the roughly linear income growth in the data well.

Figure 2.5: GDP per Capita [Logarithmic Scale]

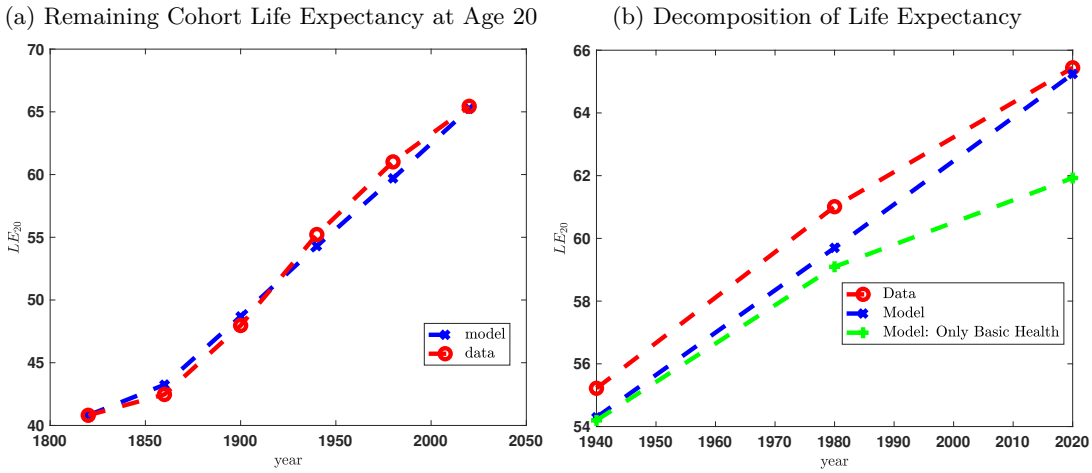


Notes: Natural logarithm of real GDP per capita in model and data. *Source:* Data sources as in Figure 2.2, own calculations.

Figure 2.6 displays the time series of cohort life expectancy at age 20 from the data (as estimated for the U.S. in Section 2.2) and from the model. The left panel shows the entire time series starting from 1820, and the right panel zooms into the period after the modern health sector emerged (from 1940 on) and additionally displays the component of life expectancy that is driven by investments into only basic health goods (green dots). The gap between the blue and the green line can be interpreted as the contribution of modern health.

We observe that the model, on the account of endogenous income growth and unbalanced endogenous growth between the final goods sector and the emergence of the modern health sector in the 1940s, implies that life expectancy grows continually throughout the last 2 centuries and matches the data on cohort life expectancy well.

Figure 2.6: Health Shares (Spending, Employment, R&D) in the U.S.



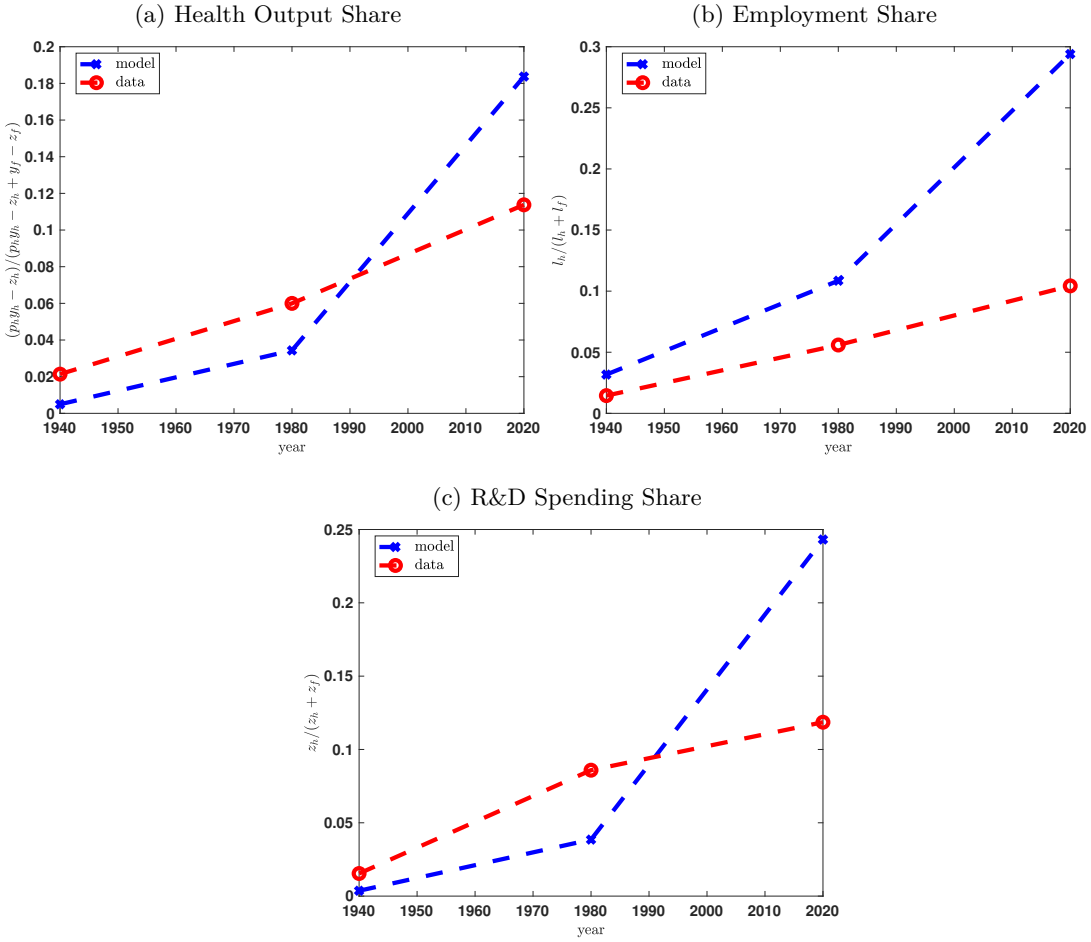
Notes: Remaining cohort life expectancy at age 20 in model (blue x mark) and data (red circles). The figure also shows a decomposition setting spending on modern health goods to zero (black dots). *Sources:* Data sources as in Figure 2.1, own calculations.

Turning to the decomposition of life expectancy improvements driven by the increase in traditional health goods and the modern high-tech health sector, we observe that according to the model, modern health goods become an important driver of life expectancy after 1980 and account for 30% of the increase in life expectancy at age 20 since 1940, translating into 3.3 additional expected years of life. The complementary implication is noteworthy, too: even in the absence of the emergence of modern medicine, income-induced growth in basic health goods (a richer, more balanced diet, better hygiene) would still have led to an increase in life expectancy between 1940 and 2020 by close to 8 years, and basic health goods are the main driver of increased longevity, according to the model, until 1980.

The health-related output-, labor- and R&D shares underlying the emergence of the modern health sector are displayed in Figure 2.7. Note that we plot these shares only from 1940 onward as they are not available empirically, and are equal to zero in the model, beforehand. The figure shows that qualitatively, the model reproduces the increasing shares of labor, R&D and consequently output accruing to the modern health sector. Quantitatively, in the model the takeoff is initially somewhat too slow between 1940 and 1980 then accelerates too much between 1980 and 2020, relative to what

the data suggests. Note that among these shares, only the average growth rate in the output share over time was targeted in the calibration, and therefore some divergence between model and data is to be expected and the ability of the model to qualitatively match the relevant time series from the data should be a considered a qualified success.

Figure 2.7: Health Shares (Spending, Employment, R&D) in the U.S.



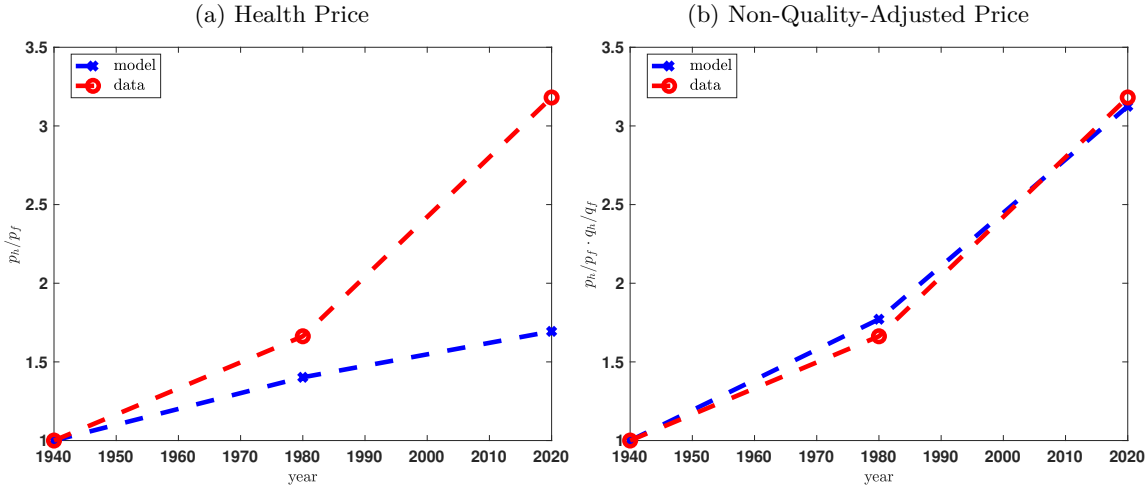
Sources: Data sources as in Figure 2.3, own calculations.

Finally, Figure 2.8 plots the relative price of health goods from Section 2.2 of the data, combined with the relative price times the growth component, $p_t \frac{q_{ht}}{q_{ft}}$, which is the model analogue price that would be observed in the data if there were no quality adjustments at all in the empirically measured relative price of health. We refer to $p_t \frac{q_{ht}}{q_{ft}}$ as the non-quality-adjusted price. According to the model, the price of one raw unit of health goods is increasing slowly over time, accounting for the

improvement in the quality of health care the non-quality-adjusted price of health goods increases at a rate similar to the data.[32] To decompose the observed relative price increase in the measured health price in the data into a component driven by rising output prices versus falling input prices, we decompose the increase in the non-quality-adjusted health price $p_t \frac{q_{ht}}{q_{ft}}$ into its two components: 1) growth in the price of health goods p_t, whose main driver is the (income-growth-induced) rising household demand for health goods relative to final goods; and 2) productivity growth in the modern health sector relative to the final goods sector, driven by endogenous technological progress.

Between 1940 and 1980, both components contribute roughly half of the overall increase in the non-quality-adjusted health price $p_t \frac{q_{ht}}{q_{ft}}$; precisely speaking, the rising demand for health goods accounts for 52%. After 1980, technological progress in the modern health sector accelerates and becomes the dominant force, and as a result, 67% of the overall growth in the non-quality-adjusted health price between 1940 and 2020 is accounted for by greater quality growth in intermediates in the modern health sector (and thus faster productivity growth in that sector) relative to the final goods sector. The relative contribution of rising demand accounts for the remaining 33%.

Figure 2.8: Health Price Index



Sources: Data sources as in Figure 2.4, own calculations.

[32] Thus, this figure implicitly assumes that the empirically observed relative price for health goods has not been appropriately quality adjusted.

2.7. Conclusion

In this paper we build a quantitative theory of income growth, the increase in life expectancy in the last two centuries, and the emergence and expansion of a modern health sector in the 20th century. Our two-sector overlapping generations model with endogenous and directed technical change endogenously determines income growth, life expectancy, and technological progress in the health sector and the final goods sector, as well as the size of the health sector and the quality and price of the goods in general equilibrium. We show that it can generate an economic path in which households are initially poor and the quality-adjusted price of health goods is prohibitively high so that demand for health goods is zero, life is short and life expectancy stagnant. As income grows, fueled by technological progress, households start consuming basic health goods, life expectancy starts to rise, and directed technological progress eventually, with a delay of ca. 100 years, leads to the emergence and expansion of a modern health sector.

Since technological progress in the health sector is endogenous, government health policies (such as the funding of basic research in the health sector or the public provision of health goods though government-run health insurance or the direct production of these goods in public hospitals) will impact the timing and speed of the development of a modern health sector. The next steps of our analysis will be to evaluate positively, and to study normatively the importance of government interventions in the health sector in the 20-th century in the U.S.

CHAPTER 3

DYNAMICS OF THE WEALTH DISTRIBUTION IN THE PRESENCE OF HIGHER-ORDER EARNINGS RISK

3.1. Introduction

The aim of this paper is to explore the long and short-term dynamics of the wealth distribution in the presence of higher-order earnings risk. A growing body of recent empirical studies using administrative and survey panel data on individual earnings finds that earnings dynamics are richer than usually assumed and modeled in quantitative macroeconomic models. Specifically, these studies document that the distribution of shocks to individual earnings exhibits sizable left-skewness and substantial excess kurtosis. This is in contrast to most standard approaches of capturing individual earnings risk. The most commonly used example is the canonical linear transitory plus persistence process with a Gaussian distribution for both idiosyncratic shocks which can neither capture any higher-order moments in the distribution of shocks to individual earnings nor any dependence of persistence or moments on the earnings history.

Earnings dynamics, coupled with the net value of asset holdings, play a central role in determining consumption responses to earnings risk and shocks over the life- and business cycle. Individuals save precautionarily when facing a high degree of earnings risk in order to at least partially insure against potential future changes in earnings and they respond to the realization of an unexpected change in earnings by altering consumption behavior. For both, the precautionary savings motive as well as the consumption response to shocks, the size and persistence of earnings changes matters in determining how much consumption and saving behavior needs to adjust to ensure a certain standard of living today and in the future. In line with that, several recent studies, among others Busch and Ludwig (2023) and De Nardi et al. (2020), show that the existence of higher-order moments in the distribution of earnings changes has important implications for consumption and saving choices. Based on these findings, I explore how the presence of higher-order earnings risk affects different parts of the cross-sectional household distribution and what the resulting equilib-

rium implications for long and short-run wealth inequality are. I then evaluate the implications of these findings for the ability of standard heterogeneous-agent models to match the observed degree of wealth inequality.

The main contribution of this paper is two-fold. First, I evaluate the role of higher-order earnings risk for consumption, earnings and wealth in the cross-section and over the business cycle. In order to do so I use an incomplete markets real business cycle model in which households face aggregate and idiosyncratic income risk, and accumulate wealth to self-insure against shocks to their earnings. To understand the impact of higher-order earnings risk I compare two economies, a canonical economy and a higher-order economy, which only differ in their earnings dynamics. Log-earnings are in both economies the sum of a standard Gaussian transitory shock and a persistent component. In the canonical economy the persistent component has regular normal innovations and, thus, the earnings distribution does not capture any higher-order moments. In the higher-order economy the persistent component is instead calibrated to match the unconditional earnings distributions from recent empirical studies, in particular Guvenen et al. (2014). The second contribution is methodological. I use a new global solution method to solve for the aggregate dynamics of heterogeneous-agent macro models. While the literature on solution methods for heterogeneous-agent macro models is rapidly growing much of the recent progress pertains to the development of perturbation methods (Auclert et al. (2021), Winberry (2018)) and less to global solution methods such as Fernández-Villaverde et al. (2023) and Schaab (2020).

First, I find that earnings inequality increases over the long run in the presence of higher-order moments. Induced by more severe negative shocks, earnings of the bottom quintiles of the earnings distribution decrease relative to the average and top earners. Moreover, matching moments of the 1 and 5-year earnings growth rate distributions induces shocks to earnings to become more persistent. Income and wealth poor households respond to lower earnings and less upside potential by reducing consumption and increasing savings. In contrast, wealthy households behave as permanent income consumers and therefore barely respond to changes to higher-order moments of shocks to their earnings. Since consumption and wealth are strongly correlated with earnings,

the rise in earnings inequality is partially passed through to larger consumption inequality and lower wealth inequality, the wealth Gini falls from 0.71 to 0.68. To put this into perspective, in order to match the same wealth Gini of 0.71 as in the canonical economy the higher-order economy requires discount heterogeneity to increase by 20%. The mild looking changes in the cross-section mask strong consumption responses by the poor. In particular, consumption as a share of cash at hand increases by roughly 13 percentage points for the bottom quintile of the wealth distribution. Thus, the presence of higher-order earnings risk increases the importance of wealth and, thus, precautionary savings as an insurance tool for the poor. This reinforces the known difficulty of standard heterogeneous-agent models to match the empirically observed dispersion of wealth when accounting for higher-order earnings risk. Instead these findings put more emphasis on alternative sources of wealth inequality such as heterogeneity in asset returns which has recently received much attention.[33]

Second, higher-order earnings risk changes the dynamics of the wealth distribution over the business cycle only for the bottom wealth quintiles, leaving the time series of economic aggregates such as capital and output mostly unchanged. As poor households increase savings to insure against worse and more persistent shocks to their earnings their capital income increases. Moreover, the higher-order moments of earnings, particularly excess kurtosis, are partially passed through to consumption and wealth holdings of the bottom wealth quintile. Both effects reduce the correlation between consumption of the wealth poor and aggregate output substantially, from 0.77 in the canonical economy to 0.19 in the higher-order economy.

Lastly, welfare costs of higher-order earnings risk are concentrated at impatient households at the bottom of the wealth distribution who would pay 1.7% of permanent consumption to remain in the canonical economy.

Methodological Contribution. I use General Polynomial Chaos Expansion (GPCE) to solve for the aggregate dynamics of the model. I build on Proehl (2017) and develop a form of GPCE that is suitable as a solution method for heterogeneous-agent macro models, as described below, and

[33]For example, Fagereng et al. (2020) and Bach et al. (2020) find substantial heterogeneity in returns to wealth, Hubmer et al. (2021) argue in a large-scale heterogeneous-agent model that asset return heterogeneity is key to matching the time series of wealth inequality.

show that it generates a law of motion which accurately forecasts aggregate prices in the presented model. As an extension I solve a model with time-varying earnings risk and show that the method also performs well in that context.

GPCE is a global projection method that expands the cross-sectional household distribution μ_t in terms of a series of orthogonal polynomials Ψ_i and thereby approximates the distribution with time-varying coefficients $\alpha_{i,t}$.

$$\mu_t(s_t) = \sum_{i=1}^{n} \alpha_{t,i} \Psi_i(\nu) \tag{3.1}$$

where s_t is a vector of individual state variables and ν is a base random variable based on which the polynomials are generated. GPCE gains its efficiency through a smart choice of the base random variable, in particular a base random variable that is distributed similar to the cross-sectional household distribution. This allows GPCE to achieve a high degree of accuracy with a low dimensional approximation and thereby overcome the curse of dimensionality. I make two contributions with respect to making GPCE suitable and efficient for heterogeneous-agent macro models:

1. One underlying assumption when using GPCE is that the model parameters are independent. When approximating the cross-sectional household distribution this assumption requires the individual state variables to be independent which does not hold in most economic settings and, particularly in heterogeneous-agent models, is in stark contrast with the empirical motivation for the models. I develop a GPCE method that allows for dependence between the individual state variables and thereby makes GPCE suitable to solving heterogeneous-agent models.

2. I show that a short outer iteration scheme that updates the base random variable such that it is similarly distributed as the ergodic household distribution leads to reduced and less biased forecast errors for a given number of aggregate state variables. Moreover, I implement a version of the method that projects on different bases depending on whether the economy is in a recession or an expansion. This is of particular importance when applying GPCE

to economies which at times move far away from the ergodic distribution since it allows to choose the appropriate base distribution and therefore polynomials to project on for different regions of the aggregate state space.

The remainder of the paper is structured as follows. The next section places this paper in the literature. Section 3.3 describes the quantitative model used to analyze the implications of higher-order earnings risk. In section 3.4, I present the global solution method as well as my contributions in detail. Sections 3.5 and 3.6 calibrate the model and present the economic results. Section 3.7 evaluates the accuracy of the computational method in contrast to previously used ones. Section 3.8 extends the computational method by allowing for a time-varying base distribution to evaluate its ability to accurately solve models which exhibit stronger aggregate non-linearities than the benchmark economy studied here. Finally, section 3.9 concludes.

3.2. Related Literature

This paper is related to several branches of the literature. First and foremost, it relates to the long-standing literature studying theories of wealth inequality starting with Bewley (1977), İmrohoroğlu and Imrohoroglu (1989), Huggett (1993), and Aiyagari (1994). The core of the Bewely-Hugget-Aiyagari (BHA) economy builds on idiosyncratic uninsurable shocks to households' earnings. Households have to accumulate non-state-contingent assets to smooth consumption. The shocks endogenously generate ex-post dispersion in wealth as households experience different shocks over time and, thus, have different ability to accumulate wealth. In the basic model wealth inequality is completely determined by the exogenous specification of the earnings process, however, properly calibrated earnings processes fall short of generating the dispersion of wealth observed in the data. An extensive literature builds on the core model and introduces additional sources of wealth inequality such as entrepreneurial ability (Cagetti and De Nardi (2006), Quadrini (2000)) or more recently heterogeneity in returns to wealth (Hubmer et al. (2021)). Starting with Krusell and Smith (1998) a large literature has also explored the dynamics of earnings and wealth inequality in the presence of business cycle fluctuations which affect households across the wealth distribution differently. I contribute to this literature by revisiting the role of earnings dynamics for wealth dispersion,

in particular focusing on the role of higher-order earnings risk.

This paper is further related to the growing literature studying the implications of higher-order earnings dynamics for consumption and wealth. A recent empirical body of papers documents that, in contrast to previous work, individual earnings shocks exhibit sizable higher-order moments in form of negative skewness and excess kurtosis. Incorporating these higher-order dynamics in richer earnings processes has been shown to matter for a variety of economic questions. Golosov et al. (2016) show that it has important implications for optimal redistribution and insurance. In a standard Aiyagari economy Civale et al. (2017) evaluate the implications of higher-order earnings risk for the aggregate capital stock. De Nardi et al. (2020) and Busch and Ludwig (2023) both analyze the implications for consumption-savings decisions over the life-cycle in partial equilibrium life-cycle models. They find that a richer process for earnings increases consumption insurance and moves insurance against persistent shocks closer to the data. Importantly, Busch and Ludwig note that the increased average consumption insurance masks the fact that insurance against negative shocks nevertheless falls. The intuition is that negative skewness coupled with excess kurtosis induces negative shocks to be larger in size and the rise in precautionary savings is not sufficient to fully offset those. I contribute to this literature by analyzing the heterogeneous role of higher-order earnings risk for consumption and wealth across the wealth distribution. However, in contrast to much of the existing literature I allow for general equilibrium effects and business cycle fluctuations. This is important as wealth inequality is not only determined by individual household responses to earnings dynamics but also through their interactions in capital and labor markets which are themselves subject to business cycle fluctuations.

My work is further related to the extensive literature on solution methods for heterogeneous-agent models, and in particular global solution methods.[34] In their seminal contribution Krusell and Smith (1998) develop a global solution method that proposes a parametric law of motion for the aggregate capital stock and is still widely used. More recently, there has been a series of papers on perturbation methods using similar finite-dimensional distribution approximations as I do (Boppart et al.

[34] e.g. Den Haan (1996), Den Haan (1997), Reiter (2010), Maliar et al. (2010), Proehl (2017), Fernández-Villaverde et al. (2023), Schaab (2020)

(2018), Winberry (2018)). I build on and my contribution is most closely related to Proehl (2017) who first introduced General Polynomial Chaos Expansion (GPCE) to an economic setting and whose paper provides important insights into how to implement GPCE in practice. Much of her focus is on the theoretical foundations of the method, proving convergence to the rational expectation equilibrium in a standard Krusell-Smith economy. My paper is also closely related to Schaab (2020) who develops a similar global solution method. While he generates the basis functions to project on in a different way, he also focuses on how to choose the basis functions optimally in order to generate a low-order efficient and non-parametric approximation for the law of motion. I contribute to that literature by developing an efficient version of GPCE that, crucially, allows for individual state variables to be dependent, thereby making it a suitable solution method for heterogeneous-agent models.

3.3. Model

The model builds on Krusell and Smith (1998), thus, it is a general equilibrium model with household heterogeneity and aggregate risk. It differs from Krusell and Smith (1998) at the microeconomic level in two important ways: First, households experience idiosyncratic shocks to their earnings rather than to their employment status. This allows the model to match the estimated earnings dynamics from the data and to obtain more realistic cross-sectional earnings and wealth distributions. Second, households differ in their permanent discount factors which is a common modification to match the cross-sectional dispersion in wealth observed in the data. Generating a realistic degree of wealth dispersion is necessary to understanding the implication of higher-order earnings risk as households across the wealth distribution respond differently to earnings risk. Moreover, calibrating the dispersion of discount factors to match cross-sectional moments of the wealth distribution gives rise to an intuitive measure of how much higher-order earnings risk affects wealth dispersion, that is, the change in discount factor dispersion required to match the same moments.

3.3.1. Technology

A representative firm maximizes profits by renting capital K_t and labor L_t from households to produce a non-storable consumption good Y_t. The firm takes factor prices for capital and labor, R_t and W_t, as given and operates a standard Cobb-Douglas technology subject to aggregate productivity shocks z_t. The static maximization problem is given by

$$\max_{K_t, L_t} Y_t - R_t K_t - W_t L_t.$$

where

$$Y = z_t K_t^{\alpha} L_t^{1-\alpha} \tag{3.2}$$

and α denotes the capital share of output. The mean-reverting productivity shock z_t is the source of aggregate uncertainty in this model and follows an AR(1) process given by

$$\log(z_t) = (1 - \rho_z)\mu_z + \rho_z \log(z_{t-1}) + \epsilon_t^z,$$

with $\mathrm{E}[z_t] = 1$. Input markets are competitive, thus, factor prices R_t and W_t are equal to their marginal products. Capital used in production depreciates at rate δ.

3.3.2. Households

The economy is populated by a continuum of households of unit mass. Households survive from each period to the next with constant probability θ as in Blanchard (1985). Carroll et al. (2015) show that this ensures the existence of an ergodic wealth distribution. Each period a mass $1 - \theta$ of new households is born with zero initial wealth, thus, leaving the overall population size unchanged.

3.3.2.1 Preferences

Households maximize expected discounted utility from consumption with standard time-separable preferences, that is, period utility $u(c)$ is continuous, strictly increasing and concave. Households differ in wealth, labor productivity and discount factors. Discount factor heterogeneity as a tool to

generate a degree of wealth inequality similar to the data was first introduced by Krusell and Smith (1998) who postulated that households face stochastic shocks to their discount factors. Instead, I follow Carroll et al. (2017) and assume households have different but permanent discount factors. This can be interpreted as capturing a variety of channels of heterogeneity such as differences in risk preferences, age or expectations that matter for consumption-savings decisions, and thereby for the resulting distribution of wealth. Discount factors are distributed uniformly on the interval $[\beta - \Delta, \beta + \Delta]$ which I discretize with three possible values. Thus, there are two parameters, the mean discount factor β and the dispersion Δ, that need to be chosen in the calibration.

3.3.2.2 Household Problem

Each period households choose how much to consume and save while supplying labor inelastically to the firm. Households are subject to uninsurable idiosyncratic earnings risk as well as aggregate uncertainty, they can save via risky capital and have otherwise only access to perfect annuity markets. Assets of the deceased are distributed equally among the surviving population and the newborn households are born with zero assets.

Let a_t and y_t denote current asset holdings and labor productivity, respectively, and β_i the permanent discount factor of households of type i. Further, let Z_t be a vector of aggregate state variables consisting of aggregate productivity z_t and the cross-sectional household distribution $\mu_t(\beta_i, a_t, y_t)$ which households need to know in order to predict future prices. Then the recursive household problem is given by

$$V(\beta_i, a_t, y_t, Z_t) = \max_{c_t, a_{t+1}} u(c_t) + \theta\beta_i \mathrm{E}_t \left[V(\beta_i, a_{t+1}, y_{t+1}, Z_{t+1})\right] \tag{3.3}$$

subject to budget constraint, borrowing constraint and aggregate law of motion

$$\begin{aligned} c_t + a_{t+1} &\leq a_t(1 + R(Z_t) - \delta) + y_t W(Z_t) \\ a_{t+1} &\geq 0 \\ \mu_{t+1} &= A(Z_t, z_{t+1}) \end{aligned}$$

Household utility is given by

$$u(c_t) = \frac{c_t^{1-\sigma}}{1-\sigma} \tag{3.4}$$

where σ as usual quantifies risk aversion and the inverse of the intertemporal elasticity of substitution.

3.3.2.3 An Earnings Process with Higher-Order Moments

The focus of this paper is the effect of higher-order idiosyncratic earnings risk on household behavior. Labor productivity y_t is the sum of a transitory and a persistent component

$$y_t = \exp\left(p_t + \epsilon_t\right), \qquad \epsilon_t \underset{iid}{\sim} \mathcal{N}(0, \sigma_\epsilon^2) \tag{3.5}$$

The persistent component p_t follows an AR(1) process given by

$$p_t = \rho p_{t-1} + \eta_t, \qquad \eta_t \underset{iid}{\sim} \mathcal{F} \tag{3.6}$$

The distribution of innovations to persistent earnings $\mathcal{F}$ is calibrated to match recent empirical findings from Guvenen et al. (2021) and exhibits negative skewness and excess kurtosis. Let $\Pi(y_{t+1}|y_t)$ denote transition probabilities and $p(y)$ the invariant distribution of overall earnings that come out of the calibration. I propose a simple way to discretizing the persistent component when matching higher-order moments of earnings building on Civale et al. (2017) which, in contrast to existing methods, does not require a simulation step.[35]

3.3.3. Recursive Competitive Equilibrium

A recursive competitive equilibrium consists of policy and value functions for households in all individual and aggregate states, $c_t(\beta_i, a_t, y_t, Z_t)$, $a_{t+1}(\beta_i, a_t, y_t, Z_t)$, $V(\beta_i, a_t, y_t, Z_t)$, functions for aggregate prices, $R(Z_t)$ and $W(Z_t)$, and an aggregate law of motion for the cross-sectional household distribution $A(Z_t, z_{t+1})$ such that

[35] See appendix C.2 for a detailed description of the discretization method.

1. Given pricing functions and the aggregate law of motion, $R(Z_t)$, $W(Z_t)$ and $A(Z_t, z_{t+1})$, $V(\beta_i, a_t, y_t, Z_t)$ solves the household problem described in (3.3), and $c_t(\beta_i, a_t, y_t, Z_t)$, $a_{t+1}(\beta_i, a_t, y_t, Z_t)$ are the corresponding decision rules.

2. Factor prices are equal to their marginal products and given by

$$R(Z_t) = z_t \alpha \left(\frac{K_t(Z_t)}{L_t} \right)^{\alpha - 1}$$
$$W(Z_t) = z_t (1 - \alpha) \left(\frac{K_t(Z_t)}{L_t} \right)^{\alpha}$$

3. Capital and Labor markets clear

$$K_t(Z_t) = \int a_t \; d\mu(\beta_i, a_t, y_t)$$
$$L_t = \bar{L} = \sum_y y p(y)$$

4. The aggregate law of motion A used by households to forecast the evolution of the cross-sectional distribution is consistent with the realized law of motion induced by household behavior and the exogenous processes for aggregate and idiosyncratic risk.

3.4. The Global Solution Method

This section discusses the methodological contribution and presents the global solution method used to solve for the aggregate dynamics of the economy. Reading this section is not required in order to follow the economic results.

It is well known that heterogeneous-agent models with aggregate non-linearities are difficult to solve. The reason is that households base current actions on forecasts of future prices and in any rational expectations equilibrium those forecasts have to be consistent with how the economy actually evolves, that is, with the future prices that materialize. If the optimal policy function for savings features a constant propensity to save out of current assets and income, the evolution of aggregate capital can be described by an average savings function of a representative household

and the current capital stocks becomes a sufficient statistic for future prices. However, if the propensity to save out of current assets and income is not constant across the individual state space, forecasting the aggregate capital stock requires forecasting the evolution of the cross-sectional household distribution. The distribution then becomes an aggregate state variable which makes the aggregate state space infinite-dimensional. The goal of any solution method is to accurately approximate the aggregate state space in finitely many dimensions, in fact in as few as possible.

3.4.1. Polynomial Chaos Expansion

General Polynomial Chaos Expansion (GPCE) is a projection method that represents some random variables of interest, such as the cross-sectional household distribution, as a function of a basic random variable with some specified distribution. The function is an infinite series of orthogonal polynomials which maps the basic random variable into the space of square-integrable random variables. Thus, the basic idea is to represent the cross-sectional household distribution as a random variable by approximating it as a series of polynomials which themselves are random variables.

I build on Proehl (2017) who first used GPCE to solve a standard Krusell-Smith economy. Their focus was largely on the theoretical underpinnings, among other things proving convergence to the rational expectations equilibrium. The central task for any projection method is to choose the projection base, here the set of polynomials. The advantage of polynomial chaos expansion over other projection methods is that when the basic random variable is chosen such that it is similar to the cross-sectional household distribution to be approximated the order of polynomials required to obtain a good approximation of the household distribution is significantly reduced. Moreover, given a base random variable the set of polynomials can be generated with standard methods. However, when approximating multivariate distributions GPCE builds on the assumption that the variables that make up the multivariate distribution are independent. When applied to economic models, this assumption requires the individual state variables to be independent. In most economic settings individual state variables are not independent, and in heterogeneous-agent models such as the one presented here generating a realistic joint distribution of income and wealth is instead a central motivation. The standard GPCE method, therefore, cannot accurately approximate the law

of motion in these models because the implied independence assumption causes the distribution of the projection base to miss important features of the cross-sectional household distribution to be approximated. To make GPCE suitable for this class of models, I develop an approach to account for dependent base variables and show that it generates an accurate law of motion.

Formally, let $\nu \sim \mathcal{F}$ be some basic random variable and Ψ_i a series of orthogonal polynomials which map ν into the space of square-integrable random variables. Then any square-integrable random variable μ_t can be written as

$$\mu_t = \sum_{i=0}^{\infty} \alpha_{t,i} \Psi_i(\nu) \tag{3.7}$$

where $\alpha_{t,i}$ is a series of scalar coefficients. Here μ_t is a cross-sectional household distribution which is fully characterized by the specific series of coefficients $\{\alpha_{t,i}\}_{i=1}^{n}$. In practice, the above polynomial expansion must be truncated for the aggregate state space to become finite-dimensional and to make the household problem tractable, and thereby computationally feasible. Let the order of truncation be denoted by n, thus, one replaces the true household distribution with an n-th order GPCE approximation. Then the evolution of the approximate household distribution $\mu_{t,n}$ is fully characterized by the evolution of coefficients $\{\alpha_{t,i}\}_{i=1}^{n}$. The number of dimensions in the aggregate state space required to approximate the cross-sectional distribution reduces to n and the household problem becomes tractable.

Naturally, truncating the infinite series in equation 3.7 introduces an approximation error. The quality of the solution method is determined by its ability to approximate the distributions of interest sufficiently well while keeping the order of truncation n low and, thus, the aggregate state space small.

3.4.1.1 Choice of the Base Random Variable

The first task is to choose the distribution of the base random variable which the polynomials take as an argument. It is crucial that this base distribution closely mimics the cross-sectional household distributions that is approximated in the dynamic model. This will allow the truncation order to

be low while still achieving a high degree of accuracy. I propose an outer iteration scheme that chooses and adjusts the base distribution in the following way:

1. Begin with an initial base distribution. Here the cross-sectional household distribution from the steady state model without aggregate risk is a good choice. First, the steady state distribution is easy to obtain and often one solves for the steady state anyway in order to obtain an initial guess for the policy functions. Second, the shape of the ergodic distribution from the dynamic model is often similar to the shape of the steady state distribution.

2. Solve the model with the polynomials based on the initial base distribution and simulate the economy. Compute forecast errors and the ergodic distribution.

3. If forecast errors are too large, update the base distribution by using the ergodic distribution from the simulation. In practice it is useful to dampen the adjustment of the base distribution by using a convex combination of the ergodic distribution and previous base distribution. This is repeated until the forecast errors of the law of motion are sufficiently small or the ergodic and base distributions are equivalent, thus, making further updates redundant. If that is the case while the forecast errors are still too large the order of truncation needs to be increased and the scheme repeated.

The overall shape of the ergodic distribution is usually not sensitive to small forecast errors in the law of motion, thus, only a small number of iterations is necessary until the base and ergodic distribution converge. For example, I use three iterations.

The implementation as described above may not work well in economic settings in which parts of the aggregate state space exhibit strong non-linearities. In that case the cross-sectional household distribution can move far away from the ergodic distribution in those parts of the aggregate state space. This is often the case in models with financial accelerators such as Brunnermeier and Sannikov (2014) or Fernández-Villaverde et al. (2023). In that case, I show that one can project on different base distributions and, thus, different sets of polynomials depending on where in the aggregate state space the economy is. For example, in models with financial accelerators the level of leverage

of the financial sector could determine the current base distribution. In section 3.8, I solve an economy with cyclical earnings risk by projecting on different bases in recessions and expansions, thus, the current base distribution depends on the aggregate level of productivity z_t. The method provides an accurate law of motion in that setting as well.

3.4.1.2 Polynomials with Standard GPCE

Given a choice for the base random variable ν one needs to generate the set of orthogonal polynomials $\{\Psi_i(\nu)\}_{i=1}^n$ on which to project. When approximating multivariate distributions, such as the cross-sectional household distribution over (β, y, a) in the presented model, the standard GPCE method prescribes to choose an independent basic random variable $(\nu_1, \nu_2, \nu_3) = (\nu_\beta, \nu_y, \nu_a)$ for each dimension. For each base random variable ν_j the respective univariate set of polynomials $\{\Psi_{j,i}(\nu_j)\}_{i=1}^n$ is generated separately. The polynomials can be generated using the three-term recurrence relation (see e.g. Gautschi (1982) or Zheng et al. (2015)).[36] The multivariate set of orthogonal polynomials $\{\Psi_i(\nu)\}_{i=1}^n$ is then a tensor product of the univariate polynomials given by

$$\begin{aligned}
\mu_t &= \sum_{i=1}^{\infty} \alpha_{t,i} \Psi_i(\nu) \\
&= \sum_{i=1}^{\infty} \alpha_{t,i} \Psi_i(\nu_\beta, \nu_y, \nu_k) \\
&= \sum_{i=1}^{\infty} \alpha_{t,i} \sum_{i_1,i_2,i_3=1}^{i} \Psi_{1,i}(\nu_\beta) \cdot \Psi_{2,i}(\nu_y) \cdot \Psi_{3,i}(\nu_a) \; \mathbb{1}\{i_1 + i_2 + i_3 = i\}
\end{aligned} \tag{3.8}$$

The corresponding probability density function $f(\nu)$ of the base distribution ν is also given by the tensor product of the univariate pdfs, that is, the joint distribution of the base random variable $\nu = (\nu_\beta, \nu_y, \nu_a)$ is generated by multiplying the marginal pdfs. However, in the presented model the individual state variables (β, y, a) are not independent but instead exhibit strong correlation. Generating the set of polynomials under the independence assumption as described above generates a base distribution with a set of corresponding polynomials that does not closely mimic the cross-sectional household distribution to be approximated. This leads to large approximating errors in the law of motion if accounting for the correlation between the individual state variables is relevant

[36] See appendix C.1 for details on generating the polynomials with the three-term recurrence relation.

for the evolution of prices. For example, this is the case if asset poor agents behave significantly different depending on whether their earnings are low or high or depending on whether they are impatient or patient. As shown and discussed in section 3.7 this applies to my model.

3.4.1.3 A Dependent GPCE Method

To allow for dependence among the base random variables, I instead propose a different way of generating the polynomials and using GPCE. Instead of approximating the joint, unconditional cross-sectional household distribution with a joint base random variable and joint set of polynomials as described in the previous section, I approximate the evolution of the conditional asset distribution $\mu_t(a|\beta, y)$, thus for each (β, y) group, separately. I choose a univariate base random variable $\nu(a|\beta, y)$ and generate a set of univariate polynomials $\{\Psi_i(\nu(a|\beta, y))\}_{i=1}^n$ for each (β, y). As a result, I effectively apply GPCE to each (β, y) asset distribution separately and, thus, the evolution of each (β, y) asset distribution is described by a set of coefficients $\{\alpha_{t,i}(\beta, y)\}_{i=1}^n$. This naturally allows the different base random variable to depend on the discount factor β as well as income y, and thereby allows individual state variables to be dependent. Solving for a law of motion for all $\{\alpha_{t,i}(\beta, y)\}_{i=1}^n$, however, would add n aggregate state variables for each (β, y) group. In order to overcome the resulting curse of dimensionality I solve for the law of motion only at specific combinations of coefficients in the aggregate state space.

I illustrate this by looking at the first-order coefficients $\alpha_{t,1}(\beta, y)$. The first-order polynomials are by default equal to 1, $\Psi_i(\nu) = 1$ and higher-order order polynomials are constructed with zero-mean thereby making the first-order coefficients $\alpha_{t,1}$ equal to the mean of the distribution at time t. First, I discretize the state space of coefficients $\alpha_{t,1}(\beta, y)$ separately for each (β, y) base using the same number of grid points for each one. This yields the first-order coefficients to have different grids across (β, y) groups. Therefore, it allows exactly defining the possible average asset holdings for each (β, y) conditional asset distribution at which the law of motion is solved for. This is not possible with the standard GPCE approach.

In order to overcome the curse of dimensionality, I only solve for points in the aggregate state space where all $\alpha_{t,1}(\beta, y)$ have the same index in their respective grids. For example, if each grid for the

first-order coefficients $\alpha_{t,1}(\beta, y)$ has a total of two grid points I only solve for the law of motion at points in the aggregate state space where all $\alpha_{t,1}(\beta, y)$ are either low or high. Economically, I thereby assume that the average asset holdings of (β, y) groups are positively correlated. For the sake of forecasting prices, it is therefore sufficient to approximate the law of motion at points in the aggregate state space where average asset holdings for all conditional asset distributions $\mu_t(a|\beta, y)$ are simultaneously low or high. Importantly, this assumption does not imply that average asset holdings by (β, y) groups simultaneously increase and decrease by the same amount or even factor but rather that the sign of changes in the average asset holdings is the same over time. The aggregate law of motion then takes as input not the first-order coefficient for all (β, y) asset distributions but instead I choose the weighted average across all groups resulting in an unconditional coefficient

$$\alpha_{t,1} = \sum_{\beta,y} \alpha_{t,1}(\beta, y) p(\beta, y) \tag{3.9}$$

The weights $p(\beta, y)$ are given by the exogenous, invariant mass of agents with individual state (β, y). As a result, the first-order coefficient $\alpha_{t,1}$ will be equal to the aggregate capital in the economy as with standard GPCE despite generating each (β, y) asset distribution with their own separate set of polynomials and coefficients.

I then apply the same assumption when generating higher-order polynomials and discretizing the state space of the corresponding coefficients. As a result, the total number of aggregate grid points used to approximate the cross-sectional distribution is again equal to the order of truncation n.

3.4.1.4 Simultaneously Solving for the Transition and Policy Functions

I will briefly outline the specific steps taken when solving for the law of motion for the unconditional coefficients $f(\{\alpha_{t,i}\}_{i=1}^n) = \{\alpha_{t+1,i}\}_{i=1}^n$.

Going back to Krusell and Smith (1998), the traditional way of finding a law of motion for the cross-sectional household distribution that is consistent with household behavior is to start with a parametric guess for the law of motion and then solve and simulate the model. The guessed law of motion is updated based on the realized law of motion in the simulation and the model is solved again with the updated law of motion. It usually requires many rounds of simulation until the

guessed and simulated laws of motion converge. A major advantage of projection methods such as Dependent GPCE as presented here or the one presented in Schaab (2020) is that they do not need the simulation step but instead solve for a nonparametric law of motion while solving for the policy functions.

First, one initializes the algorithm with some guess for the transition function just as one begins with some initial guess for the policy functions. In this class of models, a good initial guess is usually to assume that the cross-sectional household distribution does not change. Thus, given a set of coefficients $\{\alpha_{t,i}\}_{i=1}^{n}$ in the discretized aggregate state space the initial guess for the law of motion at that point is given by

$$\begin{aligned} f(\{\alpha_{t,i}\}_{i=1}^{n}) &= \{\alpha_{t+1,i}\}_{i=1}^{n} \\ &= \{\alpha_{t,i}\}_{i=1}^{n} \end{aligned}$$

In a given aggregate state $(z_t, \{\tilde{\alpha}_{t,i}\}_{i=1}^{n})$ the law of motion and current policy functions determine the possible aggregate and individual states next period and, thus, the right-hand-side of the Euler equation. One then solves the Euler equation and updates the current guess for policy functions in all individual states. The updated policy functions are then used to update the law of motion in the following way. The current conditional asset distributions $\tilde{\mu}_t(a|\beta, y)$ for all (β, y) is given by the polynomial expansion in the current conditional coefficients

$$\tilde{\mu}_t(a|\beta, y) = \sum_{i=1}^{n} \tilde{\alpha}_{t,i}(\beta, y)\Psi_i(\nu(a|\beta, y))$$

Given the policy function for asset holdings $a_{t+1}(z_t, \{\tilde{\alpha}_{t,i}\}_{i=1}^{n}, \beta, y, a)$, the conditional asset holdings next period $\tilde{\mu}_{t+1}(\bar{a}|\bar{\beta}, \bar{y})$ at some specific point $(\bar{a}, \bar{\beta}, \bar{y})$ in the individual state space evolves according to

$$\tilde{\mu}_{t+1}(\bar{a}|\bar{\beta}, \bar{y}) = \int_{\beta,y,a} p(a|\beta, y)\mathbb{1}\{a_{t+1}(z_t, \{\tilde{\alpha}_{t,i}\}_{i=1}^{n}, \beta, y, \tilde{\mu}_t(a|\beta, y)) = \bar{a}\}p(\bar{\beta})p(\bar{y}|y)d(\beta, y, a)$$

where $p(\bar{\beta})$ is the exogenous, invariant mass of agents with $\beta = \bar{\beta}$, $p(\bar{y}|y)$ is the exogenously deter-

mined transition probability for earnings, and $p(a|\beta, y)$ is the conditional base pdf of group (β, y). One then projects the realized cross-sectional household distribution $\tilde{\mu}_{t+1}$ onto each polynomial in order to obtain the realized conditional coefficients next period.

$$\begin{aligned}\alpha_{t+1,i}(\beta, y) &= \frac{\langle \tilde{\mu}_{t+1}, \Psi_i \rangle}{\langle \Psi_i, \Psi_i \rangle} \\ &= \frac{1}{\langle \Psi_i, \Psi_i \rangle} \int_{\beta,y,a} \tilde{\mu}_{t+1}(a|\beta, y)\Psi_i(a|\beta, y)dF(a|\beta, y) \qquad \forall i\end{aligned}$$

where $dF(a|\beta, y)$ denotes the conditional density of the base random variable for group (β, y). The transition function is then updated with the new unconditional coefficients $\alpha_{t+1,i}$ which follow from equation 3.9. This is repeated until policy functions and transition function converge simultaneously.

3.5. Taking the Canonical Economy to the Data

The model is calibrated to match US data at an annual frequency. I calibrate the model with the canonical specification and then use the same set of parameters in the economy with higher-order earnings risk. The distribution of discount factors is calibrated internally, the remaining parameters are calibrated externally.

Aggregate Risk and Technology. As standard in the literature, I model aggregate shocks as a two-state AR(1) process where the two states correspond to recessions and expansions $z \in \{z_r, z_e\}$. Following Krusell and Smith (1998) and much of the literature I set the standard deviation of aggregate productivity to $\sigma_z = 0.01$ to roughly match the magnitude of business cycle fluctuations in the US. This results in productivities in recessions and expansions to be $z_r = 0.99$ and $z_e = 1.01$, respectively. I follow Busch and Ludwig (2023) in their classification of recessions and expansions which gives rise to the following transition matrix for aggregate productivity

$$\pi(z_{t+1}|z_t) = \begin{pmatrix} 0.388 & 0.612 \\ 0.231 & 0.769 \end{pmatrix} \tag{3.10}$$

Table 3.1: Benchmark Calibration

	Parameter	Value	Source/Target
	Preferences		
σ	Relative Risk Aversion	1	standard
$\bar{\beta}$	Discount Factor	0.9559	K/Y=3
Δ_β	Discount Factor Dispersion	0.0203	Wealth Gini=0.71
ϕ	Survival Probability	0.9833	Avg life-time of 60
	Technology and Aggregate Uncertainty		
α	Capital Share	0.36	standard
δ	Depreciation	0.08	standard
σ_z	Standard Deviation of TFP	0.01	standard
$\bar{y}$	Average Income		E(Y)=1

and corresponding invariant distribution $\pi_z = (0.274, 0.726)$. The capital share is set to $\alpha = 0.36$ and capital depreciates at rate $\delta = 0.08$. Average income $\bar{y}$ in the economy is normalized such that average output over the business cycle is equal to 1.

Preferences and Discount Factor Distribution. Households are assumed to have logarithmic utility. I calibrate the parameters of the discount factor distribution, meaning the average discount factor β and the dispersion Δ, to match an average capital to output ratio of 3 and a Gini coefficient for wealth of 0.71. I choose a wealth Gini at the lower end of estimates for the US to prevent the degree of discount factor heterogeneity from becoming too large as it is the only source of wealth inequality in this model beyond earnings inequality. I set the survival probability ϕ to 0.9833 such that the average life-time of an agent is 60. Table 3.1 provides a summary of the model parameters in the economy with the canonical earnings process.

3.5.1. The Canonical Earnings Process

The earnings process in the canonical economy does not exhibit any higher-order order moments, the transitory as well as the persistent innovations are Gaussian. Table 3.2 provides the chosen parameters. There are three parameters to calibrate, the variances of both shocks and the persistence of the persistent component. The objective of this paper is to analyze the impact of higher-order earnings risk by comparing its predictions in a standard heterogeneous-agent model to the canonical earnings process used in much of the literature. Thus, to make my results comparable to similar

Table 3.2: Benchmark Earnings Process

Parameter	Value	Description
ρ_η	0.965	Persistence persistent shocks
σ^2_η	0.035	Variance persistent shocks
σ^2_ϵ	0.07	Variance transitory shocks

papers in the literature I set all parameters to standard values that are widely used.[37]

3.5.2. An Earnings Process with Higher-Order Moments

In order to generate higher-order moments in persistent earnings I let the innovations to the persistent component follow a mixture of normal distributions. The full earnings process is then given by

$$y_t = \exp\left(p_t + \epsilon_t\right), \qquad \epsilon_t \underset{iid}{\sim} \mathcal{N}(0, \sigma^2_\epsilon) \tag{3.11}$$

$$p_t = \rho p_{t-1} + \eta_t, \qquad \eta_t \underset{iid}{\sim} \begin{cases} \mathcal{N}(\mu_1, \sigma^2_1) \text{ with } p_1 \\ \mathcal{N}(\mu_2, \sigma^2_2) \text{ with } 1 - p_1 \end{cases} \tag{3.12}$$

I calibrate and discretize the above process with GMM by simultaneously matching a set of moments of the earnings process as well as a set of statistics of the earnings distribution (see appendix C.2 for details). The latter is possible because labor is supplied inelastically, thus, one can compute statistics of the earnings distribution directly from the earnings grid and stationary earnings distribution. For the earnings process I target estimates of earnings moments from Guvenen et al. (2021). While some related papers such as De Nardi et al. (2020) provide discretized processes and transition matrices there are two advantages from discretizing myself. First, De Nardi et al. (2020) solve a partial-equilibrium model which allows them to use a large number of grid points to obtain a better match, too large to be feasible in my general equilibrium setting with aggregate risk. Second, the provided earnings process exhibits significantly less earnings inequality than observed in the data, however, roughly matching the empirically observed earnings distribution is key when analyzing implications for wealth inequality. Table 3.3 provides the calibration results for

[37] For example, Krueger and Ludwig (2016), Busch and Ludwig (2023), Storesletten et al. (2004).

Table 3.3: Unconditional moments of log earnings for higher-order earnings process, targeted moments from Guvenen et al. (2021)

	Targets	Discretized process
Variance(x)	0.58	0.59
Variance($\Delta_1 x$)	0.23	0.19
Skewness($\Delta_1 x$)	-1.35	-1.39
Kurtosis($\Delta_1 x$)	17.8	15.72
Variance($\Delta_5 x$)	0.46	0.51
Skewness($\Delta_5 x$)	-1.01	-1.25
Kurtosis($\Delta_5 x$)	11.55	9.88

Table 3.4: Cross-sectional wealth, earnings and consumption distributions by quintiles for higher-order (H-O), canonical (Can) and Data.
Data estimates from Krueger et al. (2016) and Chang et al. (2019)

	Wealth			**Earn**			**Cons**		
	H-O	Can	Data	H-O	Can	Data	H-O	Can	Data
Bottom 20%	1.6%	1.3%	-0.5%	4.2%	5.8%	4.5%	5.7%	6.8%	6.5%
20%-40%	3.6%	3.1%	0.5%	7.5%	9.7%	9.9%	8.2%	12%	11.4%
40%-60%	6.6 %	5.5%	5.1%	16%	15%	15.3%	18%	16%	16.4%
60%-80%	14%	13%	18.7%	27%	24%	22.8%	28%	24%	23.3%
80%-100%	74%	77%	76.2%	45%	46%	47.5%	39%	42%	42.4%
Gini	0.68	0.71	0.71	0.43	0.41	0.42	0.37	0.35	0.36

the moments of the earnings process. Apart from moments for 1 and 5-year earnings growth rates from Guvenen et al. (2021), I also target the variance of log earnings from the canonical process to ensure better comparability between the economies.

3.6. The Cross-Sectional Household Distribution

I first compare the cross-sectional distributions of income, earnings and wealth across the two economies. As a benchmark, I also report the observed cross-sectional facts from the data as estimated by Krueger et al. (2016) and Chang et al. (2019). Matching these cross-sectional distributions is important in a variety of settings, particularly when moving to normative questions such as the welfare costs of business cycles in the presence of higher-order earnings risk, and policy design implications. Table 3 provides the ergodic cross-sectional distributions of wealth, earnings and consumption by their respective quintiles from simulating the economies. To begin with, the canonical model is able to match the cross-sectional distributions well while only targeting the

wealth Gini in the calibration. However, it requires the discount factor dispersion to be relatively large at $\Delta_\beta = 0.0203$ as shown in table 3.1. Specifically, the earnings and consumption distributions are close the data. As for the wealth distribution the model fits the overall shape, however, has the known issue that the bottom quintiles accumulate too much wealth. There are several reasons for that. While the earnings process is calibrated to match earnings dynamics post government transfer it still misses out on some important parts of the social security net for low-income households which reduce their precautionary savings motive. For example, in the canonical economy households in the bottom earnings state earn roughly 8.2% of average earnings, significantly less than what unemployed households usually receive through unemployment insurance in the US. As a comparison, the replacement rate of unemployment insurance for an individual who earned the average wage prior to job loss is 54% according to the OECD. As a result, households at the bottom of the wealth distribution exhibit too strong of a precautionary savings motive. In addition, households cannot borrow in this setting and, thus, the model cannot generate the negative wealth holdings of the bottom quintile observed in the data.

With the same calibration, the model with higher-order earnings risk matches the three cross-sectional distributions still fairly well but worse than the canonical model across the board. The induced earnings distribution is more unequal than in the canonical case and the data. In particular, the bottom quintiles now receive a smaller share of overall earnings. Excess kurtosis and negative skewness increase the dispersion of earnings at the bottom with earnings in the bottom earnings state falling relative to average and top earnings. As a result, more earnings are shifted from the bottom two quintiles to the top three quintiles. In addition to those changes in the earnings distribution the higher-order earnings process is also more persistent than the canonical one. This is mostly induced by matching moments of 1 and 5-year earnings growth rates. As a result, low-income households not only earn less but also are less likely to become lucky and climb up the earnings latter. Since consumption and wealth are both strongly correlated with earnings, the increase in earnings inequality is partly passed through to consumption and wealth. Low-earnings households respond to the change in their earnings prospects by increasing savings and reducing consumption. As a result, consumption of the bottom two quintiles falls while consumption of the

Table 3.5: Earnings distribution by wealth quintiles, averaged over all simulation periods. Data estimates from Krueger et al. (2016) and Chang et al. (2019)

	Higher-Order	Canonical	Data
Bottom 20%	10%	13%	9.8%
20%-40%	14%	14%	13%
40%-60%	19%	19%	18%
60%-80%	27%	24%	22%
80%-100%	30%	30%	37%

top three quintiles rises, leading to an increase in consumption inequality as measured by the Gini coefficient. Similarly, wealth dispersion falls as the bottom quintiles increase their savings which results in a decrease of the wealth Gini from 0.71 to 0.68.

This is relevant for the existing large empirical and quantitative literature that tries to identify potential mechanisms to explain and generate the degree of observed wealth dispersion. These findings suggest that the correlation between earnings and wealth dispersion is weakened by the presence of higher-order risk which in turn increases the importance of alternative sources of wealth dispersion. In order to further quantify the reduction in wealth dispersion I recalibrate the higher-order economy to match the wealth Gini. The required dispersion in discount factors rises by 20% to roughly $\Delta_\beta = 0.025$.

Table 3.5 reports the earnings distribution by wealth quintiles for the two economies as well as from the data. In line with the analysis above, the share of earnings at the bottom of the distribution falls significantly from 13% in the canonical economy to 10% in the higher-order economy. As a result, the higher-order earnings dynamics help to match the bottom of the distribution as the canonical economy generates too little earnings dispersion among the wealth poor.

In conclusion, the presence of higher-order earnings risk has quantitatively important implications for the bottom of the earnings and consumption distribution while the wealth distribution is only affected mildly. This is mostly due to the fact that the bottom quintiles of the wealth distribution hold little wealth such that even relatively large changes in their consumption-savings behavior do not translate into significant changes in wealth dispersion.

Table 3.6: Consumption share by wealth quintiles measured as consumption over cash at hand, averaged over all simulation periods.

	Higher-Order	Canonical
Bottom 20%	54%	67%
20%-40%	44%	52%
40%-60%	37%	45%
60%-80%	30%	29%
80%-100%	13%	13%

3.6.1. Consumption Responses across the Distribution

To get a better understanding of the heterogeneity in responses to higher-order earnings risk that the above-described changes in the cross-sectional household distribution imply, a closer look at consumption-savings decisions across the wealth distribution is useful. Table 3.6 reports consumption as a fraction of cash at hand by wealth quintiles averaged across the respective quintile. The table shows that the mild changes in the cross-sectional distributions mask more drastic changes in the behavior of wealth-poor households. The bottom three wealth quintiles reduce their consumption share significantly, with the share falling for the bottom quintile by 13 percentage points. In contrast, the top two wealth quintiles spend roughly the same fraction of asset holdings on consumption as in the canonical economy. Wealthy households behave as permanent income consumers and therefore react little to changes in the higher-order moments of their earnings dynamics but instead primarily care about expected earnings. In order to generate negative skewness while keeping the other moments constant, in particular the variance, additional probability mass has to be moved to high earnings states making good shocks more likely. As a result, expected earnings increase slightly and rich households respond to the increase in permanent income by reducing savings slightly. While the response by the wealthy is very weak it still offsets the increase in savings by the bottom quintiles. The reason is that the top wealth quintile holds roughly 75% of the total wealth in the economy. Aggregate capital and, thus, prices are therefore the same in both economies in equilibrium.

Table 3.7: Correlation between aggregate consumption of each wealth quintile and aggregate output as well as kurtosis of total wealth held by each quintile.

	corr(C,Y)		**Kurtosis Wealth**	
	H-O	Can	H-O	Can
Bottom 20%	0.19	0.77	6.2	2.4
20%-40%	0.37	0.34	3.5	2.7
40%-60%	0.34	0.42	2.8	3
60%-80%	0.51	0.54	2.5	2.6
80%-100%	0.66	0.61	2.6	2.6

3.6.2. Time Series Dynamics

The time series dynamics for economic aggregates are very similar in the two models. This is consistent with the above analysis showing that the presence of higher-order earnings risk primarily affects the lowest wealth quintiles while wealthier households behave as permanent income consumers and barely alter their consumption-savings behavior. The lowest wealth quintile holds only a very small share of aggregate wealth and, as a result, changes in its consumption-savings behavior are not passed through to aggregate capital, output and prices. Nevertheless, as consumption behavior of the wealth poor is altered by higher-order earnings risk so are time series dynamics of their consumption. Table 3.7 shows by wealth quintile the correlation between consumption and aggregate output as well as kurtosis of total wealth held by that quintile. Higher savings by the poor increase their ability to smooth consumption and, as a result, correlation between consumption of the bottom wealth quintile and aggregate output drops. In addition, the time series of total wealth held by the bottom wealth quintile exhibits substantial excess kurtosis in the higher-order economy. The reason is that for those households cash at hand consists primarily of earnings and, therefore, wealth holdings adopt the higher-order moments of earnings. As a result, variation in consumption of wealth poor households over time is predominantly driven by large changes to their earnings rather than business cycle fluctuations in the higher-order economy. This further reduces the correlation between their consumption and aggregate output compared to the canonical economy.

Table 3.8: Average, ergodic CEV for impatient and patient households by wealth quintile of the impatient and patient distribution, respectively.

CEV	Impatient	Patient
Bottom 20%	-1.7%	0.68%
20%-40%	-1.1%	0.42%
40%-60%	1.3%	2.1%
60%-80%	1.9%	2.4%
80%-100%	1.4%	2.2%

3.6.3. Welfare Costs of Higher-Order Earnings Risk

Given the documented heterogeneity in consumption responses across the wealth distribution it is natural to quantify the cost of higher-order earnings risk in utility terms. In order to do so, I ask in which world households prefer to live. In particular, I look at average expected life-time utility by wealth quintile for each discount factor group and calculate the consumption equivalent variation (CEV) that a household needs to receive in the canonical economy in order to be indifferent between staying in the canonical economy and moving to the higher-order economy. Thus, I assume that households keep their discount factor type and remain in the same wealth quintile among households with the same discount factor. Calculating the CEV separately for each discount factor group is useful as consumption patterns and welfare costs substantially differ between those groups. The CEV for wealth quintile Q of discount factor group β is then given by

$$W(Q,\beta) = \exp\{(1-\beta)\left(V_{\text{HoM}}(Q,\beta) - V_{\text{Can}}(Q,\beta)\right)\} - 1 \tag{3.13}$$

where $V_{\text{i}}(Q,\beta)$ is the average life-time value of households in wealth quintile Q of discount factor group β, averaged across all simulation periods. Table 3.8 reports those CEV for the patient and impatient households. Note, the wealth quintiles here refer to the wealth distribution of impatient and patient households separately, not the overall wealth distribution. Consistent with the previous analysis welfare costs of higher-order earnings risk decrease in wealth. An impatient household in the bottom wealth quintile would accept a permanent decrease in consumption of 1.7% in the canonical economy instead of moving to the higher-order economy while remaining impatient and in the bottom wealth quintile of the impatient households. In contrast, wealthy

impatient households prefer the higher-order economy and require an increase of 1.3%, 1.9%, and 1.4% in permanent consumption for the top three wealth quintiles respectively in order to stay in the canonical economy. As shown in table 3.5 the earnings share held by the fourth wealth quintile increases significantly when allowing for higher-order earnings dynamics which is the reason that those household benefit most from the presence of higher-order earnings risk. The overall pattern is similar for patient households though they prefer to live in the higher-order economy across all wealth quintiles. Recall, that these are wealth quintiles of the patient wealth distribution. Since poor patient households respond stronger to higher-order earnings risk and accumulate significantly more wealth than impatient household they are better insured against the additional earnings risk. Consistent with that, the share of patient households in the bottom quintile of the overall wealth distribution falls from 12% in the canonical economy to 8% in the higher-order economy. Wealthy households prefer the higher-order economy as they behave as permanent income consumers and permanent income increases slightly as discussed. Importantly, this analysis abstracts from potential transition dynamics between economies with different earnings dynamics and merely asks which ergodic world households prefer assuming they keep their characteristics. It is therefore intuitive that primarily poor impatient households experience significant welfare costs of higher-order earnings risk because those households stay poor in the higher-order economy.

3.7. Evaluating the Solution Method

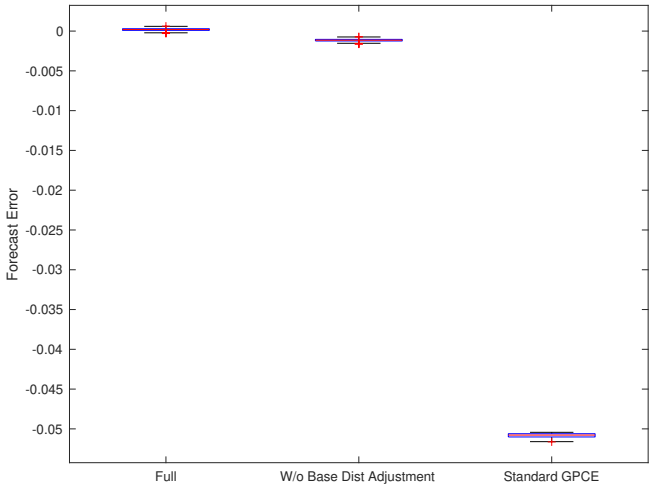
As discussed in 3.4, I make two contributions that make GPCE efficiently applicable to heterogeneous-agent models. First, my approach allows for dependent random base variables and, thus, dependent individual state variables. Second, I implement an outer iteration scheme that adjusts the base random variable, and thereby generates a projection base that more closely mimics the realized distributions in the simulation. In order to understand the corresponding gains, I solve the canonical model with three different methods: The here proposed full Dependent GPCE method with both adjustments (full), the Dependent GPCE method without adjusting the initially chosen steady state distribution as base distribution (noBDadj), and finally with the standard GPCE method as used in Proehl (2017) which builds on the independence assumption. In all cases, the order of truncating is three. While forecast errors reduce further with an order of four the gains are rela-

tively small and have no meaningful effect on economic aggregates. I use the canonical economy for this evaluation as the performance of the solution method in the two economies is essentially equivalent.

3.7.1. One-Period Ahead Forecast Errors

Figure 3.1 shows boxplots for the distributions of one-period ahead forecast errors for aggregate capital in all three models. Forecast errors are computed as the percentage deviation of forecasted from realized aggregate capital and generated by simulating the economies for 1000 periods (after burn-in).

Figure 3.1: Boxplot for one-period ahead forecasting error distribution, different GPCE approaches



The largest improvement for the accuracy of the law of motion measured at the one-period horizon comes from allowing for the base random variable to be dependent. The average forecast error falls by roughly 3 orders of magnitude, from -0.053 to 0.0002, when moving from the standard to the full Dependent GPCE method. Moreover, the standard GPCE method produces significant bias in its forecasts, systematically forecasting smaller than realized capital stocks. The independence assumption of the standard GPCE method is violated in this model due to correlation between wealth and income as well as wealth and permanent discount factor type. Particularly the latter is strong and therefore responsible for the large forecast errors generated by standard GPCE. The standard Krusell-Smith economy as solved by Proehl (2017) does not exhibit any preference

heterogeneity and idiosyncratic income risk only exists in the form of unemployment risk. The unemployed make up a small fraction of the total mass of households and are sufficiently wealthy to have similar propensities to save out of cash at hand as the employed. As a result, the evolution of the economy is almost entirely determined by the employed agents and there is no relevant heterogeneity in individual states beyond wealth dispersion among the employed. However, that is not the case in my model which results in large forecast errors with the standard GPCE method. To get a clearer picture of the gains from the second contribution figures 3.2 and 3.3 plot the forecast error distributions for the Dependent GPCE method with and without base distribution adjustment.

Figure 3.2: Full Dependent GPCE method

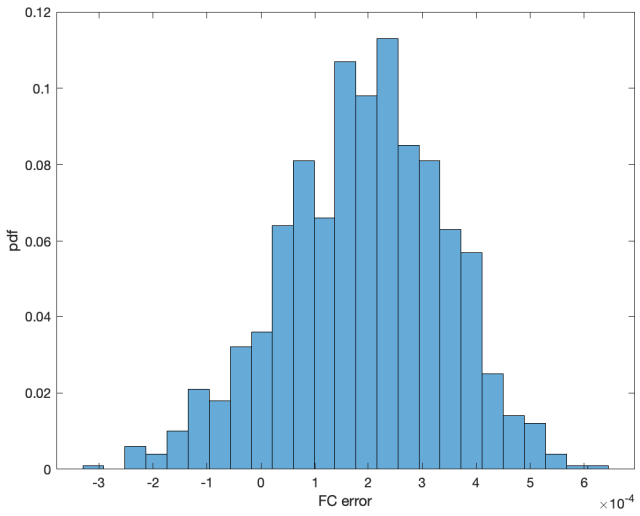
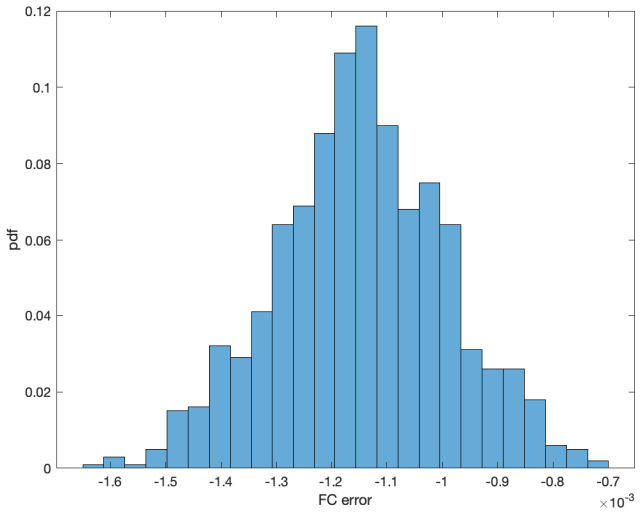
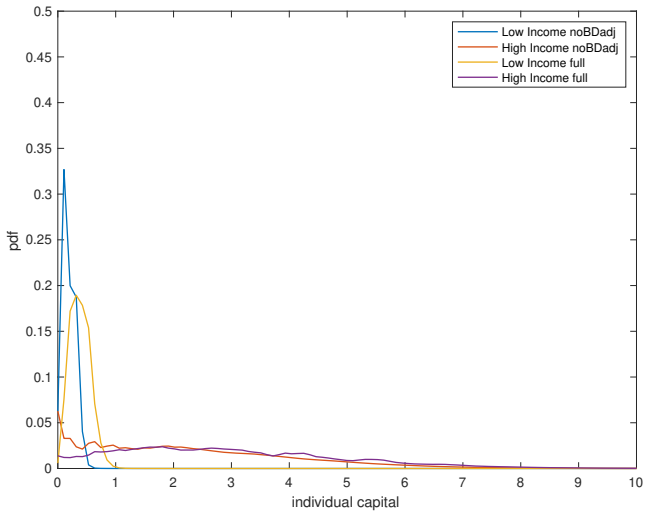


Figure 3.3: Dependent GPCE method without base distribution adjustment



Adjusting the base distribution to be distributed as the realized ergodic distribution yields two important improvements. First, average forecast errors in absolute terms fall roughly by one order of magnitude. Second, adjusting the base random variable reduces the bias in forecasts. Recall that the base random variable is the basis for the polynomials which in turn determine for a given set of coefficients the sample cross-sectional distributions on the aggregate grid based on which the approximate law of motion is derived. The significant and consistent downward bias in forecasts every period without adjustment shows that the shape of the steady state distribution differs in ways from the realized distributions in the dynamic model that are relevant for the evolution of prices. This is confirmed when comparing specific (β, y) base distributions from the two methods. Figure 3.4 plots those distributions for the lowest and highest impatient earnings groups for the Dependent GPCE method with (full) and without base distribution adjustment (noBDadj).

Figure 3.4: Comparison of base distributions for impatient low earnings and impatient high earnings group across methods



The yellow and blue graphs show the base distributions for impatient households in the lowest earnings group for the Dependent GPCE method with and without base distribution adjustment, respectively. In the dynamic model, income-poor households at the bottom of the wealth distribution have a stronger precautionary savings motive than in the stationary economy due to the presence of aggregate risk and, therefore, accumulate more wealth. Consequently, adjusting the base distribution based on the ergodic distribution significantly reduces the mass of households at or close to the borrowing constraint, resulting in the shift from blue to yellow. The same holds true for the base distributions of high earnings households. Moreover, the shape of the base distributions for low and high earnings households are very different as expected when allowing for dependent base variables because earnings are persistent and, thus, correlated with wealth. In contrast, with the standard GPCE method that relies on the independence assumption the conditional base distributions for different earnings groups are equivalent in their shape by construction.

3.7.2. Do Forecast Errors Accumulate?

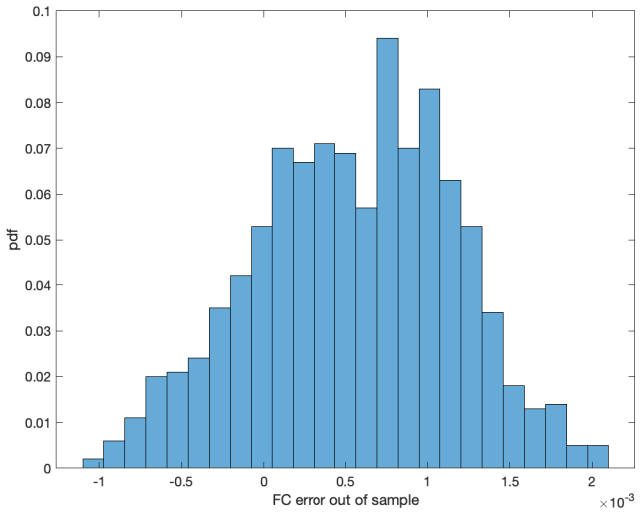
As discussed by Den Haan (2010) one-period ahead forecast errors do not properly assess whether forecast errors accumulate over time. The reason is that current aggregate state variables based on which forecasts are being made are updated every period during the simulation and are, thus,

equal to the true, realized aggregate state variables. In order to evaluate the accuracy of the law of motion further, I therefore check the accumulation of forecast errors. Following Den Haan (2010), one compares the realized aggregate capital stocks from the full simulation to the series of aggregate capital stocks generated by purely iterating over the law of motion. I start with the same initial aggregate state, number of burn-in and simulation periods as in the proper simulation and generate the out-of-sample sequence of aggregate states in the following way

$$\{\{\alpha_{0,i}\}_{i=1}^{n}, \{\alpha_{1,i}\}_{i=1}^{n}, \{\alpha_{2,i}\}_{i=1}^{n}, \{\alpha_{3,i}\}_{i=1}^{n}, ...\} =$$
$$\{\{\alpha_{0,i}\}_{i=1}^{n}, f(\{\alpha_{0,i}\}_{i=1}^{n}), f(f(\{\alpha_{0,i}\}_{i=1}^{n})), f(f(f(\{\alpha_{0,i}\}_{i=1}^{n}))), ...\}$$

Figure 3.5 plots these out-of-sample forecast error distribution for the dependent GPCE method.

Figure 3.5: Full Dependent GPCE Method



Compared to the one-period ahead forecast errors the mode of the distribution is shifted to the right with larger errors on average. However, in absolute terms the out-of-sample sample forecast errors are still small showing that the approximate law of motion is accurate over time and forecast errors do not systematically accumulate to large errors. In particular, the average out-of-sample forecast deviates from realized capital by roughly 0.07%. This does not hold for the standard GPCE

Table 3.9: Average and maximum forecast errors when evaluating the law of motion out of sample

Error\Method	Full Dependent GPCE	Standard GPCE
1-period ahead	$2 \cdot 10^{-4}$	0.0529
out-of-sample	$7 \cdot 10^{-4}$	0.2080

method. Table 3.9 documents the average absolute one-period ahead and out-of-sample forecast errors for the Dependent and standard GPCE methods, respectively. While average forecast errors between out-of-sample and one-period ahead increase by roughly the same factor the standard GPCE method starts at a significantly higher base as it already produces substantial average one-period ahead forecast errors. In particular, forecasts miss realized capital on average by 5.29%. This results in an accumulated average out-of-sample forecast error of 20.8% for the standard GPCE method. To put this into perspective, the percentage standard deviation of the aggregate capital stock in this economy is roughly 0.6%, meaning that the simulated out-of-sample economy moves to aggregate capital stocks significantly and consistently below the interval of realized capital stocks in the proper simulation when using the standard GPCE method. In contrast, the average out-of-sample forecast error of 0.07% of the Dependent GPCE method is small relative to average variations of the capital stock.

3.8. GPCE with Time-Varying Base Distributions

Many interesting economic settings are such that the cross-sectional household distribution at times moves far away from the ergodic distribution. Models with financial intermediaries often exhibit strong aggregate non-linearities between the crisis and regular region of the aggregate state space. In that case, instead of using one general base distribution and set of polynomials it is more efficient to choose different bases for different regions in the aggregate state space. In order to assess the accuracy of the dependent GPCE method when projecting on different bases, I solve a model with cyclical earnings risk and different base distributions in recessions and expansions.

The model is identical to the one presented in section 3.3 apart from the earnings process. The specification of the earnings process changes in that the variance of the persistent component now

Table 3.10: Earnings process with cyclical variance, targets based on Storesletten et al. (2004)

Parameter	Value	Description
$\sigma_y^2(R)$	0.6	Variance earnings shocks in Recession
$\sigma_y^2(E)$	0.33	Variance earnings shocks in Expansion
σ_ϵ^2	0.22	Variance transitory shocks

depends on the aggregate state z_t as in Storesletten et al. (2004).

$$y_t = \exp\left(p_t(z_t) + \epsilon_t\right), \qquad \epsilon_t \underset{iid}{\sim} \mathcal{N}(0, \sigma_\epsilon^2) \tag{3.14}$$

$$p_t = \rho p_{t-1} + \eta_t(z_t), \qquad \eta_t \underset{iid}{\sim} \mathcal{N}(0, \sigma_\eta^2(z_t)) \tag{3.15}$$

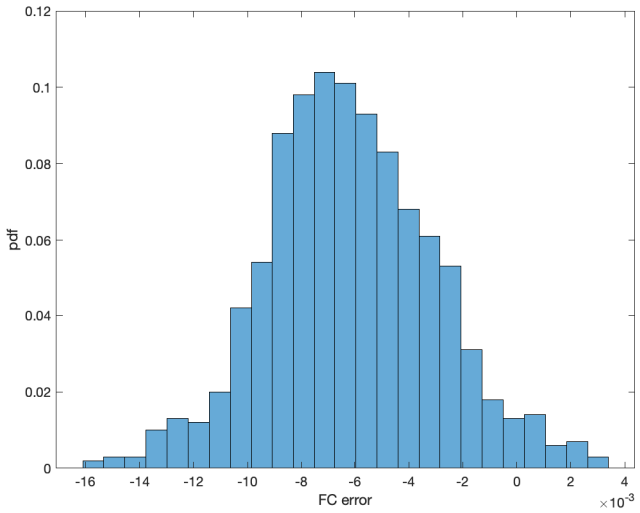
Table 3.10 depicts the calibration of the process which targets the estimates from Storesletten et al. (2004). I use their reported moments for the persistent and transitory components and compute the implied moments of the overall earnings process. I then use GMM to calibrate and discretize the persistent earnings processes in recessions and expansions while ensuring a common earnings grid. The targets are the implied moments for earnings.[38]

3.8.1. Evaluating Cyclical Dependent GPCE

I evaluate the solution method again based on forecast errors for aggregate capital. For comparability, I truncate the polynomial at order three as in the main model. Figure 3.6 shows one-period ahead forecast errors for capital.

[38] I use the same discretization method as for the higher-order process but without targeting higher-order moments. See appendix C.2 for more details.

Figure 3.6: One-period ahead forecasting error distribution with time-varying base distribution



One-period ahead forecast errors are roughly one order of magnitude larger than in the non-cyclical economy in which the base distribution is constant. On average households miss the realized capital stock with their forecasts by roughly 0.5%. While the forecast errors increase relative to the canonical economy this is to be expected as capital is also substantially more volatile with a percentage standard deviation of 3.75%. The intuition is that countercyclical earnings risk amplifies aggregate fluctuations by inducing a countercyclical precautionary savings motive.

3.9. Conclusion

This paper has introduced higher-order earnings dynamics consistent with recent empirical findings into a workhorse heterogeneous-agent real business cycle model. Compared to canonical earnings processes, higher-order earnings risk induces larger earnings inequality, particularly through more persistent and lower earnings at the bottom of the distribution. Low-income households at the bottom of the wealth distribution respond by reducing consumption as a share of total resources by roughly 13 percentage points in order to increase precautionary savings. In contrast, wealthier households are not strongly affected by changes in the higher-order moments of earnings shocks as they behave as permanent income consumers. As a result, wealth inequality falls while consumption inequality increases, reinforcing the known issue of generating the degree of wealth dispersion

observed in the data. Since effectively only poor households adjust their consumption-savings pattern the aggregate time series dynamics of the model do not change in the presence of higher-order earnings risk, leaving the dynamics of aggregate capital, output and consumption mostly unchanged. However, higher-order earnings risk does imply substantial welfare costs for wealth poor households reaching 1.7% of permanent consumption. There are two reasons for that. First, average consumption falls for the poor as the precautionary savings motive increases. Second, for poor households resources to spend consist mostly of earnings. Therefore, the higher-order moments of earnings are passed through to consumption dynamics, and particularly excess kurtosis induces larger changes in consumption which lowers welfare.

APPENDIX A

Appendix for Chapter 1

A.1. Further Empirical Evidence

A.1.1. Cohort Effects Across US States

I estimate cohort effects of low-skilled and routine workers over time across US states in order to test whether unionization has specifically affected the price of human capital of incoming cohorts during the automation transition. Following Heckman et al. (1998), and more recently Lagakos et al. (2018) and Fang and Qiu (2023), I decompose earnings growth into cohort, experience and time effects for cohorts born between 1940 and 1980 at the state and education (and occupation) level. Experience effects measure human capital growth over the life cycle, cohort effects measure the relative human capital level of a cohort at labor market entry, and time effects capture growth of the price of human capital over time. Thus, cohort effects essentially measure the value of human capital of each cohort at labor market entry in units of wages. Therefore, comparing the cohort component of earnings growth across states with high and low unionization in routine occupations quantifies by how much more the value, or the marginal product of labor, of incoming routine workers, measured in units of entry wages, has declined in states with high routine unionization relative to states with low routine unionization. It is well known that experience, cohort, and time effects cannot be separately identified without further assumptions due to perfect collinearity. In order to solve the identification issue, I closely follow the literature, and in particular Fang and Qiu (2023), by adopting the standard identification strategy first used by Heckman et al. (1998). The identifying assumption is that there is no experience growth in the final years of a worker's career which is based on theories of life cycle wage growth.[39] To see this, denote log wage $w_{i,c,t}$ of individual i from cohort c at time t as

$$w_{i,c,t} = p_t + h_{c,t} + \epsilon_{i,c,t}, \qquad \text{where } E_i[\epsilon_{i,c,t}] = 0.$$

[39] See, e.g., Rubinstein and Weiss (2006)

Further decompose the cohort component into entry level human capital $s_c = h_{c,c}$ and return to e years of experience $r_{c,e} = r_e$ according to

$$w_{c,t} = p_t + s_c + r_e,$$

where p_t reflect time effects, s_c reflect cohort effects, and r_e reflect experience effects. It is straight forward to see that perfect collinearity $e = t - c$ now results in non-identification

$$w_{c,t+\tau} - w_{c,t} = p_{t+\tau} - p_t + s_c - s_c + r_{e+\tau} - r_e.$$

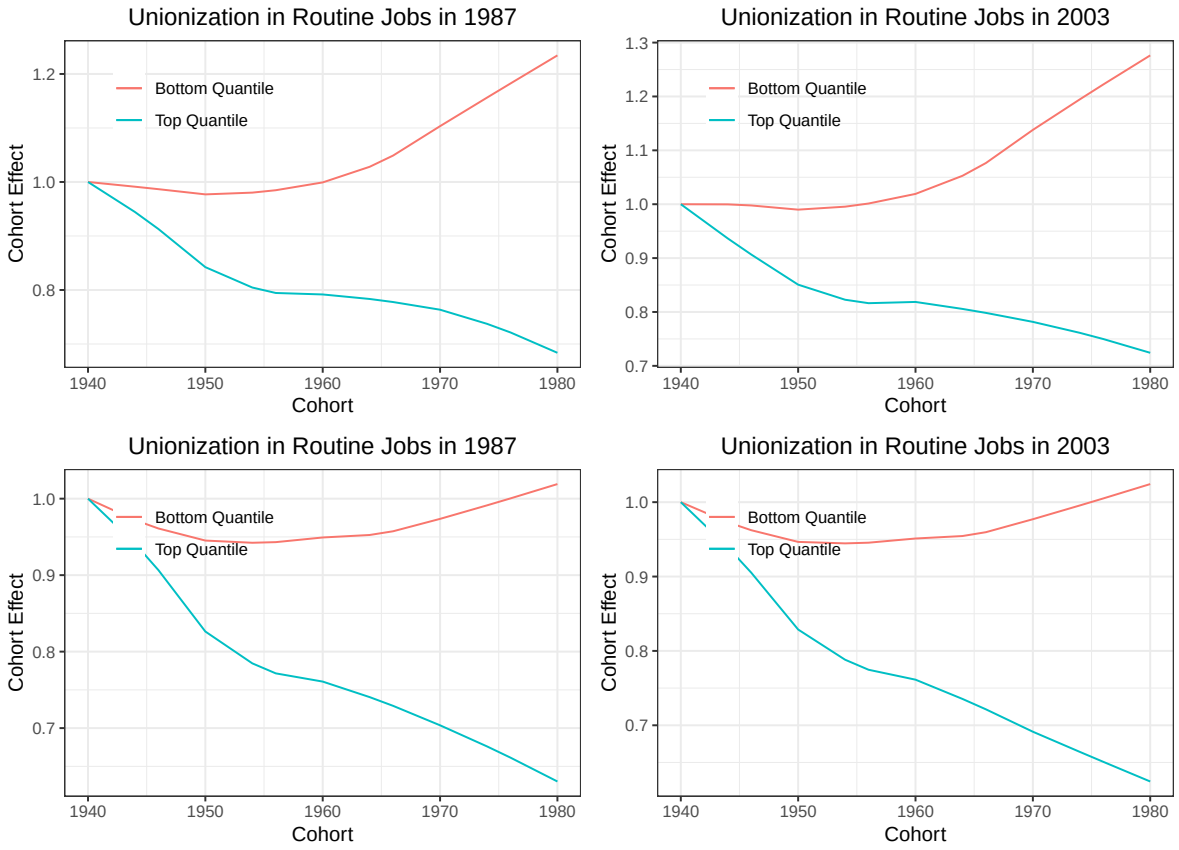
However, with the identifying assumption that there is no experience growth in final years of a worker's career, $r_e = 0$ for cohorts with $e \geq \bar{e}$, the above equation reduces to the following for workers of cohorts with $e \geq \bar{e}$:

$$w_{c,t+\tau} - w_{c,t} = p_{t+\tau} - p_t.$$

Thus, the assumption allows to identify common time effects through older cohorts. Since those time effects by definition are common across cohorts, this then allows for the identification of experience and cohort effects for all other cohorts. I apply the above estimation to decompose repeated cross-sectional annual earnings (total income) profiles from CPS non-parametrically into experience, cohort, and time effects at the state level.

Figure A.1 displays the average cohort effects for highschool dropouts for states in the bottom and top quartile of unionization. In particular, the x-axis shows cohorts by birth year, the y-axis shows the cohort component of entry wages for each cohort relative to the cohort born in 1940. For the top row of plots states are weighted equally when ranked by percentile of unionization while the bottom row shows results when weighing states by their overall routine employment decline between 1980 and 2010. In the left column of plots, unionization percentiles are computed based on routine unionization in 1987 while routine employment in 2003 is used in the right column.

Figure A.1: The panels compare the average estimated cohort effect in states with high and low routine manual unionization.



The top right panel shows that entry wages in states with high routine unionization decline relative to states with low routine unionization from 1940 onward. However, entry wages diverge more strongly from 1960 onward, that is, for cohorts that entered the labor market after 1980. Thus, labor market entry conditions for highschool dropouts have deteriorated particularly from 1980 onward in states with high routine unionization. Moreover, the results are very similar when using the 2003 routine unionization. This is an important robustness test as it bolsters the MSA level analysis which uses average routine manual unionization in a MSA between 1995 and 2005 as explanatory variable since CPS coverage at the MSA level is not sufficient before 1995.

A.1.2. Inflow-Outflow Decomposition

Following Cortes et al. (2020), I use monthly individual-level matched CPS data, and classify every individual observation into 9 mutually exclusive employment states: Non-routine cognitive (NRC), routine cognitive (RC), routine manual (RM), non-routine manual (NRM) (employed or unemployed), and not in the labor force (NLF). I then compute monthly labor market flows from 1986 to 2012 for each U.S. state between these employment states, thus, transition rates between employment states over time. In order to quantify how much the outflow rate from routine employment (ERM) out of the labor force contributed to the reduction in overall routine employment (ERM), I construct counterfactual routine employment paths as Cortes et al. (2020) in the following way:

1. Fix the outflow rate from ERM to NLF at 1986 level:

$$\hat{\mu}_t(NLF, ERM) = \mu_{1986}(NLF, ERM) \ \forall t.$$

2. Leave other transition rates as in data, only rescale such that transition rates add up to 1 ($\sum_j \mu_t(NLF, i) = 1$).

3. Construct counterfactual employment shares over time:

$$S_{t+1} = \hat{\mu}_t S_t.$$

4. Compare the counterfactual decline in ERM to the realized one:

$$F(\text{ERM} \to \text{NLF}) \equiv 1 - \frac{\Delta \hat{ERM}_{cf}}{\Delta ERM}.$$

$F(\text{ERM} \to \text{NLF})$ measures how much decline in ERM would have been avoided if the transition rate from ERM to NLF stayed at its 1986 level $\mu_{1986}(NLF, ERM)$.

I then regress $F(\text{ERM} \to \text{NLF})$ on routine manual unionization and a set of controls including the

1980 industry composition and demographics. In particular, I run the following model across US states s:

$$F(\text{ERM} \rightarrow \text{NLF}) = \beta_0 + \beta_1 U_s + \gamma X_s + u_s.$$

Table A.1 shows the results for three regressions using as independent variable the percentile of unionization for each state. The first column uses a dummy variable that measures whether a states is above or below the median of unionization, columns 2 and 3 use categorial variables that measure the quartile and quintile of a state's unionization.

Table A.1: The table shows the results of regressing the contribution of outflow from routine manual employment into non-employment for routine manual employment decline on different measures of unionization.

	Dependent variable: $F(ERM \rightarrow NLF)$		
	Union Coverage Q2	Union Coverage Q4	Union Coverage Q5
	(1)	(2)	(3)
	-0.064^{*}	-0.038^{**}	-0.031^{***}
	(0.038)	(0.018)	(0.012)
Observations	51	51	51
R^2	0.103	0.157	0.169

The results shows that the statistical contributions of outflow rates to the overall decline in routine employment is negatively correlated with routine manual unionization. As expected, the coefficient falls in magnitude from the left to the right when going to smaller percentiles while becoming better identified. This is because the change in unionization between two quintiles in column 3 is smaller than between the bottom and top half of unionization in column 1. Moreover, the estimates are economically meaningful. Going from the 1st (lowest) to the 5th (highest) quintile of unionization is associated with a 15pp increase in the share of routine manual employment decline accounted for by outflow from routine manual employment out of the labor force.

A.1.3. Additional Material for Empirical Analysis

A.1.3.1 Additional Robustness: Effect on RM Employment Decline

Figure A.2: The graphs show the effect of going from the 25th to the 75th percentile of unionization on the RM employment share over time.

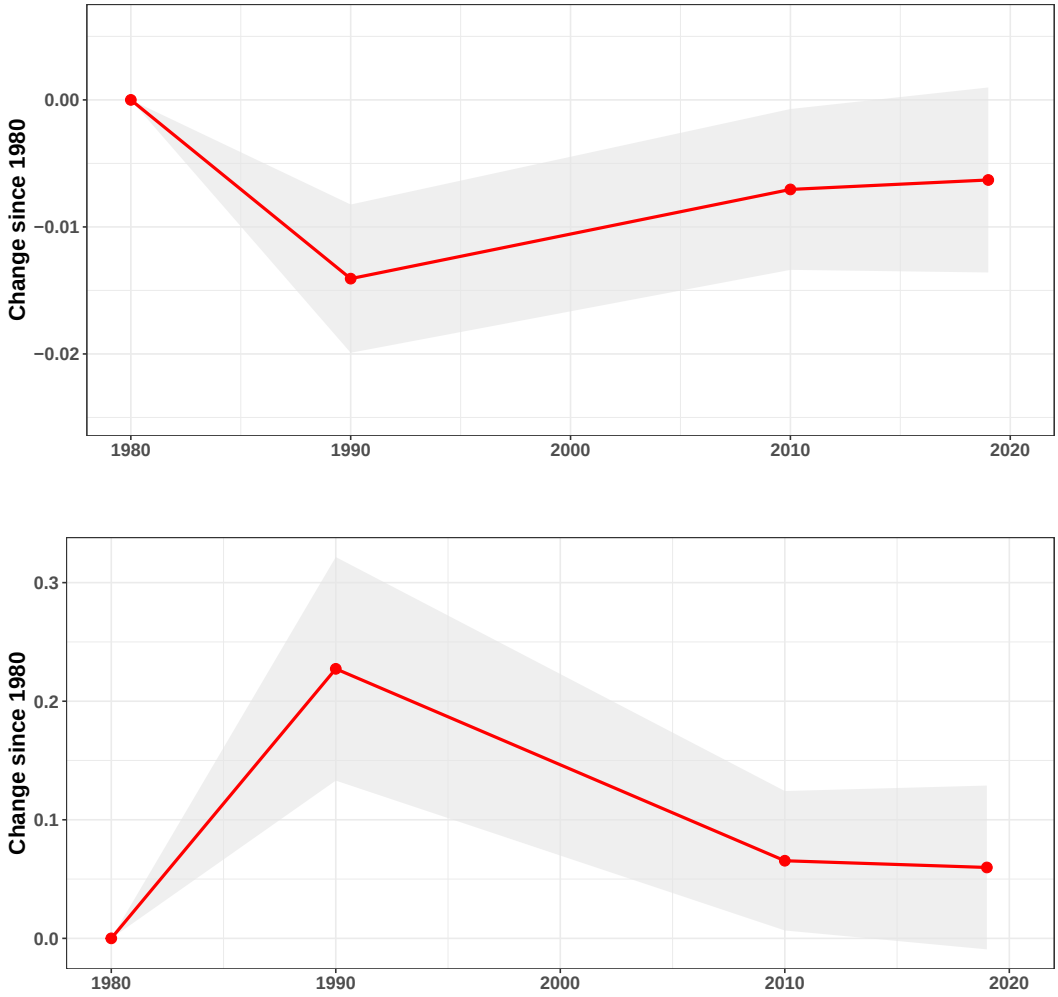


Figure A.2 repeats the main exercise of regressing the change in the routine manual employment share on routine manual unionization and the set of controls, and it plots the effect of going from a MSA at the 25th percentile to a MSA at the 75th percentile of unionization. As expected, the magnitude of the effect falls relative to the main result that compares the 10th to the 90th

percentile of unionization. However, the union effect remains significant and large. The routine manual employment share falls significantly more in high-unionized MSAs between 1980 and 1990, after which employment decline in low-unionized MSAs starts to catch up. The union effect is large, reaching almost 25% of the mean routine manual employment decline across MSAs between 1980 and 1990.

Figure A.3: The graph shows the effect of going from the 10th to the 90th percentile of unionization in 1986 at the MSA level on the RM employment share over time. Unionization is measured at the MSA level in 1986.

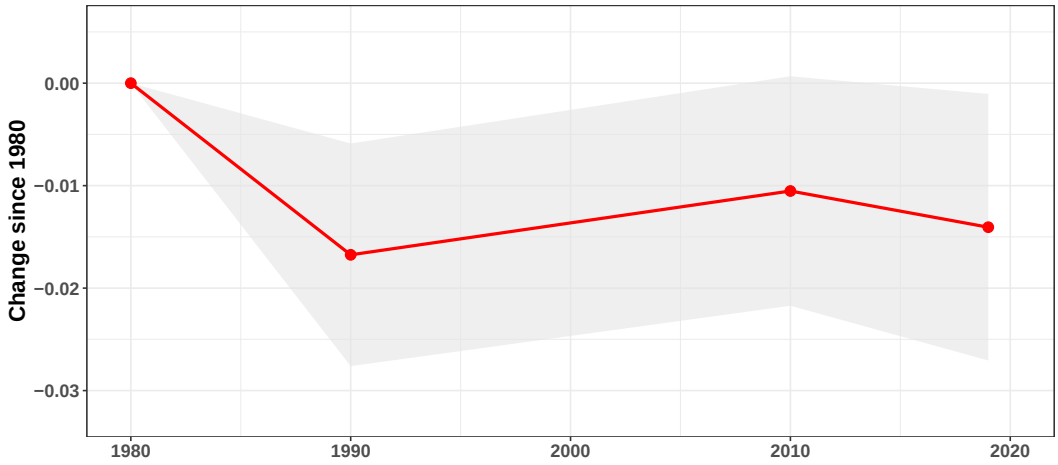
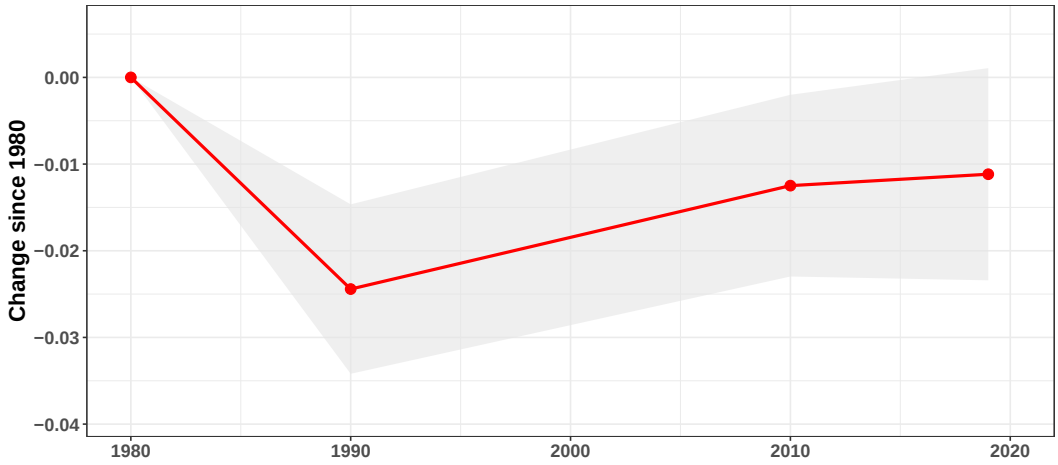


Figure A.4: The graph shows the effect of going from the 10th to the 90th percentile of unionization on the RM employment share over time. The set of controls additionally includes the change in the age composition at the MSA level as a proxy for migration.



A.1.3.2 Main Results: Union Effect on Age Composition

Table A.2: Effect of unionization on the change in the age distribution of routine manual workers between 1980 and 1990.

	Dependent variable: Change in CDF across Ages				
	Age 20	Age 30	Age 40	Age 50	Age 60
	(1)	(2)	(3)	(4)	(5)
Unionization	−0.043***	−0.126***	−0.114***	−0.062***	−0.023**
	(0.012)	(0.027)	(0.026)	(0.020)	(0.011)
Change RM 1980-1990	0.165***	0.687***	0.405***	0.111	0.088*
	(0.059)	(0.156)	(0.141)	(0.091)	(0.052)
Mean dependent	-0.072	-0.099	-0.017	0.026	0.012
Observations	200	200	200	200	200
R^2	0.314	0.474	0.367	0.261	0.260
Adjusted R^2	0.262	0.434	0.319	0.205	0.204
Note:	*p<0.1; **p<0.05; ***p<0.01				

Table A.3: Effect of unionization on the change in the age distribution of routine manual workers between 1980 and 2010.

	Dependent variable: Change in CDF Gap across Ages				
	Age 20	Age 30	Age 40	Age 50	Age 60
	(1)	(2)	(3)	(4)	(5)
Unionization	-0.044^{***}	-0.106^{***}	-0.119^{***}	-0.142^{***}	-0.046^{***}
	(0.014)	(0.035)	(0.028)	(0.030)	(0.017)
Change RM 1980-2010	0.051	0.580^{***}	0.551^{***}	0.272	0.055
	(0.065)	(0.189)	(0.155)	(0.175)	(0.094)
Mean dependent	-0.1	-0.2	-0.18	-0.083	-0.0063
Observations	200	200	200	200	200
R^2	0.260	0.334	0.331	0.364	0.256
Adjusted R^2	0.204	0.284	0.280	0.316	0.199
Note:	*p<0.1; **p<0.05; ***p<0.01				

Table A.4: Effect of unionization on the change in the age distribution of routine manual workers between 1980 and 2019.

	Dependent variable: Change in CDF Gap across Ages				
	Age 20	Age 30	Age 40	Age 50	Age 60
	(1)	(2)	(3)	(4)	(5)
Unionization	-0.037^{**}	-0.026	-0.067^{**}	-0.087^{***}	-0.084^{***}
	(0.017)	(0.029)	(0.033)	(0.031)	(0.024)
Change RM 1980-2019	0.115	0.226^{*}	0.349^{**}	0.302^{**}	0.237^{***}
	(0.078)	(0.135)	(0.153)	(0.143)	(0.086)
Mean dependent	-0.094	-0.17	-0.16	-0.11	-0.051
Observations	147	147	147	147	147
R^2	0.253	0.327	0.379	0.287	0.291
Adjusted R^2	0.174	0.256	0.313	0.211	0.216
Note:	*p<0.1; **p<0.05; ***p<0.01				

A.1.3.3 Additional Robustness: Union Effect on Age Composition

Figure A.5: The graphs show the effect of going from the 25th to the 75th percentile of unionization on the change in the routine manual age composition over time.

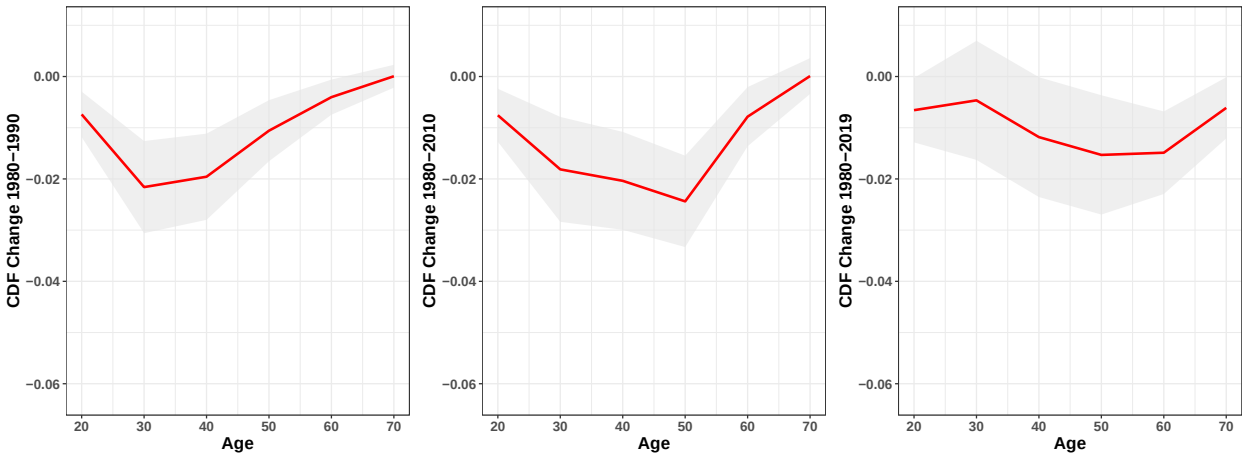


Figure A.5 repeats the main exercise of regressing the change in the age composition (cdf) of the routine manual workforce in a MSA on routine manual unionization and the set of controls, and it plots the coefficient scaled by the difference in unionization between a MSA at the 25th percentile and a MSA at the 75th percentile of unionization. As expected, the magnitude of the effect falls relative to the main result that compares the 10th to the 90th percentile of unionization. However, the union effect remains significant and large.

Figure A.6: The graphs show the effect of going from the 10th to the 90th percentile of unionization in 1986 at the MSA level on the change in the routine manual age composition over time. Unionization is measured at the MSA level in 1986.

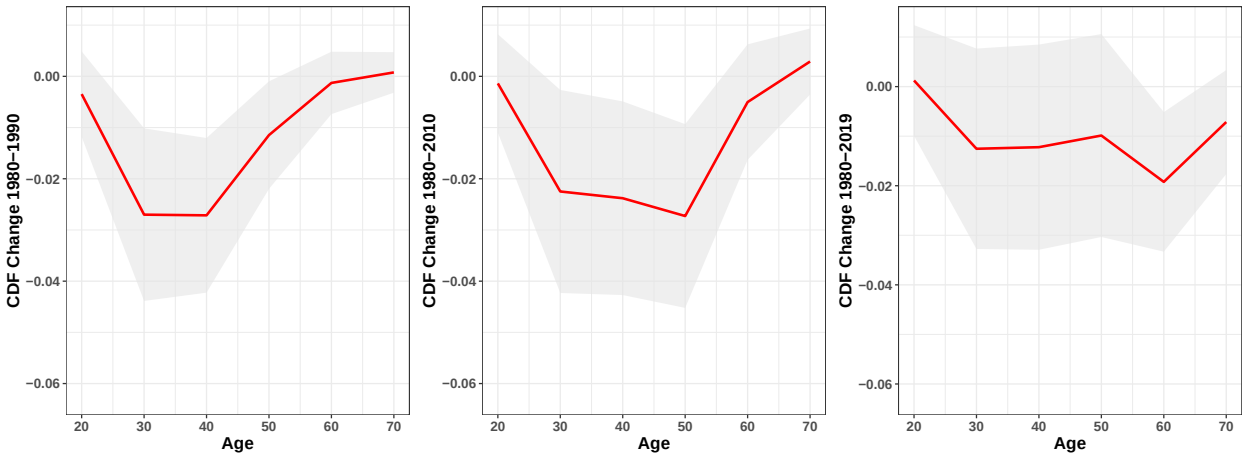
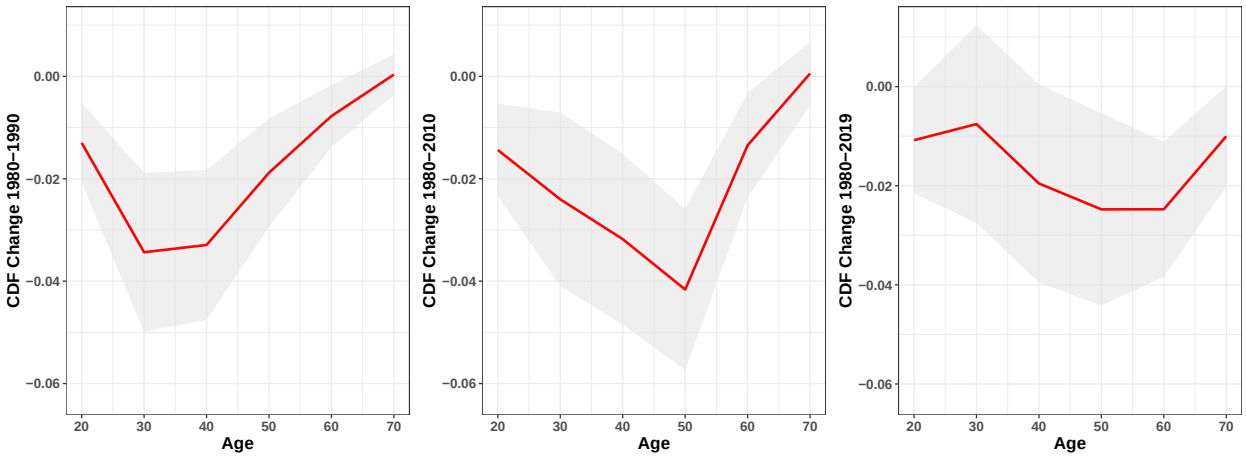


Figure A.6 repeats the main exercise of regressing the change in the age composition (cdf) of the routine manual workforce in a MSA on unionization and the set of controls. Instead of using the average share of unionization among routine manual workers between 1995 and 2005, I here use unionization in 1986. However, in order to have sufficient coverage, unionization is measured among all employed workers at the overall MSA level. Again, the magnitude of the effect falls relative to the main result as expected, but the union effect remains.

Figure A.7: The graphs show the effect of going from the 10th to the 90th percentile of unionization on the change in the routine manual age composition over time. The set of controls additionally includes the change in the age composition at the MSA level as a proxy for migration.



One concern may be that young workers in highly unionzed MSAs respond to bad employment prospects in routine manual occupations by migrating to less unionized MSAs. Figure A.7 shows results when additionally controlling for changes in the age composition of all employed workers at the MSA level as a proxy for in and out migration for the corresponding time periods. The union effect remains basically unchanged, indicating that the main results are not driven by migration.

A.1.3.4 Additional Robustness: Union Effect on Age Composition

Table A.5: Robustness: Effect of unionization on the change in the age distribution of routine manual workers between 1980 and different stages of the transition (1990, 2010, 2019). Regression uses routine manual employment share in 1980 for each MSA as weights.

	Dependent variable: Change in CDF across Ages				
	Age 20	Age 30	Age 40	Age 50	Age 60
	(1)	(2)	(3)	(4)	(5)
CDF Change 1980-1990	-0.042^{***}	-0.121^{***}	-0.110^{***}	-0.060^{***}	-0.023^{*}
	(0.012)	(0.028)	(0.027)	(0.021)	(0.012)
CDF Change 1980-2010	-0.043^{***}	-0.104^{***}	-0.113^{***}	-0.138^{***}	-0.042^{**}
	(0.014)	(0.034)	(0.029)	(0.031)	(0.018)
CDF Change 1980-2019	-0.036^{**}	-0.022	-0.063^{*}	-0.080^{**}	-0.079^{***}
	(0.017)	(0.029)	(0.034)	(0.032)	(0.025)

Table A.5 reports the results when estimating the effect of unionization on the change in the age distribution of routine manual workers between 1980 and different stages of the transition using the routine manual employment share in 1980 for each MSA as regression weights. The results are robust to reweighting.

A.2. Model Appendix

A.2.1. Workers hold fixed equity shares

Figure A.8 displays the welfare cost to routine workers along the transition in the low-unionized labor market if workers hold fixed and equal equity shares in the firms. While the overall distribution and evolution of welfare costs to routine workers is similar to the baseline economy in which workers do not own equity, the level of welfare costs is lower. If workers hold equity, they benefit from automation due increased profits, which partially offsets the earnings losses they incur.

Figure A.8: The graph shows the welfare cost of automation for routine workers along the transition in the low-unionized labor market when workers hold fixed and equal equity shares.

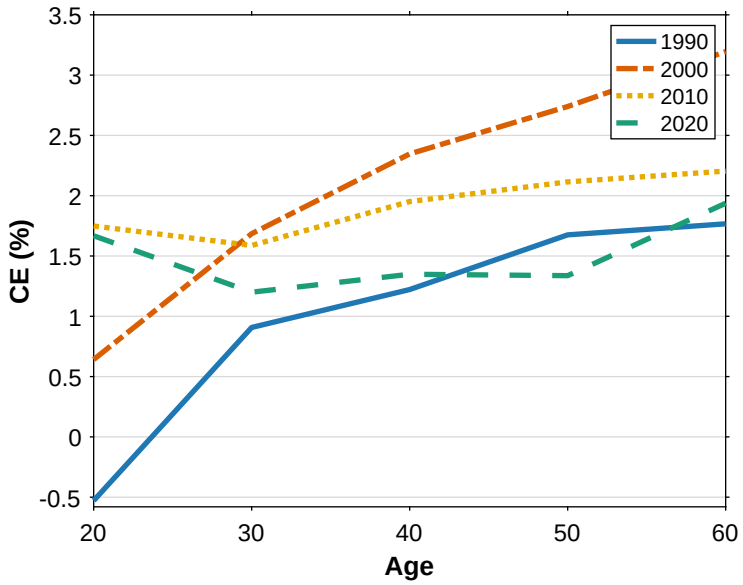
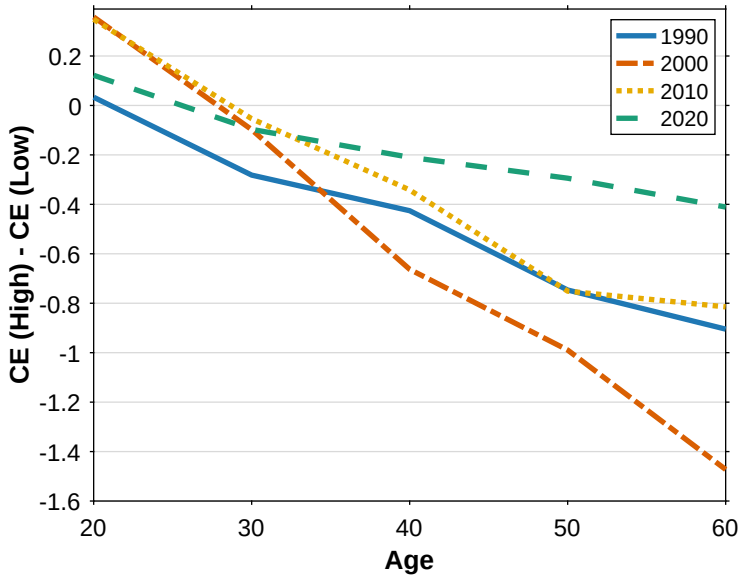


Figure A.9 displays the union effect on welfare cost to routine workers along the transition for the fixed equity case. Again, the shape is similar to the baseline economy, unions shift the cost from older, incumbent cohorts to young workers. However, the union effect falls as wage income becomes a smaller component of workers' income, which in turn reduces the importance of lower layoff risk and limited earnings losses due to high unionization.

Figure A.9: The graph shows the union effect on the welfare cost of automation along the transition.



APPENDIX B

Appendix for Chapter 2

B.1. Data Appendix

Our data to construct cohort life-expectancy comes from three different sources. First, we use historical mortality rates for the US that were originally collected and imputed by Haines (1994), which were updated by Hacker (2010). This data covers the time period 1790 to 1899 and comes at a decennial frequency for the age groups $\{0, 1-4, 5-9, \ldots, 80\}$, where the authors report mortality rates of 1 for age 80 onward. We carry out two transformations to this data: (i) we hold mortality rates constant within age group and compute age specific mortality rates in each age group i, m^i_j such that m^i is the respective geometric average within this age group, i.e., $m^i_j = 1 - (1 - m^i)^{1/5}$; (ii) to obtain estimates of mortality rats above (and including) age 80, we estimate per period a Gompertz–Makeham mortality model on the (constructed) age specific mortality rates and set the mortality rates for all ages 80 and older to the predicted values. Second, for the years 1900 to 1932 we use data from the Human Life-Table Database, which have age specific mortality data for ages $0, 1, \ldots, 105$. We append those to the historical mortality rates and smooth the resulting mortality rates over years 1790 to 1932 and ages 5 and older with a kernel density smoother and a bandwidth parameter of 5. Third, for the years 1933 to 2021 we use data from the Human Mortality Database with age and time specific mortality rates up to age 110. To predict future mortality rates (needed for the computation of cohort life expectancy) we estimate future trends in mortality by a Lee-Carter method—assuming a deterministic trend of the single index—which we apply to the postwar data (years 1950-2021). Finally, over the entire period (including the predicted mortality rates), we filter the data by applying a Hodrick-Prescott filter with a bandwidth parameter of 100.

The bootstrapped confidence intervals shown in Panels (c) to (e) in Figure 2.1 are computed by bootstrapping along the time dimension of the cross-sectional mortality rates, whereby we ignore the uncertainty of the predicted mortality estimates from the Lee-Carter method. We implement the bootstrap procedure as a block bootstrap and set the width of the blocks according to the

standard rule of thumb $bw = T^{1/3}$. We bootstrap on the entire data sequence, starting in year 1790 and ending in 2152[40], thus $T = 363$ and $bw = 7.13$, of which we take the ceil.

B.2. Model Appendix

B.2.1. Analytical Solution to the Household Problem

Households derive utility from consumption in young age c_t^y,, and old age c_{t+1}^o,, they survive from the first to the second period of their life with probability ψ which depends on their investment i_t into health goods when young. Expected lifetime utility is given by

$$(1-\beta)u(c_t^y) + \beta\psi\left(i_t\right)u(c_{t+1}^o).$$

The maximization of the utility function is subject to the constraints:

$$c_t^y + i_{ft} + p_t i_{ht} + s_t = x_t \equiv w_{ft} l_{ft} + w_{ht} l_{ht} + T_t \tag{B.1a}$$

$$i_t = f(i_{ft}, i_{ht}) \tag{B.1b}$$

$$1 = g(l_{ft}, l_{ht}) \tag{B.1c}$$

$$c_{t+1}^o = R s_t. \tag{B.1d}$$

Since optimal saving is always strictly positive, potential borrowing constraints never bind and the period budget constraints can be consolidated to the lifetime budget constraint

$$c_t^y + i_{ft} + p_t i_{ht} + \frac{c_{t+1}^o}{R} = w_{ft} l_{ft} + w_{ht} l_{ht} + T_t \equiv x_t. \tag{B.2}$$

B.2.1.1 Division of Health Investment

We first derive the optimal split between final goods and health goods for a given amount of health expenditures e_t.

[40] Data from the Human Mortality Database range to year 2021 and to age 110. We add 20 additional years. The horizon is thus $2021 + (110 + 1) + 20 = 2152$.

Corner Solution with $i_{ht} = 0$, $i_{ft} = e_t$: In the corner solution the first-order conditions do not hold with equality. From the budget constraint we get $i_{ht} = 0$, $i_{ft} = e_t$. Then health investment i_t is given by

$$i_t = f(i_{ht}, i_{ft}) = i_{ht} + (\nu + i_{ft})^{\zeta} = (\nu + e_t)^{\zeta}$$

Note, for $e_t = 0$ the corner solution corresponds to stage 1 (no health expenditures), for $e_t > 0$ it corresponds to stage 2 (positive health expenditures that are fully invested into final goods).

Interior Solution: In the interior solution the first-order conditions hold with equality, they are given by

$$\begin{aligned} \zeta (\nu + i_{ft})^{\zeta - 1} &= \lambda \\ 1 &= \lambda p_t \end{aligned}$$

Combining yields

$$\begin{aligned} \zeta (\nu + i_{ft})^{\zeta - 1} &= \frac{1}{p_t} \\ \nu + i_{ft} &= (\zeta p_t)^{\frac{1}{1-\zeta}} \\ i_{ft} &= (\zeta p_t)^{\frac{1}{1-\zeta}} - \nu \end{aligned}$$

Define $\tilde{\lambda}_t \equiv (\zeta p_t)^{\frac{1}{1-\zeta}}$. Then

$$i_{ft} = \tilde{\lambda}_t - \nu$$

The quasi-linear specification means that in the interior solution there are no wealth effects for the final health good i_{ft}. Thus, $i_{ft} = \tilde{\lambda}_t - \nu$ is independent of e_t. Using the budget constraint, we get

$$i_{ht} = \frac{e_t - (\tilde{\lambda}_t - \nu)}{p_t}$$

For $\zeta \to 0$ we get $\tilde{\lambda}_t \to 0$ and, thus, taking into account the non-negativity constraints $(i_{ft}, i_{ht}) \geq 0$

$$\begin{aligned} i_{ft} &\to 0 \\ i_{ht} &\to \frac{e_t}{p_t} \\ i_t &\to \frac{e_t}{p_t} + \nu^{\zeta} \end{aligned}$$

Thus, ζ needs to be sufficiently large for stage 2 to exist, otherwise the marginal benefit from final goods investment is too small.

Existence of Stage 2: We want to ensure that initial health investment will be allocated towards i_{ft}, that is, that the corner solution with $i_{ht} = 0$ and $i_{ft} = e_t$ characterized above exists. We get the undesired corner solution with $i_{ht} > 0$ and $i_{ft} = 0$ (stage 2 is skipped with initial health investment directly being allocated towards i_{ht}) if the marginal cost of investing into i_{ft} exceeds the marginal benefit, evaluated at $i_{ft} = 0$:

$$\begin{aligned} \zeta \left(\nu + i_{ft}\right)^{\zeta - 1} &\leq \tfrac{1}{p_t} \\ \zeta \nu^{\zeta - 1} &\leq \tfrac{1}{p_t} \\ \left(\zeta p_t\right)^{\frac{1}{1-\zeta}} &\leq \nu \\ \tilde{\lambda}_t &\leq \nu \end{aligned}$$

Then stage 2 exists as long as the optimal interior solution computed above yields $i_{ft} = \tilde{\lambda}_t - \nu > 0$ which requires the non-homotheticity factor ν to be sufficiently small relative to health sector price p_t and ζ.

Characterizing the Stages: For a given level of health expenditures e_t we can fully characterize the stages now. Assuming $\tilde{\lambda}_t > \nu$ for existence of stage 2, the stages are then characterized by

$$\text{Stage} = \begin{cases} 1, & \text{if} \quad e_t = 0 \\ 2, & \text{if} \quad e_t \in (0, \tilde{\lambda}_t - \nu] \\ 3, & \text{if} \quad e_t > \tilde{\lambda}_t - \nu \end{cases}$$

B.2.1.2 Level of Health Expenditures

Given the optimal division of health investment we now optimize over the allocation of cash x_t into consumption in young age, savings s_t, and health expenditures e_t. The household solves

$$\max_{0\leq c_t^y, e_t \leq x_t} (1-\beta)u(c_{t,y}) + \beta\psi\left(i_t(p_t, e_t)\right) u(R\left[x_t - e_t - c_{t,y}\right]$$

We can rewrite the household problem in terms of expenditure shares. Define the share of consumption when young and the share of health expenditures in old age spending, respectively, as

$$\begin{aligned}
\vartheta_{t,c} &= \frac{c_t^y}{x_t} \in [0,1], \\
\vartheta_{t,e} &= \frac{e_t}{e_t + s_t} = \frac{e_t}{(1-\vartheta_{t,c})x_t} = \frac{p_t i_{ht} + i_{ft}}{(1-\vartheta_{t,c})x_t} \in [0,1],
\end{aligned}$$

Then the maximization problem can be rewritten in terms of those two spending shares

$$\max_{0\leq \vartheta_{t,c}, \vartheta_{t,e} \leq x_t} (1-\beta)u(\vartheta_{t,c}x_t) + \beta\psi\left(i_t(p_t, (1-\vartheta_{t,c})\vartheta_{t,e}x_t)\right) u(Rx_t\left(1-\vartheta_{t,c}\right)\left(1-\vartheta_{t,e}\right))$$

The first-order conditions are given by

$$\begin{aligned}
[\vartheta_{t,c}] \quad 0 &= (1-\beta)u_y' x_t - \beta\psi u_o' R x_t (1-\vartheta_{t,e}) + \beta\psi' \frac{\partial i_t}{\partial \vartheta_{t,c}} u_o \\
[\vartheta_{t,e}] \quad 0 &= \psi' \frac{\partial i_t}{\partial \vartheta_{t,e}} u_o - \psi u_o' R x_t (1-\vartheta_{t,c})
\end{aligned}$$

Note, $\frac{\partial i_t}{\partial \vartheta_{t,e}}$ depends on the optimal division of health investment and, thus, the level of $\vartheta_{t,e}$. Rearranging yields

$$\begin{aligned}
(1-\beta)u_y' &= \beta\psi u_o' R(1-\vartheta_{t,e}) - \beta\psi' \frac{\partial i_t}{\partial \vartheta_{t,c}} \frac{u_o}{x} \\
\frac{\psi}{\psi'} x_t R(1-\vartheta_{t,c}) &= \frac{u_o}{u_o'} \frac{\partial i_t}{\partial \vartheta_{t,e}}
\end{aligned}$$

The second optimality condition describes the tradeoff between health investment and consumption

in old age.

B.2.1.3 Corner Solution of Optimal Health Investment

In stage 1 all resources are spent on consumption without any health investment, $\vartheta_{t,e} = 0$. Moreover, $\frac{\partial i_t}{\partial \vartheta_{t,e}}$ depends on the optimal division of health investment, that is, it depends on how the first marginal unit of health investment e_t is allocated between i_{ht} and i_{ft}. We assume $\tilde{\lambda}_t \geq \nu$ which we later verify in equilibrium. As shown above, this ensures that phase 2 exists and initial health investment is allocated towards the final good sector i_{ft}. Then health investment and its derivatives are given by

$$\begin{aligned} i_0 &= \nu^{\zeta} \\ \frac{\partial i_t}{\partial \vartheta_{t,e}}|_{\vartheta_{t,e}=0} &= \frac{\partial}{\partial \vartheta_{t,e}} \left(\nu + (1-\vartheta_{t,c})\vartheta_{t,e}x\right)^{\zeta} |_{\vartheta_{t,e}=0} = \zeta(1-\vartheta_{t,c})x_t\nu^{\zeta-1} \\ \frac{\partial i_t}{\partial \vartheta_{t,c}} &= -\zeta\nu^{\zeta-1}\vartheta_{t,e}x|_{\vartheta_{t,e}=0} = 0 \end{aligned}$$

Consumption in young and old age are given by

$$\begin{aligned} c_t^y &= \vartheta_{t,c}x_t \\ c_{t+1}^o &= R(1-\vartheta_{t,c})x_t \end{aligned}$$

The intertemporal Euler equation can be solved analytically for the optimal share of consumption

in young age during phase 1. It is given by

$$
\begin{aligned}
(1-\beta)u_y' &= \beta\psi_0 u_o' R(1-\vartheta_{t,e}) - \beta\psi_0' \frac{\partial i_t}{\partial \vartheta_{t,c}} \frac{u_o}{x} \\
(1-\beta)u_y' &= \beta\psi_0 u_o' R \\
(\vartheta_{t,c}x)^{-\sigma} &= \frac{\beta}{1-\beta}\psi_0 \left(R(1-\vartheta_{t,c})x\right)^{-\sigma} R \\
\left(\frac{1-\vartheta_{t,c}}{\vartheta_{t,c}}\right)^{\sigma} &= \frac{\beta}{1-\beta}\psi_0 R^{1-\sigma} \\
\vartheta_{t,c} &= \left[1 + \left(\frac{\beta}{1-\beta}\psi_0 R^{1-\sigma}\right)^{\frac{1}{\sigma}}\right]^{-1} \equiv \vartheta_{\text{phase 1},c}
\end{aligned}
$$

where $\psi_0 = 1 - \left(1+\nu^{\zeta}\right)^{-\xi}$ is the base survival probability during phase 1. We can further use the 2. FOC characterizing the tradeoff between health investment and consumption in old age to find the threshold cash level at which phase 2 begins. During stage 1 without any health expenditure, the 2. FOC does not hold with equality, evaluated at $i_t(p_t, e_t = 0)$, such that

$$
\begin{aligned}
\frac{\psi}{\psi'} x_t R(1-\vartheta_{t,c}) &\geq \frac{u_o}{u_o'} \frac{\partial i_t}{\partial \vartheta_{t,e}} \\
\frac{1}{\xi}(1+i_t)\left((1+i_t)^{\xi} - 1\right) R x_t (1-\vartheta_{t,c}) &\geq \left(\frac{1}{1-\sigma} + b(c_{t+1}^o)^{\sigma-1}\right) c_{t+1}^o \frac{\partial i_t}{\partial \vartheta_{t,e}}
\end{aligned}
$$

Plugging $i_t(p_t, e_t = 0)$ and $\frac{\partial i_t}{\partial \vartheta_{t,e}}$ in yields

$$
\begin{aligned}
\frac{1}{\xi}\left(1+\nu^{\zeta}\right)\left(\left(1+\nu^{\zeta}\right)^{\xi} - 1\right) R x_t (1-\vartheta_{t,c}) &\geq \left(\frac{1}{1-\sigma} + b(c_{t+1}^o)^{\sigma-1}\right) c_{t+1}^o \zeta (1-\vartheta_{t,c}) x_t \nu^{\zeta-1} \\
\frac{1}{\xi}\left(1+\nu^{\zeta}\right)\left(\left(1+\nu^{\zeta}\right)^{\xi} - 1\right) R &\geq \left(\frac{1}{1-\sigma} + b(c_{t+1}^o)^{\sigma-1}\right) c_{t+1}^o \zeta \nu^{\zeta-1}
\end{aligned}
$$

Rewriting consumption in old age in terms of overall cash in the first stage ($\vartheta_{t,e} = 0$) yields

$$
c_{t+1}^o = R s_t = R(1-\vartheta_{t,e})(1-\vartheta_{\text{phase 1},c}) x_t = R(1-\vartheta_{\text{phase 1},c}) x_t = R\tilde{x}_t
$$

where

$$\tilde{x}_t = (1 - \vartheta_{\text{phase 1},c})x_t = \left(1 - \left[1 + \left(\frac{\beta}{1-\beta}\psi_0 R^{1-\sigma}\right)^{\frac{1}{\sigma}}\right]^{-1}\right) x_t$$

is total cash spent on health investment and old age consumption during phase 1. Plugging back in then gives

$$\frac{1}{\xi}\left(1+\nu^{\zeta}\right)\left(\left(1+\nu^{\zeta}\right)^{\xi}-1\right) R \geq \left(\frac{1}{1-\sigma} + b\left(R\tilde{x}_t\right)^{\sigma-1}\right) R\tilde{x}_t \zeta \nu^{\zeta-1} \tag{B.3}$$

$$\frac{1}{\xi}\left(1+\nu^{\zeta}\right)\left(\left(1+\nu^{\zeta}\right)^{\xi}-1\right) \geq \left(\frac{1}{1-\sigma} + b\left(R\tilde{x}_t\right)^{\sigma-1}\right) \tilde{x}_t \zeta \nu^{\zeta-1} \tag{B.4}$$

This equation characterizes the region of cash $\tilde{x}_t$ for which health expenditures are zero, $e_t = 0$, and the economy is in phase 1. For convenience define

$$A_1 \equiv \frac{1}{\xi}\frac{\left(1+\nu^{\zeta}\right)\left(\left(1+\nu^{\zeta}\right)^{\xi}-1\right)}{\zeta\nu^{\zeta-1}} \begin{cases} = 0, & \text{if } \nu = 0 \\ > 0, & \text{if } \nu > 0 \end{cases}$$

Thus, $A_1 \geq 0$ and in our case we have $A_1 > 0$ because $\nu > 0$ is necessary for the existence of stage 2. Plugging A_1 back into the condition for the corner solution yields

$$A_1 \geq \left(\frac{1}{1-\sigma} + b\left(R\tilde{x}_t\right)^{\sigma-1}\right)\tilde{x}_t$$

To make progress analytically we need to make assumptions on the parameters. Assume $\sigma = 2$, then

$$\begin{aligned} A_1 &\geq -\tilde{x}_t + bR\tilde{x}_t^2 \\ 0 &\geq \tilde{x}_t^2 - \frac{1}{bR}\tilde{x}_t - \frac{A_1}{bR} \end{aligned}$$

The two solutions to this quadratic equation are given by

$$\tilde{x}_{1/2} = \frac{1}{2bR}\left(1 \pm \sqrt{1+4bRA_1}\right)$$

$A_1 \geq 0$ means that $\sqrt{1+4bRA_1} \geq 1$. Thus, $\tilde{x}_2 \leq 0$, then $\tilde{x}_1$ is the only economically relevant solution. The first kick-off separating stage 1 and stage 2 is then given by

$$\begin{aligned}\tilde{x}_{\text{kickoff1}} &= \frac{1}{2bR}\left(1+\sqrt{1+4bRA_1}\right) \\ &= \frac{1}{bR}\left(\frac{1}{2}\left(1+\sqrt{1+4bRA_1}\right)\right) \\ &\equiv \frac{1}{bR}\Delta(b,R,A_1) \end{aligned} \tag{B.5}$$

Notice that $\nu > 0$ ensures $A_1 > 0$ which in turn ensures $\Delta(b,R,A_1) > 1$. Further, $\tilde{x}_{\text{lowerbound}} = \frac{1}{b}$ is the lower bound on cash such that there is no suicide. Then the interval

$$\tilde{x} \in [\tilde{x}_{\text{lowerbound}}, \tilde{x}_{\text{kickoff1}}] \tag{B.6}$$

is non-empty for $\nu > 0$ and characterizes the cash region for stage 1 without suicide. Finally, we can map this characterization of phase 1 in terms of cash spent in old age $\tilde{x}$ back into overall cash x.

$$x \in [x_{\text{lowerbound}}, x_{\text{kickoff1}}] = \left[\frac{\tilde{x}_{\text{lowerbound}}}{(1-\vartheta_{\text{phase 1},c})}, \frac{\tilde{x}_{\text{kickoff1}}}{(1-\vartheta_{\text{phase 1},c})}\right] \tag{B.7}$$

B.2.1.4 Interior Solution of Optimal Health Investment

In the interior solution we cannot solve for the shares analytically. Instead we have a system of two equations given by

$$\begin{aligned}(1-\beta)u_y' &= \beta\psi u_o' R(1-\vartheta_{t,e}) - \beta\psi' \frac{\partial i_t}{\partial \vartheta_{t,c}} \frac{u_o}{x} \\ \frac{\psi}{\psi'} x_t R(1-\vartheta_{t,c}) &= \frac{u_o}{u_o'} \frac{\partial i_t}{\partial \vartheta_{t,e}}\end{aligned}$$

Where health investment i_t and its derivatives ($\frac{\partial i_t}{\partial \vartheta_{t,c}}, \frac{\partial i_t}{\partial \vartheta_{t,e}}$) both depend on whether we are in stage 2 or 3 (both stages are interior solutions with respect to the optimal $\vartheta_{t,e}$). We know that we move from stage 2 to stage 3 when e_t is sufficiently large, formally when $e_t = (1-\vartheta_{t,c})\vartheta_{t,e}(x_t)x_t > \tilde{\lambda}_t - \nu$. For a given x_t we can solve the FOCs and find $\vec{\vartheta}(x_t) = (\vartheta_{t,c}(x_t), \vartheta_{t,e}(x_t))$ with the following steps

1. Guess $\hat{\vec{\vartheta}}$

2. Determine the current stage through $\hat{e} = (1-\hat{\vartheta}_{t,c})\hat{\vartheta}_{t,e}x$. Specifically,

$$\text{Stage} = \begin{cases} 2, & \text{if} \quad \hat{e} \in (0, \tilde{\lambda}_t - \nu] \\ 3, & \text{if} \quad \hat{e} > \tilde{\lambda}_t - \nu \end{cases}$$

3. Compute $\hat{c}^y_t$, $\hat{c}^o_{t+1}$, $\hat{i}$, $\frac{\partial \hat{i}}{\partial \hat{\vartheta}_{t,e}}$ and $\frac{\partial \hat{i}}{\partial \hat{\vartheta}_{t,c}}$ based on the stage and check if the FOCs hold

$$\hat{i}_f = \begin{cases} \hat{e} = (1-\hat{\vartheta}_{t,c})\hat{\vartheta}_{t,e}x, & \text{if Stage} = 2 \\ \tilde{\lambda} - \nu, & \text{if Stage} = 3 \end{cases}$$

$$\hat{i}_h = \begin{cases} 0, & \text{if Stage} = 2 \\ \frac{\hat{e}-\hat{i}_f}{p_h} = \frac{(1-\hat{\vartheta}_{t,c})\hat{\vartheta}_{t,e}x-\hat{i}_f}{p_h}, & \text{if Stage} = 3 \end{cases}$$

$$\frac{\partial \hat{i}}{\partial \hat{\vartheta}_{t,e}} = \begin{cases} \zeta(\nu+\hat{i}_f)^{\zeta-1}(1-\hat{\vartheta}_{t,c})x, & \text{if Stage} = 2 \\ \eta_h \frac{(1-\hat{\vartheta}_{t,c})x}{p_h}, & \text{if Stage} = 3 \end{cases}$$

$$\frac{\partial \hat{i}}{\partial \hat{\vartheta}_{t,c}} = \begin{cases} -\zeta(\nu+\hat{i}_f)^{\zeta-1}\hat{\vartheta}_{t,e}x, & \text{if Stage} = 2 \\ -\eta_h \frac{\hat{\vartheta}_{t,e}x}{p_h}, & \text{if Stage} = 3 \end{cases}$$

4. Repeat until $\hat{\vec{\vartheta}}$ is found that solves the FOCs

B.2.1.5 Interior Solution on the BGP

We solve for the BGP using $x_t \to \infty$ and $\psi(i_t) = 1$. In the interior solution of the BGP we have

$$\frac{\partial i_t}{\partial \vartheta_{t,e}} = \frac{\partial}{\partial \vartheta_{t,e}} \frac{1}{p_t} \left(\vartheta_{t,e}(1-\vartheta_{t,c})x_t + \nu_t + \frac{1-\zeta}{\zeta}\tilde{\lambda}_t \right) = \frac{(1-\vartheta_{t,c})x_t}{p_t}$$

Then the 2. FOC becomes

$$\frac{1}{\xi}(1+i_t)\left((1+i_t)^{\xi}-1\right)x_t R(1-\vartheta_{t,c}) = \left(\frac{1}{1-\sigma} + bc_{t+1}^{\sigma-1}\right) c_{t+1} \frac{(1-\vartheta_{t,c})x_t}{p_t}$$

$p_t = p$ is constant, x_t converges to infinity and $i_t \to \eta_h i_{ht} \to \frac{e_t}{p_t} = \frac{(1-\vartheta_{t,c})\vartheta_e x}{p_t}$. Thus i_t and c_{t+1} are both constant shares of cash x_t on the BGP. For a BGP to exist we therefore require $\xi = \sigma - 1$. Plugging in

$$i = \frac{(1-\vartheta_c)\vartheta_e x}{p}$$
$$c = R(1-\vartheta_e)(1-\vartheta_c)x$$

Solving for the limit case where $x \to \infty$ we can find the health expenditure share on the BGP

$$\vartheta_e^* = \left(1 + \left[\frac{(pR)^{1-\sigma}}{b\xi}\right]^{\frac{1}{\sigma}}\right)^{-1}$$

Plugging ϑ_e^* into the Euler equation yields the BGP share of young consumption in cash

$$\vartheta_c^* = \left[1 + \left(\frac{\beta}{1-\beta}\psi\left[R(1-\vartheta_e^*)\right]^{1-\sigma}\right)^{\frac{1}{\sigma}}\right]^{-1}$$

B.2.2. Model without Young Consumption: Characterization of Health Investment

If we abstract from consumption at young age, $\beta = 1$, we can characterize the full household solution analytically along the transition, including the cash levels at which both kickoffs happen. We again first solve for the optimal split between final goods and health goods for a given amount of health expenditures e_t. Then we solve for the optimal amount of health expenditures e_t. Thus, we start by solving

$$\begin{aligned} i_t &= i_t(p_t, e_t) = \max_{i_{ft}, i_{ht}} f(i_{ft}, i_{ht}) \\ s.t.\ p_t i_{ht} + i_{ft} &= e_t \\ i_{ft}, i_{ht} &\geq 0 \\ f(i_{ft}, i_{ht}) &= i_{ht} + (\nu + i_{ft})^{\zeta} \end{aligned}$$

Corner Solution with $i_{ht} = 0$, $i_{ft} = e_t$: In the corner solution the first-order conditions do not hold with equality. From the budget constraint we get $i_{ht} = 0$, $i_{ft} = e_t$. Then health investment i_t is given by

$$i_t = f(i_{ht}, i_{ft}) = i_{ht} + (\nu + i_{ft})^{\zeta} = (\nu + e_t)^{\zeta}$$

Note, for $e_t = 0$ the corner solution corresponds to stage 1 (no health expenditures), for $e_t > 0$ it corresponds to stage 2 (positive health expenditures but fully invested into final goods).

Interior Solution: In the interior solution the first-order conditions hold with equality, they are given by

$$\begin{aligned} \zeta\,(\nu + i_{ft})^{\zeta - 1} &= \lambda \\ 1 &= \lambda p_t \end{aligned}$$

Combining yields

$$\begin{aligned} \zeta\left(\nu + i_{ft}\right)^{\zeta-1} &= \frac{1}{p_t} \\ \nu + i_{ft} &= \left(\zeta p_t\right)^{\frac{1}{1-\zeta}} \\ i_{ft} &= \left(\zeta p_t\right)^{\frac{1}{1-\zeta}} - \nu \end{aligned}$$

Define $\tilde{\lambda}_t \equiv \left(\zeta p_t\right)^{\frac{1}{1-\zeta}}$. Then

$$i_{ft} = \tilde{\lambda}_t - \nu$$

The quasi-linear specification means that in the interior solution there are no wealth effects for the concave good i_{ft}. Thus, $i_{ft} = \tilde{\lambda}_t - \nu$ is constant and independent of e_t. Using the budget constraint, we get

$$i_{ht} = \frac{e_t - (\tilde{\lambda}_t - \nu)}{p_t}$$

For $\zeta \to 0$ we get $\tilde{\lambda}_t \to 0$ and, thus, taking into account the non-negativity constraints $(i_{ft}, i_{ht}) \geq 0$

$$\begin{aligned} i_{ft} &\to 0 \\ i_{ht} &\to \frac{e_t}{p_t} \\ i_t &\to \frac{e_t}{p_t} + \nu^{\zeta} \end{aligned}$$

Thus, ζ needs to be sufficiently large for stage 2 to exist, otherwise the marginal benefit from final goods investment is too small.

Existence of Stage 2: We want to ensure that initial health investment will be allocated towards i_{ft}, that is, that the corner solution with $i_{ht} = 0$ and $i_{ft} = e_t$ characterized above exists. We get the undesired the corner solution with $i_{ht} > 0$ and $i_{ft} = 0$ (stage 2 is skipped with initial health investment directly being allocated towards i_{ht}) if the marginal cost of investing into i_{ft} exceeds

marginal benefit, evaluated at $i_{ft} = 0$:

$$\begin{aligned} \zeta\left(\nu + i_{ft}\right)^{\zeta-1} &\leq \tfrac{1}{p_t} \\ \zeta\left(\nu\right)^{\zeta-1} &\leq \tfrac{1}{p_t} \\ \left(\zeta p_t\right)^{\frac{1}{1-\zeta}} &\leq \nu \\ \tilde{\lambda}_t &\leq \nu \end{aligned}$$

Then stage 2 exists as long as the optimal interior solution computed above yields $i_{ft} = \tilde{\lambda}_t - \nu > 0$ which requires the non-homotheticity factor ν to be sufficiently small relative to health sector price p_t and ζ.

Characterizing the Stages: For a given level of health expenditures e_t we can fully characterize the stages now. Assuming $\tilde{\lambda}_t > \nu$ for existence of stage 2, the stages are then characterized by

$$\text{Stage} = \begin{cases} 1, & \text{if} \quad e_t = 0 \\ 2, & \text{if} \quad e_t \in (0, \tilde{\lambda}_t - \nu] \\ 3, & \text{if} \quad e_t > \tilde{\lambda}_t - \nu \end{cases}$$

B.2.2.1 Level of Health Expenditures

Given the optimal division of health investment we now optimize over the allocation of cash x_t into savings s_t and health expenditures e_t. That is, the household now solves

$$\max_{0 \leq e_t \leq x_t} \psi\left(i_t(p_t, e_t)\right) u(r_{t+1}\left[x_t - e_t\right]).$$

Define the share of health expenditures in total spending as

$$\vartheta_t = \frac{e_t}{x_t} = \frac{p_t i_{ht} + i_{ft}}{x_t} \in [0, 1],$$

Then we can rewrite the above problem in terms of ϑ_t

$$\max_{\vartheta_t \in [0,1]} \psi\left(i_t(p_t, \vartheta_t x_t)\right) u\left(R_{t+1} x_t \left(1 - \vartheta_t\right)\right).$$

The first-order condition is given by

$$0 = \psi' \frac{\partial i_t}{\partial \vartheta_t} u - \psi u' R_{t+1} x_t$$

Note, the first-order condition depends on $\frac{\partial i_t}{\partial \vartheta_t}$ which in turn depends on the optimal division of health investment and, thus, the level of ϑ_t. Rearranging yields

$$\frac{\psi}{\psi'} x_t R_{t+1} = \frac{u}{u'} \frac{\partial i_t}{\partial \vartheta_t} \tag{B.8}$$

B.2.2.2 Corner Solution of Optimal Health Investment

We get the corner solution with $\vartheta = 0$, corresponding to stage 1 without any health expenditure, when the FOC does not hold with equality evaluated at $i_t(p_t, e_t = 0)$ such that

$$\begin{aligned} \frac{\psi}{\psi'} x_t R_{t+1} &\geq \frac{u}{u'} \frac{\partial i_t}{\partial \vartheta_t} \\ \text{or} \quad \frac{1}{\xi}\left(1 + i_t\right)\left(\left(1 + i_t\right)^{\xi} - 1\right) x_t R_{t+1} &\geq \left(\frac{1}{1-\sigma} + b c_{t+1}^{\sigma-1}\right) c_{t+1} \frac{\partial i_t}{\partial \vartheta_t} \end{aligned}$$

Thus, we get the corner solution when the marginal benefit from saving one more unit is larger than the marginal benefit from the first unit of health investment. To further characterize this condition we can plug in $i_t(p_t, e_t = 0) = \mu_t \nu^{\zeta}$. However, $\frac{\partial i_t}{\partial \vartheta_t}$ depends on the optimal division of health investment, that is, it depends on how the first marginal unit of health investment e_t is allocated between i_{ht} and i_{ft}. We assume $\tilde{\lambda}_t \geq \nu$ which, as shown above, ensures that initial health investment is allocated towards the final good sector i_{ft}. Then the derivative evaluated at $\vartheta_t = 0$ is given by

$$\frac{\partial i_t}{\partial \vartheta_t} = \frac{\partial}{\partial \vartheta_t}\left(\nu + e_t\right)^{\zeta} = \frac{\partial}{\partial \vartheta_t}\left(\nu + \vartheta_t x_t\right)^{\zeta} = \zeta x_t \nu^{\zeta - 1}$$

Plugging $i_t(p_t, e_t = 0)$ and $\frac{\partial i_t}{\partial \vartheta_t}$ back into the corner solution condition yields

$$\frac{1}{\xi}\left(1+\nu^{\zeta}\right)\left(\left(1+\nu^{\zeta}\right)^{\xi}-1\right)R_{t+1} \geq \left(\frac{1}{1-\sigma}+bc_{t+1}^{\sigma-1}\right)c_{t+1}\zeta\nu^{\zeta-1}$$

Rewriting consumption in terms of overall cash in the first stage ($\vartheta_t = 0$) yields

$$c_{t+1} = R_{t+1}s_t = R_{t+1}(1-\vartheta_t)x_t = R_{t+1}x_t$$

Plugging back in then gives

$$\frac{1}{\xi}\left(1+\nu^{\zeta}\right)\left(\left(1+\nu^{\zeta}\right)^{\xi}-1\right) \geq \left(\frac{1}{1-\sigma}+b\left(R_{t+1}x_t\right)^{\sigma-1}\right)x_t\zeta\nu^{\zeta-1} \tag{B.9}$$

Thus, this equation characterizes the first stage: The region of cash x_t for which health expenditures are zero, $e_t = 0$. For convenience define

$$A_1 \equiv \frac{1}{\xi}\frac{\left(1+\nu^{\zeta}\right)\left(\left(1+\nu^{\zeta}\right)^{\xi}-1\right)}{\zeta\nu^{\zeta-1}} \begin{cases} = 0, & \text{if } \nu = 0 \\ > 0, & \text{if } \nu > 0 \end{cases}$$

Thus, $A_1 \geq 0$ and in our case we have $A_1 > 0$ because $\nu > 0$ is necessary for the existence of stage 2. Plugging A_1 back into the condition for the corner solution yields

$$A_1 \geq \left(\frac{1}{1-\sigma}+b\left(R_{t+1}x_t\right)^{\sigma-1}\right)x_t$$

To make progress analytically we need to make assumptions on the parameters. Assume $\sigma = 2$, then

$$\begin{aligned} A_1 &\geq -x_t + bR_{t+1}x_t^2 \\ 0 &\geq x_t^2 - \frac{1}{bR_{t+1}}x_t - \frac{A_1}{bR_{t+1}} \end{aligned}$$

The two solutions to this quadratic equation are given by

$$x_{1/2} = \frac{1}{2bR_{t+1}} \left(1 \pm \sqrt{1 + 4bR_{t+1}A_1}\right)$$

$A_1 \geq 0$ means that $\sqrt{1 + 4bR_{t+1}A_1} \geq 1$. Thus, $x_2 \leq 0$, then x_1 is the only economically relevant solution. The first kick-off separating stage 1 and stage 2 is then given by

$$\begin{aligned} x_{\text{kickoff1}} &= \frac{1}{2bR_{t+1}} \left(1 + \sqrt{1 + 4bR_{t+1}A_1}\right) \\ &= \frac{1}{bR_{t+1}} \left(\frac{1}{2} \left(1 + \sqrt{1 + 4bR_{t+1}A_1}\right)\right) \\ &\equiv \frac{1}{bR_{t+1}} \Delta(b, R_{t+1}, A_1) \end{aligned} \tag{B.10}$$

Notice that $\nu > 0$ ensures $A_1 > 0$ which in turn ensures $\Delta(b, R_{t+1}, A_1) > 1$. Further, $x_{\text{lowerbound}} = \frac{1}{bR_{t+1}}$ is the lower bound on cash such that there is no suicide. Then the interval

$$x \in [x_{\text{lowerbound}}, x_{\text{kickoff1}}] \tag{B.11}$$

is non-empty for $\nu > 0$ and characterizes the cash region for stage 1 without suicide.

B.2.2.3 Interior Solution of Optimal Health Investment

In the interior solution with positive health expenditures, $\vartheta_t > 0$, the FOC holds with equality

$$\frac{1}{\xi} (1 + i_t) \left((1 + i_t)^{\xi} - 1\right) x_t R_{t+1} = \left(\frac{1}{1 - \sigma} + b c_{t+1}^{\sigma - 1}\right) c_{t+1} \frac{\partial i_t}{\partial \vartheta_t}$$

Plugging in $c_{t+1} = R_{t+1} s_t = R_{t+1}(1 - \vartheta_t) x_t$ yields

$$\frac{1}{\xi} (1 + i_t) \left((1 + i_t)^{\xi} - 1\right) = \left(\frac{1}{1 - \sigma} + b \left(R_{t+1}(1 - \vartheta_t) x_t\right)^{\sigma - 1}\right) (1 - \vartheta_t) \frac{\partial i_t}{\partial \vartheta_t}$$

As before, in order to solve this equation for the share of health expenditure as a function of cash $\vartheta(x_t)$ we need to plug in i_t and $\frac{\partial i_t}{\partial \vartheta_t}$ which now both depend on whether we are in stage 2 or 3 (both stages are interior solutions with respect to the optimal ϑ_t). We know that we move from stage 2

to stage 3 when e_t is sufficiently large, formally when $e_t = \vartheta_t(x_t)x_t > \tilde{\lambda}_t - \nu$. For a given $x_t = \hat{x}$ we can solve the FOC and find $\vartheta(\hat{x})$ with the following steps

1. Guess $\hat{\vartheta}(\hat{x})$

2. Determine the current stage through $\hat{e} = \hat{\vartheta}(\hat{x})\hat{x}$. Specifically,

$$\text{Stage} = \begin{cases} 2, & \text{if} \quad \hat{e} \in (0, \tilde{\lambda}_t - \nu] \\ 3, & \text{if} \quad \hat{e} > \tilde{\lambda}_t - \nu \end{cases}$$

3. Compute $\hat{i}$ and $\frac{\partial \hat{i}}{\partial \hat{\vartheta}}$ based on the stage and check if the FOC holds

4. Repeat until $\hat{\vartheta}(\hat{x})$ that solves the FOC is found

In order to further characterize the level of cash x_{kickoff2} at which the second kick-off happens, we can utilitze that at the second kick-off total health expenditures e_t exactly equal the interior level of final good investment

$$\tilde{\lambda}_t - \nu = \vartheta_t(x_{\text{kickoff2}})x_{\text{kickoff2}}$$

Moreover, at the second kick-off health investment and its derivative w.r.t. the expenditure share are given by

$$\begin{aligned} i_t &= \lambda_t^{\zeta} \\ \frac{\partial i_t}{\partial \vartheta_t} &= \frac{\lambda_t - \nu}{p_t \vartheta_t} \end{aligned}$$

Plugging those two into the FOC yields

$$\frac{1}{\xi}\left(1 + \lambda_t^{\zeta}\right)\left(\left(1 + \lambda_t^{\zeta}\right)^{\xi} - 1\right) = \left(\frac{1}{1-\sigma} + b\left(R_{t+1}(1-\vartheta_t)x_t\right)^{\sigma-1}\right)(1-\vartheta_t)\frac{\lambda_t - \nu}{p_t \vartheta_t}$$

Now using the fact that at the second kick-off $\tilde{\lambda}_t - \nu = \vartheta_t(x)x$ we get

$$\frac{1}{\xi}\left(1+\lambda_t^{\zeta}\right)\left(\left(1+\lambda_t^{\zeta}\right)^{\xi}-1\right) = \left(\frac{1}{1-\sigma} + b\left(R_{t+1}x - R_{t+1}(\lambda_t-\nu)\right)^{\sigma-1}\right)(1-\vartheta_t)\frac{\lambda_t-\nu}{p_t\vartheta_t}$$

As before assume $\sigma = 2$, then

$$\frac{1}{\xi}\left(1+\lambda_t^{\zeta}\right)\left(\left(1+\lambda_t^{\zeta}\right)^{\xi}-1\right) = \left(-1 + b\left(R_{t+1}x - R_{t+1}(\lambda_t-\nu)\right)\right)\frac{(1-\vartheta_t)}{\vartheta_t}\frac{\lambda_t-\nu}{p_t}$$

We can again use $\tilde{\lambda}_t - \nu = \vartheta_t(x)x$ to derive

$$\frac{(1-\vartheta_t)}{\vartheta_t} = \frac{x-(\lambda_t-\nu)}{(\lambda_t-\nu)} = \frac{x}{(\lambda_t-\nu)} - 1$$

and thereby eliminate the remaining ϑ_t to get

$$\frac{1}{\xi}\left(1+\lambda_t^{\zeta}\right)\left(\left(1+\lambda_t^{\zeta}\right)^{\xi}-1\right) = \left(-1 + b\left(R_{t+1}x - R_{t+1}(\lambda_t-\nu)\right)\right)\left(\frac{x}{(\lambda_t-\nu)} - 1\right)\frac{\lambda_t-\nu}{p_t}$$

Similar to before when finding the first kick-off, again define for convenience

$$A_2 \equiv \frac{p_t}{\xi}\left(1+\lambda_t^{\zeta}\right)\left(\left(1+\lambda_t^{\zeta}\right)^{\xi}-1\right)\begin{cases} = 0, & \text{if } \nu = 0 \\ > 0, & \text{if } \nu > 0 \end{cases}$$

Then the FOC at the second kick-off reduces to

$$A_2 = \left(x-(\lambda_t-\nu)\right)\left(bR_{t+1}x - bR_{t+1}(\lambda_t-\nu) - 1\right)$$

This is again a quadratic equation in x, simplifying further yiels

$$0 = x^2 - x\frac{2(\lambda_t-\nu)bR_{t+1}+1}{bR_{t+1}} + \frac{(\lambda_t-\nu)^2 bR_{t+1} + (\lambda_t-\nu) - A_2}{bR_{t+1}}$$

The roots of this equation are given by

$$x_{1/2} = \frac{2(\lambda_t - \nu)bR_{t+1} + 1}{2bR_{t+1}} \pm \sqrt{\left(\frac{2(\lambda_t - \nu)bR_{t+1} + 1}{2bR_{t+1}}\right)^2 - \frac{(\lambda_t - \nu)^2 bR_{t+1} + (\lambda_t - \nu) - A_2}{bR_{t+1}}}$$

which simplifies to

$$\begin{aligned} x_{1/2} &= \frac{1}{2bR_{t+1}}\left(2(\lambda_t - \nu)bR_{t+1} + 1 \pm \sqrt{1 + 4A_2 bR_{t+1}}\right) \\ &= (\lambda_t - \nu) + \frac{1}{2bR_{t+1}}\left(1 \pm \sqrt{1 + 4A_2 bR_{t+1}}\right) \end{aligned}$$

This looks very similar to the solution for the first kick-off. The new term in the beginning $(\lambda_t - \nu)$ is exactly the lower bound for the second kick-off because by definition it has to be that $x_{\text{kickoff2}} \geq i_{ft} = (\lambda_t - \nu)$. Since as before $4A_2 bR_{t+1} > 0$ if $\nu > 0$ the negative solution violates that condition and can be ruled out. Then the second kick-off is given by

$$\begin{aligned} x_{\text{kickoff2}} &= (\lambda_t - \nu) + \frac{1}{2bR_{t+1}}\left(1 + \sqrt{1 + 4bR_{t+1}A_2}\right) \\ &= (\lambda_t - \nu) + \frac{1}{bR_{t+1}}\left(\frac{1}{2}\left(1 + \sqrt{1 + 4bR_{t+1}A_2}\right)\right) \\ &\equiv (\lambda_t - \nu) + \frac{1}{bR_{t+1}}\Delta(b, R_{t+1}, A_2) \\ &= i_{ft} + \frac{1}{bR_{t+1}}\Delta(b, R_{t+1}, A_2) \end{aligned} \tag{B.12}$$

Notice that the wedge $\Delta(b, R_{t+1}, A)$ appears exactly the same way in both kick-offs. The only difference is the value of A plugged into the wedge function. To compare the relative size of those

wedges first note that the wedge is increasing in A and further notice that

$$\begin{aligned} A_2 &= \frac{p_t}{\xi}\left(1+\lambda_t^{\zeta}\right)\left(\left(1+\lambda_t^{\zeta}\right)^{\xi}-1\right) \\ &> \frac{p_t}{\xi}\left(1+\nu^{\zeta}\right)\left(\left(1+\nu^{\zeta}\right)^{\xi}-1\right) \\ &> \frac{1}{\xi}\frac{\left(1+\nu^{\zeta}\right)\left(\left(1+\nu^{\zeta}\right)^{\xi}-1\right)}{\zeta\nu^{\zeta-1}} \\ &= A_1 \end{aligned}$$

where the last inequality comes from the FOC of the optimal division of health investment $\zeta\left(\nu+i_{ft}\right)^{\zeta-1}=\frac{1}{p_t}$ and, thus, $p_t>\zeta\left(\nu\right)^{\zeta-1}$ for $i_{ft}>0$. We can directly link the wedges to savings at the kick-offs. Then the amount by which savings grow between the two kick-offs is given by

$$\Delta s = s_{\text{kickoff2}} - s_{\text{kickoff1}} = \frac{1}{bR_{t+1}}\left(\Delta(b,R_{t+1},A_2)-\Delta(b,R_{t+1},A_1)\right)$$

Thus, it makes sense that we get $\Delta(b,R_{t+1},A_2)>\Delta(b,R_{t+1},A_1)$. Also note that if $\lambda_t\leq\nu$ (stage 2 does not exist), we get $A_2=A_1$ and the two kick-offs derived above collapse to the same cash value. (not yet sure expressing quantities in terms of these wedges is particularly useful but it is a good sanity check).

B.2.2.4 Interior Solution on the BGP

In the interior solution of the BGP we have

$$\frac{\partial i_t}{\partial \vartheta_t} = \frac{\partial}{\partial \vartheta_t}\frac{1}{p_t}\left(\vartheta_t x_t+\nu_t+\frac{1-\zeta}{\zeta}\tilde{\lambda}_t\right)=\frac{x_t}{p_t}$$

Then the FOC for ϑ_t becomes

$$\frac{1}{\xi}\left(1+i_t\right)\left(\left(1+i_t\right)^{\xi}-1\right)p_t R_{t+1}=\left(\frac{1}{1-\sigma}+bc_{t+1}^{\sigma-1}\right)c_{t+1}$$

p_t is constant, x_t converges to infinity and $i_t \to i_{ht} \to \frac{e_t}{p_t}$. Thus i_t and c_{t+1} are both constant shares of cash x_t on the BGP. For a BGP to exist we therefore require $\xi = \sigma - 1$. Plugging in

$$\begin{aligned} i &= \frac{\vartheta x}{p} \\ c &= R(1-\vartheta)x \end{aligned}$$

and solving for the limit case where $x \to \infty$ we can find the health expenditure share on the BGP

$$\vartheta^* = \left(1 + \left[\frac{(pR)^{1-\sigma}}{b\xi}\right]^{\frac{1}{\sigma}}\right)^{-1}$$

B.2.2.5 Summarizing the Phases

To summarize, the three phases are characterized by three separate region of cash at hand. The thresholds are given by

$$\begin{aligned} x_{\text{LB}} &= \frac{1}{bR_{t+1}} = \frac{1}{bR_2} \\ x_{\text{KO1}} &= \frac{1}{bR_{T_1+1}}\Delta(b, R_{T_1+1}, A_1) \\ x_{\text{KO2}} &= (\lambda_t - \nu) + \frac{1}{bR_{T_2+1}}\Delta(b, R_{T_2+1}, A_2) \end{aligned}$$

where T_1 and T_2 are the points in time at which the two kick-offs occur. The second equality in the first equation comes from the assumption that cash starts at its lower bound in period 1 $x_1 = \frac{1}{bR_2}$. Recall that

$$\begin{aligned} A_1 &= \frac{1}{\xi}\frac{\left(1+\nu^\zeta\right)\left(\left(1+\nu^\zeta\right)^\xi - 1\right)}{\zeta\nu^{\zeta-1}} \\ A_2 &= \frac{p_t}{\xi}\left(1+\lambda_t^\zeta\right)\left(\left(1+\lambda_t^\zeta\right)^\xi - 1\right) \\ \Delta(b, R_{t+1}, A) &= \frac{1}{2}\left(1 + \sqrt{1+4bR_{t+1}A}\right) \end{aligned}$$

Note that A_1 only depends on time-invariant parameters. Then the three regions are given by

$$\text{Phase} = \begin{cases} 1, & \text{if} \quad x_t \in [x_{\text{LB}}, x_{\text{KO1}}) \\ 2, & \text{if} \quad x_t \in [x_{\text{KO1}}, x_{\text{KO2}}) \\ 3, & \text{if} \quad x_t \in [x_{\text{KO2}}, \infty) \end{cases}$$

APPENDIX C

Appendix for Chapter 3

C.1. Generating the Polynomials

Given a base random variable ν the corresponding polynomials are generated with the three-term recurrence (see e.g. Gautschi (1982) or Zheng et al. (2015)) relation given by

$$\Psi_{i+1}(\nu) = (\nu - \gamma_i)\Psi_i(\nu) - \beta_i \Psi_{i-1}(\nu), \quad i \geq 1$$

where $\Psi_0(\nu) = 0$. The sequence is initialized with $\Psi_1(\nu) = 1$. As discussed, the first polynomial is always constant and equal to 1 which yields that the first coefficient $\alpha_{t,1}$ is equal to the mean of the distribution. The parameters $\gamma_i, \beta_i \in \mathbb{R}$ can be generated in different ways as discussed in Zheng et al. (2015). Following Proehl (2017), I use the Stieltjes method explained in detail in Gautschi (1982).

$$\gamma_i = \frac{\langle \Psi_i, \nu\Psi_i \rangle}{\langle \Psi_i, \Psi_i \rangle}$$
$$\beta_i = \frac{\langle \Psi_i, \Psi_i \rangle}{\langle \Psi_{i-1}, \Psi_{i-1} \rangle}$$

where $\langle :, : \rangle$ is the standard inner product in L^2 with respect to the base distribution ν.

C.2. Discretizing the Higher-Order Process

C.2.1. Targets and Calibrated Parameters

The earnings process is characterized as follows

$$y_t = \exp\left(p_t + \epsilon_t\right), \qquad \epsilon_t \underset{iid}{\sim} \mathcal{N}(0, \sigma_\epsilon^2)$$
$$p_t = \rho p_{t-1} + \eta_t, \qquad \eta_t \underset{iid}{\sim} \begin{cases} \mathcal{N}(\mu_1, \sigma_1^2) \text{ with } p_1 \\ \mathcal{N}(\mu_2, \sigma_2^2) \text{ with } 1 - p_1 \end{cases}$$

Table C.1: Unconditional moments of log earnings for higher-order earnings process, targets from Guvenen et al. (2021)

	Targets	Discretized process
Variance(x)	0.58	0.59
Variance($\Delta_1 x$)	0.23	0.19
Skewness($\Delta_1 x$)	-1.35	-1.39
Kurtosis($\Delta_1 x$)	17.8	15.72
Variance($\Delta_5 x$)	0.46	0.51
Skewness($\Delta_5 x$)	-1.01	-1.25
Kurtosis($\Delta_5 x$)	11.55	9.88

I use GMM to solve for the calibration of the above process as well as the grid for the persistent component p_t to match a set of moments of the earnings process. Let $\vec{p}$ denote the grid of persistent earnings then the set of parameters to choose is $\theta = (p_1, \mu_1, \mu_2, \sigma_1, \sigma_2, \sigma_\epsilon, \vec{p})$. Given the standard deviation σ_ϵ the transitory process is discretized with Tauchen [41]. I first use a global solver and afterwards improve on the initial solution with a more powerful local solver. The local solver targets the same set of moments but only chooses the grid for persistent earnings $\theta_{local} = \vec{p}$, not the calibration parameters for the process.

Let x denote log-earnings and $\Delta_j x = x_t - x_{t-j}$ the j-year growth rate of log-earnings. Table C.1 shows the set of targets as well as corresponding moments of the discretized earnings process.

C.2.2. Moments of the Discretized Process

This section briefly derives the targeted moments of the discretized earnings process given a set of parameters θ. Let $\mathcal{F}$ and $\mathbf{f}$ denote the cdf and pdf of the Gaussian mixture distribution of the innovations of the persistent component, given a set of calibration parameters. Further, let $\vec{p}$ denote the grid of persistent earnings and n_p, n_ϵ the grid sizes of persistent and transitory components, respectively.

Based on $\vec{p}$ one generates another grid $\vec{d}$ across the real numbers with buckets for each persistent earnings state given by

$$\vec{d}(i) = \frac{\vec{p}(i-1) + \vec{p}(i)}{2}, \qquad i \geq 2$$

[41] I discretize the transitory component into 3 possible values and the persistent component into a 7 point Markov chain. This generates a 21 point Markov chain for overall earnings

where $\vec{d}(1) = -\infty$ and $\vec{d}(n_p) = \infty$. Then the transition matrix T_p for persistent earnings can be generated in the following way

$$T_p(r, c) = \mathcal{F}(\vec{d}(c+1) - \rho\vec{d}(r)) - \mathcal{F}(\vec{d}(c) - \rho\vec{d}(r)).$$

Given the transition matrix for persistent earnings, it is straight forward to derive the corresponding invariant distribution of persistent earnings π_p. The transitory component is discretized with Tauchen's method which yields the corresponding grid $\vec{\epsilon}$, transition matrix T_ϵ, and invariant distribution π_ϵ. The possible values and, thus, the grid for overall earnings $\vec{y}$ follows from all possible combinations of persistent and transitory component according to the specified process $y(p, \epsilon) = \exp(p + \epsilon)$. The transition probabilities for earnings T_y are given by

$$T_y(y'(p', \epsilon')|y(p, \epsilon)) = T_p(p'|p) \cdot T_\epsilon(\epsilon'|\epsilon)$$

As with the persistent component the invariant distribution of earnings π_y can be derived by iterating over the transition matrix T_y until convergence. The grid $\vec{y}$, transition matrix T_y and invariant distribution π_y fully characterize the AR(1) process for earnings and are sufficient to derive the resulting moments.

BIBLIOGRAPHY

Daron Acemoglu. Technical change, inequality, and the labor market. *Journal of Economic Literature*, 40(1):7–72, 2002.

Daron Acemoglu and Veronica Guerrieri. Capital Deepening and Nonbalanced Economic Growth. *Journal of Political Economy*, 116(3):467–498, 2008. ISSN 0022-3808. doi: 10.1086/589523. URL http://www.journals.uchicago.edu/doi/10.1086/589523.

Daron Acemoglu and Joshua Linn. Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry*. *The Quarterly Journal of Economics*, 119(3):1049–1090, 2004.

Daron Acemoglu and Pascual Restrepo. Robots and jobs: Evidence from us labor markets. *Journal of Political Economy*, 128(6):2188–2244, 2020. ISSN 1537534X. doi: 10.1086/705716.

Rodrigo Adao, Martin Beraja, and Nitya Pandalai-Nayar. Fast and Slow Technological Transitions. *Journal of Political Economy Macroeconomics*, 2024.

Philippe Aghion and Peter Howitt. A Model of Growth through Creative Desctruction. *Econometrica*, 60:323–351, 1992.

Philippe Aghion and Peter Howitt. *Endogenous Growth Theory*. MIT Press, Cambridge, MA, 1998.

S. R. Aiyagari. Uninsured Idiosyncratic Risk and Aggregate Saving. *The Quarterly Journal of Economics*, 109(3):659–684, 1994. ISSN 0033-5533. doi: 10.2307/2118417.

Gerard F Anderson, Uwe E Reinhardt, and Peter S Hussey. It's The Prices, Stupid : Why The United States Is So Different From Other Countries. *Health Affairs*, 22(3):89–105, 2003. doi: 10.1377/hlthaff.22.3.89.

Adrien Auclert, Bence Bardóczy, Matthew Rognlie, and Ludwig Straub. Using the Sequence-Space Jacobian to Solve and Estimate Heterogeneous-Agent Models. *Econometrica*, 89(5):2375–2408, 2021. ISSN 0012-9682. doi: 10.3982/ecta17434.

David Autor. The Polarization of Job Opportunities in the U. S. Labor Market. *Community Investments*, 23(April):360–361, 2010. URL http://econ-www.mit.edu/files/5554.

David Autor, Frank Levy, and Richard Murnane. The Skill Content of Recent Technological Change: An Empirical Exploration. *The Quarterly Journal of Economics*, 118(4):1279–1333, 2003.

David Autor, David Dorn, Gordon Hanson, and Kaveh Majlesi. Importing political polarization? The electoral consequences of rising trade exposure. *American Economic Review*, 110(10):3139–3183, 2020. ISSN 19447981. doi: 10.1257/aer.20170011.

David H. Autor, Lawrence F Katz, and Alan B Krueger. Computing Inequality: Have Computers Changed the Labor Market ? *The Quarterly Journal of Economics*, 113(4):1169–1213, 1998.

Laurent Bach, Laurent E. Calvet, and Paolo Sodini. Rich Pickings? Risk, Return, and Skill in Household Wealth. *American Economic Review*, 110(9):2703–2747, 2020. ISSN 19447981. doi: 10.1257/aer.20170666.

George Baker and Bengt Holmstrom. Internal Labor Markets: Too Many Theories, Too Few Facts. *The American Economic Review*, 85(2):255–259, 1995.

George Baker, Michael Gibbs, and Bengt Holmstrom. The Internal Economics of the Firm : Evidence from Personnel Data. *Quarterly Journal of Economics*, 109(4):881–919, 1994.

Isaac Baley, Lars Ljungqvist, and Thomas J Sargent. Quit Turbulence and Unemployment. From the perspective of a particle physicist. 2020.

Isaac Baley, Lars Ljungqvist, and Thomas J. Sargent. Cross-phenomenon restrictions: Unemployment effects of layoff costs and quit turbulence. *Review of Economic Dynamics*, 50:43–60, 2023. ISSN 10942025. doi: 10.1016/j.red.2023.07.004. URL https://doi.org/10.1016/j.red.2023.07.004.

Andrea Bassanini and Romain Duval. Employment Patterns in OECD Countries: Reassessing the Role of Policies and Institutions. *OECD Social, Employment and Migration Working Papers*, (4):1–129, 2006. ISSN 02550822.

Christian Bayer and Moritz Kuhn. Job levels and Wages. 2023. URL https://papers.ssrn.com/sol3/papers.cfm?abstract{_}id=4464590.

Martin Beraja and Nathan Zorzi. Inefficient Automation. *The Review of Economic Studies*, 2022. doi: 10.2139/ssrn.4140871.

Eli Berman, John Bound, and Zvi Griliches. Changes in the Demand for Skilled Labor within U . S . Manufacturing: Evidence from the Annual Survey of Manufacturers. *The Quarterly Journal of Economics*, 109(2):367–397, 1994.

Ernst R Berndt, David M Cutler, Richard G Frank, Zvi Griliches, Joseph P Newhouse, and Jack E Triplett. Medical Care Prices and Output. In A J Cuyler and Joseph P Newhouse, editors, *Handbook of Health Economics*, volume 1, chapter 3. Elsevier Science B.V.

James E. Bessen, Maarten Goos, Anna Salomons, and Wiljan Van den Berge. Automatic Reaction - What Happens to Workers at Firms that Automate? *Review of Economics and Statistics*, 2023. doi: 10.2139/ssrn.3328877.

Truman Bewley. The permanent income hypothesis: A theoretical formulation. *Journal of Economic Theory*, 16(2):252–292, 1977. ISSN 10957235. doi: 10.1016/0022-0531(77)90009-6.

Olivier J. Blanchard. Debt, Deficits, and Finite Horizons. *Journal of Political Economy*, 93(2): 223–247, 1985. ISSN 0022-3808. doi: 10.1086/261297.

Olivier J. Blanchard and Lawrence H. Summers. Hysteresis and the European unemployment problem. *NBER Macroeconomics Annual*, 1(July):1–23, 1986.

Sebastian Böhm, Volker Grossmann, and Holger Strulik. R&D-driven medical progress, health care costs, and the future of human longevity. *Journal of the Economics of Ageing*, 18, 2021. ISSN 2212828X. doi: 10.1016/j.jeoa.2020.100286.

Timo Boppart, Per Krusell, and Kurt Mitman. Exploiting MIT shocks in heterogeneous-agent economies: the impulse response as a numerical derivative. *Journal of Economic Dynamics and Control*, 89:68–92, 2018. ISSN 01651889. doi: 10.1016/j.jedc.2018.01.002. URL https://doi.org/10.1016/j.jedc.2018.01.002.

Christine Borger, Thomas F Rutherford, and Gregory Y Won. Projecting long term medical spending growth. *Journal of Health Economics*, 27:69–88, 2008. doi: 10.1016/j.jhealeco.2007.03.003.

Juan Botero, Simeon Djankov, Rafael La Porta, Florencio Lopez-De-Silanes, and Andrei Shleifer. The regulation of labor* j. *The Quarterly Journal of Economics*, (November):1339–1382, 2004.

By Markus K Brunnermeier and Yuliy Sannikov. A Macroeconomic Model with a Financial Sector. *American Economic Review*, 104(2):379–421, 2014.

Christopher Busch and Alexander Ludwig. Higher-Order Income Risk Over the Business Cycle. *International Economic Review*, 0(0), 2023. ISSN 14682354. doi: 10.1111/iere.12685.

Marco Cagetti and Mariacristina De Nardi. Entrepreneurship , Frictions , and Wealth. *Journal of Political Economy*, 114(5):835–870, 2006.

O Cardi and R Restout. Imperfect Mobility of Labor Across Sectors: A Reappraisal of the Balassa-Samuelson Effect. *Journal of International Economics*, 97(2):249–265, 2015.

Christopher Carroll, Jiri Slacalek, Kiichi Tokuoka, and Matthew N. White. The distribution of wealth and the marginal propensity to consume. *Quantitative Economics*, 8(3):977–1020, 2017. ISSN 1759-7331. doi: 10.3982/qe694.

Christopher D. Carroll, Jiri Slacalek, and Kiichi Tokuoka. Buffer-stock saving in a Krusell-Smith world. *Economics Letters*, 132:97–100, 2015. ISSN 01651765. doi: 10.1016/j.econlet.2015.04.021. URL http://dx.doi.org/10.1016/j.econlet.2015.04.021.

Alan A Carruth and Andrew J Oswald. On Union Preferences and Labour Market Models: Insiders and Outsiders. *The Economic Journal*, 97(386):431–445, 1987.

Alan Cartter. *Theory of wages and employment*. Homewood; Ill.: R.D. Irwin, 1959.

Matteo Cervellati and Uwe Sunde. Human Capital Formation, Life Expectancy, and the Process of Development. *The American Economic Review*, 95(5):1653–1672, 2005.

Amitabh Chandra and Jonathan Skinner. Technology Growth and Expenditure Growth in Health Care. *Journal of Economic Literature*, 50(3):645–680, 2012. ISSN 0022-0515. doi: 10.1257/jel.50.3.645. URL http://pubs.aeaweb.org/doi/10.1257/jel.50.3.645.

Yongsung Chang, Sun Bin Kim, Kyooho Kwon, and Richard Rogerson. 2018 Klein Lecture: Individual and Aggregate Labor Supply in Heterogeneous Agent Economies With Intensive and Extensive Margins. *International Economic Review*, 60(1):3–24, 2019. ISSN 14682354. doi: 10.1111/iere.12377.

Simon Civale, Luis Diez-Catalan, and Fatih Fazilet. Discretizing a Process with Non-zero Skewness and High Kurtosis. 2017.

Fabrizio Colonna. Does Union Membership increase Job Security? Evidence from British Panel Data. 2008.

Congressional Budget Office. Technological Change and the Growth of Health Care Spending. 359 (1):50–60, 2008. ISSN 1533-4406. doi: 10.1056/NEJMsa0802005.

Zack Cooper, Stuart V Craig, Martin Gaynor, and John Van Reenen. The price ain't right? hospital prices and health spending on the privately insured. *Quarterly Journal of Economics*, 134(1): 51–107, 2019. ISSN 15314650. doi: 10.1093/qje/qjy020.

Guido Matias Cortes, Christopher J. Nekarda, Nir Jaimovich, and Henry E. Siu. The dynamics of disappearing routine jobs: A flows approach. *Labour Economics*, 65(October 2019):101823, 2020. ISSN 09275371. doi: 10.1016/j.labeco.2020.101823. URL https://doi.org/10.1016/j.labeco.2020.101823.

Arnaud Costinot and Iván Werning. Robots, Trade, and Luddism: A Sufficient Statistic Approach to Optimal Technology Regulation. *Review of Economic Studies*, 90(5):2261–2291, 2023. ISSN 1467937X. doi: 10.1093/restud/rdac076.

David M Cutler and Mark McClellan. Is technological change in medicine worth it? *Health Affairs*, 20(5):11–29, 2001. ISSN 02782715. doi: 10.1377/hlthaff.20.5.11.

Josh Davis, Cristian Fuenzalida, Leon Huetsch, Benjamin Mills, and Alan M. Taylor. Global Natural Rates in the Long Run: Postwar Macro Trends and the Market-Implied R* in 10 Advanced Economies. *Journal of International Economics*, 2024. ISSN 18730353. doi: 10.2139/ssrn.4603121. URL https://doi.org/10.1016/j.jinteco.2024.103919{%}0A.

Mariacristina De Nardi, Giulio Fella, and Gonzalo Paz-Pardo. Nonlinear Household Earnings

Dynamics, Self-Insurance, and Welfare. *Journal of the European Economic Association*, 18(2): 890–926, 2020. ISSN 15424774. doi: 10.1093/jeea/jvz010.

Wouter J. Den Haan. Heterogeneity, Aggregate Uncertainty, and the Short-Term Interest Rate. *Journal of Business & Economic Statistics*, 14(4):399–411, 1996. ISSN 0091-6749. URL http://dx.doi.org/10.1016/j.jaci.2012.05.050.

Wouter J. Den Haan. Solving dynamic models with aggregate shocks and heterogeneous agents. *Macroeconomic Dynamics*, 1(2):355–386, 1997. ISSN 13651005. doi: 10.1017/s1365100597003040.

Wouter J. Den Haan. Assessing the accuracy of the aggregate law of motion in models with heterogeneous agents. *Journal of Economic Dynamics and Control*, 34(1):79–99, 2010. ISSN 01651889. doi: 10.1016/j.jedc.2008.12.009. URL http://dx.doi.org/10.1016/j.jedc.2008.12.009.

Peter A Diamond. National Debt in a Neoclassical Growth Model. *American Economic Review*, 55:1126–1150, 1965.

Peter Doeringer and Michael Piore. *Internal labor markets and manpower analysis*. 1985. ISBN 9781000122572. doi: 10.4324/9781003069720.

Thomas J. Dohmen, Ben Kriechel, and Gerard A. Pfann. Monkey bars and ladders: The importance of lateral and vertical job mobility in internal labor market careers. *Journal of Population Economics*, 17(2):193–228, 2004. ISSN 09331433. doi: 10.1007/s00148-004-0191-4.

J T Dunlop. *Wage Determination Under Trade Unions*. Reprints of economic classics. Macmillan Company, 1944. URL https://books.google.com/books?id=WtlfAAAAIAAJ.

Maximiliano A Dvorkin and Alexander Monge-Naranjo. Occupation Mobility, Human Capital and the Aggregate Consequences of Task-Biased Innovations. 2019.

Isaac Ehrlich and Yong Yin. Equilibrium Health Spending and Population Aging in a Model of Endogenous Growth: Will the GDP Share of Health Spending Keep Rising? *Journal of Human Capital*, 7(4):411–447, 2013. ISSN 1932-8575. doi: 10.1086/675640. URL http://www.journals.uchicago.edu/doi/10.1086/675640.

Andreas Fagereng, Luigi Guiso, Davide Malacrino, and Luigi Pistaferri. Heterogeneity and Persistence in Returns to Wealth. *Econometrica*, 88(1):115–170, 2020.

Hanming Fang and Xincheng Qiu. "Golden Ages": A Tale of the Labor Markets in China and the United States. *Journal of Political Economy Macroeconomics*, 1(4), 2023. doi: 10.2139/ssrn.3973698.

Maria Feldman and Nick Pretnar. The Causal Factors Driving the Rise in Health Services Prices, 2023.

W Fellner. *Competition Among the Few: Oligopoly and Similar Market Structures*. A Borzoi book. A.A. Knopf, 1949. URL https://books.google.com/books?id=yWwsAAAAMAAJ.

Jesús Fernández-Villaverde, Samuel Hurtado, and Galo Nano. Financial Frictions and the Wealth Distribution. *Econometrica*, 91(3):869–901, 2023.

Raquel Fonseca, Pierre-Carl Michaud, Titus Galama, and Arie Kapteyn. Accounting for the Rise of Health Spending and Longevity. *Journal of the European Economic Association*, 19(1):536–579, 2021.

Raquel Fonseca, François Langot, Pierre Carl Michaud, and Thepthida Sopraseuth. Understanding Cross-Country Differences in Health Status and Expenditures: Health Prices Matter. *Journal of Political Economy*, 131(8):1949–1993, 2023. ISSN 1537534X. doi: 10.1086/724113.

Ivan Frankovic and Michael Kuhn. Health Insurance, Endogenous Medical Progress, and Health Expenditure Growth. *Journal of Health Economics*, 87:doi: 10.1016/j.jhealeco.2022.102717, 2023.

Ivan Frankovic, Michael Kuhn, and Stefan Wrzaczek. Medical Progress, Demand for Health Care, and Economic Performance. Working Paper, 2017.

Oded Galor. *Unified Growth Theory*. Princeton University Press, Princeton, 2011.

Walter Gautschi. On Generating Orthogonal Polynomials. *SIAM Journal on Scientific and Statistical Computing*, 3(3):289–317, 1982. ISSN 0196-5204. doi: 10.1137/0903018.

Marta Giagheddu and Andrea Papetti. Demograpics and the Real Exchange Rate. Working Paper, 2020.

Claudia Goldin and Lawrence F. Katz. *The Race Between Education and Technology*. Belknap„ Cambridge, Mass. ; London :, 2008. ISBN 0-674-02867-8.

Mikhail Golosov, Maxim Troshkin, and Aleh Tsyvinski. Redistribution and social insurance. *American Economic Review*, 106(2):359–386, 2016. ISSN 00028282. doi: 10.1257/aer.20111550.

Maarten Goos and Alan Manning. Lousy and Lovely Jobs : The Rising Polarization of Work in Britain. *The Review of Economics and Statistics*, 89(1):118–133, 2007.

Maarten Goos, Alan Manning, and Anna Salomons. Explaining job polarization: Routine-biased technological change and offshoring. *American Economic Review*, 104(8):2509–2526, 2014. ISSN 19447981. doi: 10.1257/aer.104.8.2509.

Robert F Graboyes. Medical Care Price Indexes. *FRB Richmond Economic Quarterly*, 80(4):69–89, 1994.

Georg Graetz and Guy Michaels. Robots at work. *The Review of Economics and Statistics*, 100

(5):753–768, 2018. doi: 10.1162/rest.

Michael Grossman. On the Concept of Health Capital and the Demand for Health. *Journal of Political Economy*, 80(2):223–255, 1972.

Joao Guerreiro, Sergio Rebelo, and Pedro Teles. Should Robots Be Taxed? *Review of Economic Studies*, 89(1):279–311, 2022. ISSN 1467937X. doi: 10.1093/restud/rdab019.

Adam Guren, David Hémous, and Morten Olsen. Trade dynamics with sector-specific human capital. *Journal of International Economics*, 97(1):126–147, 2015. ISSN 18730353. doi: 10.1016/j.jinteco.2015.04.003. URL http://dx.doi.org/10.1016/j.jinteco.2015.04.003.

Fatih Guvenen, Serdar Ozkan, and Jae Song. The Nature of Countercyclical Income Risk. *Journal of Political Economy*, 122(3):1–59, 2014. ISSN 1556-5068. doi: 10.2139/ssrn.2046642. URL http://www.ssrn.com/abstract=2046642.

Fatih Guvenen, Fatih Karahan, Serdar Ozkan, and Jae Song. What Do Data on Millions of U.S. Workers Reveal about Life-Cycle Earnings Dynamics? *Econometrica*, 89(5):2303–2339, 2021.

Wouter J Den Haan, Christian Haef, and Garey Ramey. Turbulence and Unemployment in a Job Matching Model. *Journal of the European Economic Association*, 3(6):1360–1385, 2005.

J David Hacker. Decennial Life Tables for the White Population of the United States, 1790-1900. *Historical Methods*, 43(2):45–79, 2010. ISSN 01615440. doi: 10.1080/01615441003720449.

Michael R Haines. Estimated Life Tables for the United States, 1850-1900, 1994.

Robert E Hall and Charles I Jones. The Value of Life and the Rise in Health Spending. *Quarterly Journal of Economics*, 122(1):39–72, 2007.

G Hansen and Edward C Prescott. Malthus to Solow. *American Economic Review*, 92 (4):1205–1217, 2002.

James J Heckman, Lance Lochner, and Christopher Taber. Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogeneous Agents. *Review of Economic Dynamics*, 1:1–58, 1998.

John Hejkal, B Ravikumar, and Guillaume Vandenbroucke. Technology Adoption, Mortality, and Population Dynamics. 2022.

Alex R Horenstein and Manuel S Santos. Understanding growth patterns in US health care expenditures. *Journal of the European Economic Association*, 17(1):284–326, 2019. ISSN 15424774. doi: 10.1093/jeea/jvx059.

Joachim Hubmer. The Race Between Preferences and Technology. *Econometrica*, 91(1):227–261,

2023. ISSN 0012-9682. doi: 10.3982/ecta18580.

Joachim Hubmer, Per Krusell, and Anthony A Smith. Sources of us wealth inequality: Past, present, and future. *NBER Macroeconomics Annual*, 35(1):392–455, 2021. ISSN 15372642. doi: 10.1086/712332.

Mark Huggett. The risk-free rate in heterogeneous-agent incomplete-insurance economies. *Journal of Economic Dynamics and Control*, 17(5-6):953–969, 1993. ISSN 01651889. doi: 10.1016/0165-1889(93)90024-M.

Anders Humlum. Robot Adoption and Labor Market Dynamics. 2021.

Zhen Huo and José Víctor Ríos-Rull. Demand induced fluctuations. *Review of Economic Dynamics*, 37:S99–S117, 2020. ISSN 10942025. doi: 10.1016/j.red.2020.06.011.

Jones. Why Have Health Expenditures as a Share of GDP Risen So Much? 2004.

Charles Jones. Life and Growth. *Journal of Political Economy*, 124(2):1–36, 2016. ISSN 0022-3808. doi: 10.3386/w17094.

Lawrence F. Katz and David H. Autor. Chapter 26 Changes in the wage structure and earnings inequality. *Handbook of Labor Economics*, 3 PART(1):1463–1555, 1999. ISSN 15734463. doi: 10.1016/S1573-4463(99)03007-2.

Michael Koch, Ilya Manuylov, and Marcel Smolka. Robots and Firms. *Economic Journal*, 131 (638):2553–2584, 2021. ISSN 14680297. doi: 10.1093/ej/ueab009.

Ralph S J Koijen, Tomas J Philipson, and Harald Uhlig. Financial Health Economics. *Econometrica*, 84(1):195–242, 2016. ISSN 0012-9682. doi: 10.3982/ECTA11182. URL https://www.econometricsociety.org/doi/10.3982/ECTA11182.

Dirk Krueger and Alexander Ludwig. On the Optimal Provision of Social Insurance: Progressive Taxation versus Education Subsidies in General Equilibrium. *Journal of Monetary Economics*, 77, 2016. ISSN 03043932. doi: 10.1016/j.jmoneco.2015.11.002. URL forthcoming.

Dirk Krueger, Kurt Mitman, and Fabrizio Perri. On the Distribution of the Welfare Losses of Large Recessions. *NBER Working Paper*, 2016. doi: 10.1017/9781108227223.006.

Per Krusell and Anthony A. Smith. Income and wealth heterogeneity in the macroeconomy. *Journal of Political Economy*, 106(5):867–896, 1998. ISSN 00223808. doi: 10.1086/250034.

Michael Kuhn and Klaus Prettner. Growth and Welfare Effects of Health Care in Knowledge-based Economies. *Journal of Health Economics*, 46:100–119, 2016. ISSN 18791646. doi: 10.1016/j.jhealeco.2016.01.009. URL http://dx.doi.org/10.1016/j.jhealeco.2016.01.009.

David Lagakos, Benjamin Moll, Nancy Qian, and Todd Schoellman. Life-Cycle Human Capital Accumulation across Countries : Lessons from US Immigrants. *Journal of Human Capital*, 12 (2), 2018.

Daniel Lawver. Measuring Quality Increases in the Medical Sector. Working Paper, 2011.

Frank Levy and Richard J Murnane. With What Skills Are Computers a Complement? *The American Economic Review*, 86(2):258–262, 1996.

Assar Lindbeck and Dennis J. Snower. Harassment , and Involuntary Unemployment: An Insider-Outsider Approach. *The American Economic Review*, 78(1):167–188, 1988.

Lars Ljungqvist and Thomas J Sargent. The European Unemployment Dilemma. *Journal of Political Economy*, 106(3):514–550, 1998.

Lars Ljungqvist and Thomas J Sargent. Two Questions about European Unemployment. *Econometrica*, 76(1):1–29, 2008.

Lilia Maliar, Serguei Maliar, and Fernando Valli. Solving the incomplete markets model with aggregate uncertainty using the Krusell-Smith algorithm. *Journal of Economic Dynamics and Control*, 34(1):42–49, 2010. ISSN 01651889. doi: 10.1016/j.jedc.2009.03.009. URL http://dx.doi.org/10.1016/j.jedc.2009.03.009.

N Gregory Mankiw and Stephen P Zeldes. The consumption of stockholders and nonstockholders. *Journal of Financial Economics*, 29(1):97–112, 1991. ISSN 00301124.

Nolan M McCarty, Keith T Poole, and Howard Rosenthal. *Polarized America: the dance of ideology and unequal riches*. MIT Press Cambridge, Mass., 2016. ISBN 0262134640; 9780262134644. doi: LK-https://worldcat.org/title/69022000.

Daniel McFadden. Conditional Logit Analysis of Qualitative Choice Behavior. 1973.

Atif Mian, Amir Sufi, and Francesco Trebbi. Resolving Debt Overhang : Political Constraints in the Aftermath of Financial Crises. *American Economic Journal: Macroeconomics*, 6(2):1–28, 2014.

MIT MIT Election Data and Science Lab. U.S. Presidential Elections Data 1976–2020, 2017. URL https://doi.org/10.7910/DVN/42MVDX.

S Parker. *A Short History of Medicine*. Penguin Random House, London, 2019.

Donald O. Parsons. The Emergence of Private Job Displacement Insurance in the United States: Severance Pay Plans 1930-1954. *SSRN Electronic Journal*, 2005a. doi: 10.2139/ssrn.872331.

Donald O. Parsons. Private Job Displacement Insurance in the United States, 1954-1979: Expansion

and Innovation. *SSRN Electronic Journal*, 2005b. doi: 10.2139/ssrn.872334.

Donald O. Parsons. Benefit Generosity in Voluntary Severance Plans: The U.S. Experience. *SSRN Electronic Journal*, 2005c. doi: 10.2139/ssrn.877903.

Pew Research Center. Political Polarization in the American Public. Technical Report 1, 2014. URL http://dx.doi.org/10.1016/j.biochi.2015.03.025{%}0Ahttp://dx.doi.org/10.1038/nature10402{%}0Ahttp://dx.doi.org/10.1038/nature21059{%}0Ahttp://journal.stainkudus.ac.id/index.php/equilibrium/article/view/1268/1127{%}0Ahttp://dx.doi.org/10.1038/nrmicro2577{%}0Ahttp://.

Pew Research Center. The Partisan Divide on Political Values Grows Even Wider. Technical Report 1, 2017. URL http://i-lib.ugm.ac.id/jurnal/download.php?dataId=2227{%}0A???{%}0Ahttps://ejournal.unisba.ac.id/index.php/kajian{_}akuntansi/article/view/3307{%}0Ahttp://publicacoes.cardiol.br/portal/ijcs/portugues/2018/v3103/pdf/3103009.pdf{%}0Ahttp://www.scielo.org.co/scielo.ph.

Brooks Pierce. Using the National Compensation Survey to Predict Wage Rates. *Compensation and Working Conditions*, (Winter 1999):8–16, 1999.

Elisabeth Proehl. Approximating Equilibria with ex-post Heterogeneity and Aggregate Risk. 2017.

Vincenzo Quadrini. Entrepreneurship, Saving, and Social Mobility. *Review of Economic Dynamics*, 3(1):1–40, 2000. ISSN 10942025. doi: 10.1006/redy.1999.0077.

Michael Reiter. Solving the incomplete markets model with aggregate uncertainty by backward induction. *Journal of Economic Dynamics and Control*, 34(1):28–35, 2010. ISSN 01651889. doi: 10.1016/j.jedc.2008.11.009. URL http://dx.doi.org/10.1016/j.jedc.2008.11.009.

Donna Retzlaff-Roberts, Cyril F Chang, and Rose M Rubin. Technical efficiency in the use of health care resources: A comparison of OECD countries. *Health Policy*, 69(1):55–72, 2004. ISSN 01688510. doi: 10.1016/j.healthpol.2003.12.002.

Yona Rubinstein and Yoram Weiss. Chapter 1 Post Schooling Wage Growth: Investment, Search and Learning. *Handbook of the Economics of Education*, 1(06):1–67, 2006. ISSN 15740692. doi: 10.1016/S1574-0692(06)01001-4.

Steven Ruggles, J Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B Schroeder, and Matthew Sobek. Integrated public use microdata series: Version 5.0 [Machine-readable database]. *Minneapolis: University of Minnesota*, 42, 2010.

Andreas Schaab. Micro and Macro Uncertainty. 2020.

Kjetil Storesletten, Chris I Telmer, and Amir Yaron. Cyclical Dynamics in Idiosyncratic Labor Market Risk ABI/INFORM Global pg. 695. *The Journal of Political Economy*, 112, 2004. URL

http://pages.stern.nyu.edu/{%}7B{~}{%}7Dcedmond/phd/StoreslettenTelmerYaronJPE2004.pdfhttp://pages.stern.nyu.edu/{~}cedmond/phd/StoreslettenTelmerYaronJPE2004.pdf.

Silvia Strub, Michael Gervin, and Aline Bütikofer. Vergleichende Analyse der Löhne von Frauen und Män- nern anhand der Lohnstrukturerhebungen 1998 bis 2006. Untersuchungen im Rahmen der Evaluation der Wirksamkeit des Gleichstellungsgesetzes. 2008.

Sharon Traiberman. Occupations and import competition: Evidence from Denmark. *American Economic Review*, 109(12):4260–4301, 2019. ISSN 19447981. doi: 10.1257/aer.20161925.

John Voorheis, Nolan McCarty, and Boris Shor. Unequal Incomes, Ideology and Gridlock: How Rising Inequality Increases Political Polarization. 2015.

B Weisbrod. The Health Care Quadrilemma: An Essay on Technological Change, Insurance, Quality of Care, and Cost Containment. *Journal of Economic Literature*, 29:523–552, 1991.

Thomas Winberry. A method for solving and estimating heterogeneous agent macro models. *Quantitative Economics*, 9(3):1123–1151, 2018. ISSN 1759-7323. doi: 10.3982/qe740.

Kai Zhao. Social Security and the Rise in Health Spending : A Macroeconomic Analysis. *Journal of Monetary Economics*, 64:21–37, 2014.

Mengdi Zheng, Xiaoliang Wan, and George Em Karniadakis. Adaptive multi-element polynomial chaos with discrete measure: Algorithms and application to SPDEs. *Applied Numerical Mathematics*, 90:91–110, 2015. ISSN 01689274. doi: 10.1016/j.apnum.2014.11.006. URL http://dx.doi.org/10.1016/j.apnum.2014.11.006.

Ayse İmrohoroğlu and Ayse Imrohoroglu. Cost of Business Cycles with Indivisibilities and Liquidity Constraints. *Journal of Political Economy*, 97(6):1364–1383, 1989.