1-1-2013

Essays on Macroeconomic Effects of Taxation and Health Policies

Soojin Kim
University of Pennsylvania, soojinkim06@gmail.com

Follow this and additional works at: http://repository.upenn.edu/edissertations
Part of the Economics Commons

Recommended Citation
http://repository.upenn.edu/edissertations/651

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/edissertations/651
For more information, please contact libraryrepository@pobox.upenn.edu.
Essays on Macroeconomic Effects of Taxation and Health Policies

Abstract
This dissertation consists of two essays that study the macroeconomic effects of taxation and health policies. In the first chapter, I study the effects of international labor migration on optimal taxation. I develop an open economy model with global financial markets and international labor migration, in which governments engage in an international tax competition. By quantitatively applying the model to the United Kingdom and a set of Continental European countries, I find that strategic interaction between governments, and mobility of labor are important determinants of optimal taxation of capital and labor in open economies. The second chapter (co-authored with Hal Cole and Dirk Krueger) studies the short- and long-run effects of the labor and health insurance market policies in the United States. Motivated by recent legislations aimed at reducing households' exposure to health risks during their working lives, we model the trade-offs between short-run insurance benefits and long-run incentive costs of the social insurance policies. Our quantitative results show that there are non-trivial incentive costs to implementing both labor and health insurance market policies in the long run, leading to a severe deterioration of population health distribution.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Economics

First Advisor
Harold L. Cole

Second Advisor
Dirk Krueger

Subject Categories
Economics

This dissertation is available at ScholarlyCommons: http://repository.upenn.edu/edissertations/651
ESSAYS ON MACROECONOMIC EFFECTS OF
TAXATION AND HEALTH POLICIES

Soojin Kim

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

2013

Supervisor of Dissertation
Harold L. Cole, Professor of Economics

Co-Supervisor of Dissertation
Dirk Krueger, Professor of Economics

Graduate Group Chairperson
George J. Mailath, Professor of Economics

Dissertation Committee:
Harold L. Cole, Professor of Economics
Hanming Fang, Professor of Economics
Dirk Krueger, Professor of Economics
ESSAYS ON MACROECONOMIC EFFECTS OF TAXATION AND HEALTH POLICIES

COPYRIGHT

2013

Soojin Kim
To my parents and my husband

for their unconditional love.
Acknowledgements

I am indebted to my advisors Hal Cole and Dirk Krueger. Their insights and enthusiasm in economics have inspired me and helped shape my research greatly. I am immensely grateful for their generous support and encouragement. I also thank my committee member Hanming Fang, and Ufuk Akcigit and Andrew Shephard for many constructive criticisms and detailed suggestions. Their comments have broadened my view in completing this dissertation. I thank all of Penn faculty with whom I had privileges to share my ideas and discuss economics. I especially thank the late Dave Cass. Although I only had little time with him, his encouragement helped me greatly during the early stages of my life as a graduate student.

I am also thankful to Hongseok Choi, Matt Hoelle, Karam Kang, Fatih Karahan, Kurt Mitman, Suryun Rhee, Cezar Santos, Eun-Young Shim, and Dave Weiss for their friendship. They made McNeil an enjoyable and exciting place to come to.

My deepest thanks go to my family. My parents’ unconditional love and sacrifice are what made this possible. I cannot thank my parents enough for their support and encouragement. My sister has always been with me as a friend, during the difficult times. I am also very grateful to my parents-in-law for their help and consideration. Finally, I thank my husband, Dong Hye, whose support and love was crucial in completing the last years of my graduate studies; and my son, Nathan, who is (still) having a hard time understanding why mommy has to go to school everyday.
This dissertation consists of two essays that study the macroeconomic effects of taxation and health policies. In the first chapter, I study the effects of international labor migration on optimal taxation. I develop an open economy model with global financial markets and international labor migration, in which governments engage in an international tax competition. By quantitatively applying the model to the United Kingdom and a set of Continental European countries, I find that strategic interaction between governments, and mobility of labor are important determinants of optimal taxation of capital and labor in open economies. The second chapter (co-authored with Hal Cole and Dirk Krueger) studies the short- and long-run effects of the labor and health insurance market policies in the United States. Motivated by recent legislations aimed at reducing households’ exposure to health risks during their working lives, we model the trade-offs between short-run insurance benefits and long-run incentive costs of the social insurance policies. Our quantitative results show that there are non-trivial incentive costs to implementing both labor and health insurance market policies in the long run, leading to a severe deterioration of population health distribution.
## Contents

Acknowledgements iv  
Abstract v  
Contents vi  
List of Tables x  
List of Figures xii  

1 The Effects of Labor Migration on Optimal Taxation: An International Tax Competition Analysis 1  
1.1 Introduction 2  
1.2 The Model 10  
1.2.1 Model Description 10  
1.2.2 Competitive Equilibrium of the Economy 14  
1.2.3 Tax Competition and Nash Equilibrium 19  
1.3 Calibration 21  
1.3.1 Descriptive Statistics of the Macroeconomic Data in the UK and CE 21  
1.3.2 Public Policies in the UK and CE 22  
1.3.3 Calibration Strategy 24  
1.3.4 Calibration Results 29  
1.3.5 Computation of Nash Equilibrium 31
1.4 Results ......................................................... 32
  1.4.1 Understanding Household Behavior ......................... 32
  1.4.2 From Status Quo to the Nash Equilibrium .................... 33
  1.4.3 Nash Equilibrium of the Model ............................ 35
  1.4.4 Effects of Labor Mobility on Optimal Taxation and Aggregate Outcomes 39
1.5 Conclusions .................................................. 42
1.6 Tables and Figures ........................................... 44

2 Analyzing the Effects of Insuring Health Risks: On the Tradeoff between Short-Run Insurance Benefits vs. Long-Run Incentive Costs 56
  2.1 Introduction ............................................... 57
    2.1.1 Institutional Background ............................... 59
    2.1.2 Related Literature .................................... 62
  2.2 The Model ................................................... 65
    2.2.1 Endowments and Preferences ............................ 65
    2.2.2 Technology ........................................... 65
    2.2.3 Time Line of Events .................................. 67
    2.2.4 Market Structure without Government .................... 67
    2.2.5 Government Policies .................................. 68
  2.3 Analysis of the Static Model ................................ 74
    2.3.1 Social Planner Problem ............................... 74
    2.3.2 Competitive Equilibrium ............................... 76
    2.3.3 Competitive Equilibrium with a No Prior Condition Law .... 78
    2.3.4 Competitive Equilibrium with a No Wage Discrimination Law ... 80
    2.3.5 Competitive Equilibrium with Both Policies ............... 81
    2.3.6 Summary of the Analysis of the Static Model ............. 82
  2.4 Analysis of the Dynamic Model ................................ 82
    2.4.1 Social Planner Problem ............................... 83
    2.4.2 Competitive Equilibrium without Policy .................. 85
B.5 Computation of the Equilibrium with
   a No-Prior-Conditions Law and/or
   a No-Wage Discrimination Law ........................................... 138
B.6 Details for Data and Calibration ........................................... 139
   B.6.1 Details of the Augmented Model Analysis: Inclusion of the z-shock . 139
   B.6.2 Descriptive Statistics of the PSID Data ........................... 140
   B.6.3 Health Shocks, Distribution of Medical Expenditures, and Discussion
       of Categorization of Health Shocks ................................... 143
   B.6.4 Estimation Results .................................................. 145
   B.6.5 Calibration Results ................................................ 148
B.7 Additional Quantitative Results ........................................... 151
   B.7.1 Model Fit .......................................................... 151
   B.7.2 Policy Implications ................................................. 152
B.8 Sensitivity Analysis ....................................................... 154
   B.8.1 Robustness of Results with Respect to Age and Gender ............ 154
   B.8.2 Benefits of Effort Not Related to Labor Productivity ............... 155

Bibliography 157
List of Tables

1.1 Summary Statistics in the United Kingdom and Continental Europe . . . . 44
1.2 Calibration Targets . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 45
1.3 Calibrated Parameters . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 46
1.4 Model Fit . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 47
1.5 Effects of Changes in the Tax Code – United Kingdom . . . . . . . . . . . 47
1.6 Unilateral Reform of the United Kingdom and Aggregate Outcomes . . . . 48
1.7 Unilateral Reform of Continental Europe and Aggregate Outcomes . . . . 48
1.8 Aggregate Outcomes in the Nash Equilibrium . . . . . . . . . . . . . . . . . 49
1.9 Effects of Labor Mobility on Optimal Taxation in the United Kingdom . . . 50
1.10 Effects of Labor Mobility on Optimal Taxation in the Continental Europe . 50
1.11 Effects of Changes to the Capital Income Tax Rate in the United Kingdom
   under Different Mobility Assumptions . . . . . . . . . . . . . . . . . . . . . . 50
1.12 Effects of Changes to Tax Codes in the United Kingdom and Continental
   Europe at Nash Equilibrium . . . . . . . . . . . . . . . . . . . . . . . . . . . 51
1.13 Effects of Labor Mobility on Aggregate Outcomes . . . . . . . . . . . . . . 51
2.1 Aggregate Welfare Comparisons . . . . . . . . . . . . . . . . . . . . . . . . . 116
2.2 Welfare Comparison in the Dynamic Economy Conditional on Health . . . 116
A.1 Summary Statistics of Continental European Countries . . . . . . . . . . . 119
A.2 Capital Income Tax Rates in Continental European Countries . . . . . . . 120
B.1 Mapping between Data and Model . . . . . . . . . . . . . . . . . . . . . . . 140
B.2 Descriptive Statistics of Key Variables in PSID ........................................ 140
B.3 Health Transition over 6 years ................................................................. 141
B.4 Average Medical Expenditure by Health Shock Categories ....................... 143
B.5 Health Shock Probabilities by Health Status .............................................. 146
B.6 Hours Worked and Labor Income by Health Shock ..................................... 147
B.7 Data Targets ............................................................................................. 148
B.8 Data Targets (continued) ............................................................................ 149
B.9 Calibrated Parameters ................................................................................ 150
B.10 Cross-Subsidy by Health Level under Different Policy Regimes: Young ...... 152
B.11 Cross-Subsidy by Health Level under Different Policy Regimes:Old ........... 152
B.12 Welfare Comparisons in Static Economy .................................................... 153
B.13 Welfare Comparisons in Dynamic Economy .............................................. 153
B.14 Lifetime Welfare Comparisons in the Dynamic Economy Conditional on Type and Health ................................................................. 153
B.15 Moments for the Subsample of Population ............................................... 155
# List of Figures

1.1 Timing of the Model ........................................... 52
1.2 Average Labor Income Tax Rates in the Continental Europe and United Kingdom ........................................... 52
1.3 Human Capital Investment Decisions of Households ........................................... 53
1.4 Differences in the Value of Living in the United Kingdom versus Continental Europe ........................................... 54
1.5 Average Labor Income Tax Rates in the Nash Equilibrium ........................................... 54
1.6 Progressivity Wedge in the *Status Quo* and in the Nash Equilibrium ........................................... 55

2.1 Production Function $F(h, \varepsilon - x)$ ........................................... 112
2.2 Production Function $F(h, \varepsilon - x)$ for fixed $\varepsilon$ ........................................... 112
2.3 Timing of the Model ........................................... 112
2.4 Optimal Health Expenditure and Production ........................................... 113
2.5 Average Effort in Model and Data ........................................... 113
2.6 Consumption Dispersion ........................................... 114
2.7 Subsidy: Excellent ........................................... 114
2.8 Subsidy: Fair ........................................... 114
2.9 Effort ........................................... 114
2.10 Average Health ........................................... 114
2.11 Cutoffs ........................................... 115
2.12 Health Spending ........................................... 115
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.13</td>
<td>Production</td>
<td>115</td>
</tr>
<tr>
<td>2.14</td>
<td>Consumption</td>
<td>115</td>
</tr>
<tr>
<td>A.1</td>
<td>Capital Income Tax Rates</td>
<td>118</td>
</tr>
<tr>
<td>A.2</td>
<td>Labor Income Tax Rates</td>
<td>118</td>
</tr>
<tr>
<td>A.3</td>
<td>Trends in Top Labor Income Tax Rates in OECD Countries</td>
<td>118</td>
</tr>
<tr>
<td>A.4</td>
<td>Average Labor Income Tax Rates in Continental European countries</td>
<td>120</td>
</tr>
<tr>
<td>B.1</td>
<td>Density of Light Activity</td>
<td>142</td>
</tr>
<tr>
<td>B.2</td>
<td>Density of Heavy Activity</td>
<td>142</td>
</tr>
<tr>
<td>B.3</td>
<td>Light Activity</td>
<td>142</td>
</tr>
<tr>
<td>B.4</td>
<td>Heavy Activity</td>
<td>142</td>
</tr>
<tr>
<td>B.5</td>
<td>Average Medical Expenditure Distribution</td>
<td>144</td>
</tr>
<tr>
<td>B.6</td>
<td>Average Expenditure with Health Shock</td>
<td>144</td>
</tr>
<tr>
<td>B.7</td>
<td>Average Expenditure w/ z-shock</td>
<td>144</td>
</tr>
<tr>
<td>B.8</td>
<td>Average Expenditure w/ ε-shock</td>
<td>144</td>
</tr>
<tr>
<td>B.9</td>
<td>GMM: Transition of Excellent Health</td>
<td>146</td>
</tr>
<tr>
<td>B.10</td>
<td>GMM: Transition of Very Good Health</td>
<td>146</td>
</tr>
<tr>
<td>B.11</td>
<td>GMM: Transition of Good Health</td>
<td>146</td>
</tr>
<tr>
<td>B.12</td>
<td>GMM: Transition of Fair Health</td>
<td>146</td>
</tr>
<tr>
<td>B.13</td>
<td>Fair</td>
<td>151</td>
</tr>
<tr>
<td>B.14</td>
<td>Good</td>
<td>151</td>
</tr>
<tr>
<td>B.15</td>
<td>Very Good</td>
<td>151</td>
</tr>
<tr>
<td>B.16</td>
<td>Excellent</td>
<td>151</td>
</tr>
</tbody>
</table>
Chapter 1

The Effects of Labor Migration on Optimal Taxation: An International Tax Competition Analysis

Summary

This paper develops a two-country, open-economy model with labor mobility and a global financial market to study optimal taxation. Governments in this economy engage in a tax competition in which they choose a potentially progressive labor income tax code and a capital income tax rate, taking as given the tax policies in the other country. A quantitative application of the model to the United Kingdom (UK) and Continental European countries (CE) shows that factor mobility and competition between governments are crucial in the design of optimal policies. First, in the Nash equilibrium, the UK and CE use progressive labor income taxes as dominant sources of tax revenue. Second, in order to isolate the effects of competition, I compare the tax rates that the UK (CE) sets, assuming that the other country does not respond to its policy changes, to the Nash equilibrium taxes. I find that capital income tax rates are higher with competition. Third, I study the importance of labor migration on optimal taxation in open economies, by comparing the optimal taxation in an economy with only capital mobility to that of the benchmark economy. In this model, incorporating labor migration leads to a divergence in the optimal tax system: Unlike in
an economy with only capital mobility, where both countries use similar capital income tax rates, the optimal capital income tax rate in the UK is lower than that in the CE when both capital and labor are mobile. This is due to the differences in productivity between the two countries. In the calibrated economy, the UK, whose productivity is higher than that of the CE, attracts more capital and labor through migration (higher population). Thus, given a relatively small population (labor), the welfare-maximizing level of capital in the CE is smaller: In the Nash equilibrium, the CE uses a higher capital income tax rate than the UK does. The steady-state welfare gain from implementing the Nash equilibrium policies is about 11 percent of consumption of the status quo economy.

1.1 Introduction

In September of 2012, French president François Hollande introduced his plans for a 75 percent income tax on “the rich.” The announcement immediately incited concerns about the exodus of talented workers from France. These concerns were not unfounded: While awaiting the French tax reform earlier in June, David Cameron, the Prime Minister of the United Kingdom, had offered to “roll out the red carpet and welcome more French businesses to Britain.”

The controversy caused by this French legislation (and Britain’s ready response to it) demonstrates the recent policy debates taking place in many countries. These debates have centered on the issues surrounding the optimal degree of income redistribution through taxation, and the competition between governments in attracting scarce resources – capital and high-skilled labor. Although a progressive tax system provides income insurance to households, it can lead to a flight of resources in open economies. Thus, governments may engage in tax competitions in an effort to attract the factors of production. This paper

---

1On December 29, 2012, however, France’s Constitutional Council ruled that the policy was unconstitutional. In response to the ruling, Prime Minister Jean-Marc Ayrault said that the government “will present a new proposal in line with the principles laid down by the Constitutional Court.”

2David Cameron made this comment at the G20 summit in Mexico on June 19, 2012. (Murphy, Richard. “In wooing French tax exiles, Cameron makes a mockery of democracy,” The Guardian, 19 June 2012.)

3The distortionary effects of implementing progressive tax systems are also present in open economies (as in closed economies): Households have less incentives to work and to accumulate human capital.
studies the optimal progressivity of labor income tax codes and capital income tax rates in economies confronting international factor mobility and competition between governments.

While economists have long studied the effects of capital or labor mobility on taxation in an international tax competition framework, a joint analysis of the mobility of both factors on optimal taxation has not been conducted. However, since capital and labor are complementary in production, considering both factors is important: A tax system that induces accumulation of one factor increases the marginal product of the other. Thus, optimal policies should weigh the tradeoffs of taxing capital and labor. Moreover, human capital can differ in its quality (skilled and unskilled labor, for example), and the returns to investment are heterogeneous across households. Therefore, when countries differ in the relative efficiency of unskilled and skilled labor, the optimal progressivity of labor income tax codes are affected by the possibility of labor mobility. Thus, the progressivity of the labor income tax code is an important component in the design of government policies in open economies.

In order to address these issues, I build a two-country, open-economy model with international labor mobility and a global financial market. Countries in the model differ in their production technologies – the relative efficiency of unskilled and skilled labor – and human capital production technologies.

Households, who reside in one of the countries, are altruistic toward their descendants and live for two periods, one as a child, and the other as a parent; they are only able to make economic decisions during the latter period. The households (parents) are heterogeneous in several dimensions: country of residence, skill level, the ability of their child, assets, and idiosyncratic labor productivity. Households can either be skilled or unskilled, with market wages that differ for each. The ability of child follows a first-order Markov process and it increases the probability that the child becomes a skilled worker (conditional on human capital investment). Moreover, the higher the ability of child, the higher the return

---

4I discuss these papers in the related literature below.
5Galor (1992) compares the welfare implications of allowing for labor or capital mobility, but not both.
6When I calibrate the model, I define “skilled” households as those who have graduated from college.
to human capital investment is. The parents make consumption and leisure decisions, and choose to invest in their offspring by leaving physical assets and/or investing in their human capital. Although households can invest and trade international bonds, I assume that an explicit insurance market for idiosyncratic labor risk does not exist. Modeling heterogeneity across households allows the model to capture any differential responses of households to tax systems. Moreover, since financial markets are incomplete, the governments can provide consumption insurance by using a progressive labor income tax system to redistribute income across households. At the end of the parental life, after observing the migration cost shock, parents make migration decisions.

Governments in this economy maximize the steady-state welfare of households who are physically residing in their countries, by choosing a labor income tax code (allowing for progressivity) and a flat capital income tax rate. They finance a fixed amount of revenue per capita, with which they provide public education and transfer back the rest of the proceeds to households in a lump-sum fashion. Since I model the endogenous accumulation of human capital and the impact of tax policies on skill distribution, public education is an important part of government policy in this analysis.

I also take into consideration, the interaction between governments in choosing their tax systems: They engage in a tax competition game. Governments maximize the steady-state welfare of households, taking as given the other country's tax system. They are only allowed to change their tax systems once (one-shot tax competition game), and I

---

7 As will be more clear in the following sections, in this paper, the financial market structure I assume (which is the approach taken by Mendoza and Tesar in their 1998 and 2005 papers) does not strictly allow the households to invest in both countries. However, there is a financial market for international bonds. This structure allows the tax rates to have reallocative effect of capital in both countries in a simple way, which is the key element for the analysis in this paper. From hereon, I use the term capital mobility to indicate that there is a global financial market.

8 The revenue requirement is chosen as the level of revenue per capita collected in the calibrated status quo economy. I use a per capita revenue requirement since population sizes are endogenously determined by mobility of labor. Using a fixed revenue requirement (regardless of the population size) creates incentives for the governments to maximize population in order to minimize the tax burden.

9 In this paper, I take as given the public education policies in both countries, and governments do not choose public education spending as a policy tool. However, since progressive labor income taxation and public education spending can interact with each other, it will be an interesting policy study to allow for an endogenous choice of education policies for the governments. I discuss this issue briefly in the conclusion.
assume that they can commit to policies, by for example, an institutional agreement.\footnote{Klein, Quadrini, and Rios-Rull (2005) study time-consistent taxation in open economies.} This tax competition framework I use builds on the one developed by Mendoza and Tesar (2005); the focus of their paper is the role of capital mobility (assuming that labor is immobile), which they analyze in a representative agent model. Incorporating heterogeneity and allowing for a progressive labor income tax code makes the model suitable for the analysis of the differential welfare impacts across households, which governments take into account in designing optimal policies.

In the quantitative analysis, I apply the model to the United Kingdom (denoted as the “UK” hereafter) and Continental European countries (an aggregate of France, Germany, Italy, Spain, and Sweden, denoted as the “CE” hereafter). I calibrate the parameters to match allocations with the observed intergenerational linkages of human capital, aggregate outcomes, and migration statistics, taking the \textit{status quo} tax systems and public education spending of the UK and CE from data as given. In the calibrated economy, the UK has higher overall efficiency, and higher relative efficiency of skilled labor in production. On the other hand, in the CE, the return to human capital investment is higher than that in the UK. These calibration results are necessary to match higher skill premium and higher percentage of college graduates (which I define as “skilled” in the data) in the UK compared to the CE. Moreover, the parameters for human capital production affect the skill distributions and education spending in the UK and CE. Given the CE’s relatively low productive efficiency of skilled labor and low skill premium, higher return to human capital investment is necessary to match its skill distribution.

In the Nash equilibrium, the UK subsidizes capital income by 32 percent, and taxes labor income at 52 percent (on average). On the other hand, the CE levies a 6 percent tax on capital income, and a 50 percent tax on labor income (on average). Both countries use progressive labor income taxes, while the progressivity is higher in the UK.

In order to isolate the importance of competition and labor mobility on optimal determination of the tax systems, I conduct two analyses. First, to investigate the effects of
competition, I let the UK (CE) unilaterally reform its tax system, taking as a given the
status quo tax system of the other government in the economy. In the unilateral reform, I
find that the UK and CE act more aggressively in setting their policies: The capital income
tax rates are lower in the unilateral reform case. This is because the benefit of a lower
capital income tax rate is higher when the other country’s capital income tax rate is higher.

Second, to analyze the effects of labor mobility on optimal taxation, I compare the
optimal tax systems in an open economy with only capital mobility, and then in an economy
with both capital and labor mobility. The Nash equilibrium capital income tax rates are
similar in the economy where only capital is mobile. However, as countries open up to labor
mobility, the capital income tax rates diverge: The UK lowers its capital income tax rate
further, while the CE does the reverse. Since the UK has higher efficiency in production, the
benefit of using a lower capital income tax rate is higher: With higher capital, the marginal
product of labor in the UK increases more than it would in the CE. This in turn generates
a larger population in the UK. On the other hand, for the CE, whose population is smaller,
the welfare-maximizing level of capital is relatively lower. Moreover, when capital income
tax is lowered, the increase in labor income tax necessary to balance the budget is too high
and depresses the working hours of households. Thus, the CE finds it optimal to use a
relatively higher capital income tax rate than the UK does.

Finally, I find that implementing Nash equilibrium policies leads to welfare gain of 11 and
13 percent of consumption in the status quo economy in the UK and CE, respectively. Under
the Nash equilibrium policies, after-tax return to capital increases since capital income tax
rates are lower. Both countries have lower fraction of the skilled workforce, and households
enjoy more leisure (work less). Unskilled households benefit more from the reform than the
skilled households do, by about twofold.

**Related Literature** This paper is located at the intersection of several strands of litera-
ture, including that on international tax competition, optimal taxation of capital and labor,
and the effects of labor mobility on macroeconomic outcomes.
There is a wide literature that studies international tax competition in capital and labor.\textsuperscript{11} With regard to the capital income tax dimension, Gordon (1986) and Razin and Sadka (1991) theoretically study two different types of capital income taxation: source versus residence-based taxation, as well as taxation of different kinds of capital. As for labor income tax competition, Mirrlees (1982) and Bhagwati and Hamada (1982) are seminal papers that focus on the taxation of foreign and domestic labor income in less-developed countries. Razin and Sadka (2010, 2011) also conduct analytical studies of labor income tax competition, but take the flow of migrants as exogenous.\textsuperscript{12} These papers offer insightful theoretical analysis. However, they do not endogenize the migration decisions of households, which I model, and lack the quantitative dimension that I provide in this paper.

The paper that most closely relates to my work is a 2005 study by Mendoza and Tesar. In it, the authors study a tax competition game in a dynamic general equilibrium framework. They only allow for capital mobility, and the capital income tax rate is chosen to maximize welfare in the balanced growth path, while the labor income tax (or a consumption tax) is a tool used to ensure fiscal solvency. They find that in the UK and Continental Europe (which they define as France, Germany, and Italy), Nash equilibrium tax rates are consistent with observed tax rates, when labor income tax is used to resolve fiscal solvency.\textsuperscript{13} However, when consumption taxes are used to maintain fiscal solvency, the optimal policy is to subsidize capital income. Unlike Mendoza and Tesar’s analysis, this paper incorporates labor mobility, heterogeneity in households, and a possibility of using progressive labor income taxes to allow for a richer analysis of optimal policies in a tax competition framework.

Another related paper is Armenter and Ortega (2010), which quantitatively studies the effects of labor mobility on the redistributive ability of regional governments in the United States. Rather than modeling non-cooperative tax competition, the authors find redistributive policies that are incentive-compatible for skilled workers to stay in a region.

\textsuperscript{11}There are also tax competition papers that deal with tax on consumption goods; among them are Chari and Kehoe (1990) and Kanbur and Keen (1993).
\textsuperscript{12}The model in Tiebout (1956) also studies labor mobility and its impact on the provision of public goods in multiple communities.
\textsuperscript{13}In their calibrated model, the implied Frisch elasticity of labor supply is around 1.7, whereas in my model, the Frisch elasticity is around 0.3. Thus, distortions from taxing labor are higher in their model.
The analysis shows that the competition between regions brings about a convergence in tax rates.

The topics studied in this paper also intersect with the dynamic optimal taxation literature. Starting with Judd (1985) and Chamley (1986), dynamic optimal taxation has been widely studied, both theoretically and quantitatively. Most of these studies have been conducted in a closed economy, with complete markets and a representative household framework. My model, on the other hand, studies taxation in large open economies, each of which has incomplete asset markets.

Aiyagari (1995) studies optimal capital taxation in an incomplete markets model. He finds that governments can improve the welfare of households by taxing capital income, which induces households to decrease their precautionary savings. This result is driven by the fact that the government endogenously chooses its expenditure level. A more recent study by Davila et al. (2012) examines constrained efficiency in incomplete markets, which has implications for capital income taxation. They find that if the consumption-poor in a country have little wealth (but relatively high labor income), social planners can improve overall welfare by increasing capital (which results in higher wages) – an argument for lower (or negative) capital income tax. Also related is a study of the dynamic optimal taxation in large open economies, analyzed by Gross in a 2012 paper. He finds that when large open economies use a territorial tax system, a zero capital income tax rate in the long run is still optimal. Among others, Conesa et al. (2009) quantitatively studies optimal taxation allowing for a progressive labor income tax system. They find that positive capital income tax is optimal, in an overlapping generations model. This paper complements the existing literature by quantitatively studying the optimal taxation of both capital and labor in large open economies with incomplete markets.

Moreover, this paper is related to the studies of optimal progressivity of the labor in-

---

14 Atkeson et al. (1999) study the robustness of the zero capital income tax result and conclude that it also applies to open economies.
15 Ha and Sibert (1997) also study strategic capital taxation in large open economies. They distinguish between corporate and savings taxes, and find that capital importers set positive tax rates, while exporters subsidize capital.
come tax code and its impact on human capital accumulation. Recent papers by Erosa and Koreshkova (2007) and Guvenen et. al (2011) relate the effects of progressive tax systems on household behaviors and aggregate outcomes, such as inequality and intergenerational persistence in earnings. Erosa and Koreshkova formulate a human capital production function that requires both time and goods input and quantitatively analyze the effects of progressive taxation on intergenerational persistence and inequality. While Erosa and Koreshkova focus on the intergenerational inequality, Guvenen et. al study the life-cycle effects of taxation and compare its implications on inequality using cross-country data. The papers in the literature focus on closed economies; here, I consider how the possibility of migration affects the incentives for human capital accumulation.

Lastly, this paper both relates to and expands on existing studies of the effects of labor mobility on macroeconomic outcomes. Two studies by Klein and Ventura (2007, 2009) use a general equilibrium model to quantify the implications for output and welfare between countries of different total factor productivities when labor is mobile. They find large gains from removing barriers to labor mobility. On a similar note, Benhabib and Jovanovic (2011) study optimal migration in a global perspective. They solve the social planning problem of maximizing the welfare of rich and poor country residents, and show that migration should be higher to achieve optimum welfare. However, these papers do not consider the taxes nor human capital accumulation that I endogenize in this paper. As this paper demonstrates, enriching the analysis to allow for the effects on human capital yields important implications for the impact of taxes on labor mobility and skill distribution in both countries.

The paper proceeds as follows. In the next section, I describe the model in detail. The model description comprises defining the competitive equilibrium of the open-economy model and the Nash equilibrium of the tax competition game between the governments. Following the model, I present the calibration strategies and results in section 1.3. In section 1.4, quantitative results are discussed in several steps. I first analyze key household

---

16 There are also papers that relate immigration to resolving fiscal imbalances in the US. Storesletten (2000) uses a general equilibrium overlapping generations model to capture the fiscal impact of immigration and claims, that immigration can be used as a tool to solve fiscal imbalance in the US. Auerbach and Oreopoulos (1999), on the other hand, finds a very small fiscal impact from immigration.
behaviors: the human capital investment and migration decisions. Then, in section 1.4.2, I discuss the effects of changes in the tax code on aggregate outcomes, which provides insights into the results of the unilateral reforms of the UK and CE governments. In 1.4.3, I present the Nash equilibrium policies and discuss the effects of competition and aggregate outcomes in the Nash equilibrium. In section 1.4.4, I compare the optimal taxation results with only capital mobility and with both capital and labor mobility to study the effects of labor mobility in isolation. Finally, I conclude in section 1.5.

1.2 The Model

In this section, I describe the model and define the competitive equilibrium.17

1.2.1 Model Description

I consider a two-country, large open-economy model where both physical and human capital are endogenously accumulated and mobile. Time is discrete and there is a fixed measure 1 of total households (the sum of households residing in both countries is constant). Agents in the economy live for two periods, one as a child and one as a parent, but only make decisions as a parent.

Endowments and Preferences. Households (parents) are endowed with one unit of time which they can divide between leisure and work.

Each household (in country $i$) is heterogeneous in his child’s ability ($z$), human capital ($\theta$), capital goods ($k$), international bonds ($b$), and idiosyncratic labor productivity shock ($\varepsilon$). Ability and human capital operate as described in the previous section. Each household can invest in domestic capital markets by amount $k$ which yields return of $r^i$, and can also hold international bonds $b$ whose gross return is $R$. The fact that households can only invest in domestic capital markets (but not in foreign capital market) is based on assumptions about the financial market structure, which I discuss in more detail in the next subsection.

17I formulate the dynamic problem recursively and conduct a steady state analysis. Thus, I abstract from time subscripts in the notations below.
Domestic capital investment, $k$, and international bonds, $b$, compose the wealth portfolio bequeathed to the households by their parents. Consequently, parents have two ways of leaving their wealth to their offspring: they can either leave assets (in capital or bonds) or invest in the human capital of their child to increase the child’s labor earnings.

The labor productivity shock $\varepsilon$ is independently and identically distributed (i.i.d.) across time as well as across households and has distribution $\pi_\varepsilon(\varepsilon)$. In each period, a household with skill level $\theta$ and i.i.d. labor productivity shock $\varepsilon$, who works $l$ hours, earns $w^i(\theta)l\varepsilon$ in country $i$.

When households decide to migrate to the other country, they incur a utility cost of moving $h$. The migration decision is made after the realization of the moving cost shock, which has a probability distribution function (pdf) of $f(h)$. In the model, I interpret $h$ as incorporating both moving cost and location preference, and thus use location preference and moving cost interchangeably to denote $h$ hereafter.

Preferences of households are assumed to be represented by

$$u(c, l) + \beta \mathbb{E}V'$$

where $c$ and $l$ are consumption and labor hours, respectively. The term $\mathbb{E}V'$ represents the expected utility of the offspring, which is weighted by the altruism factor $\beta$. The expectation is taken with respect to the ability of the grandchild ($z'$), human capital ($\theta'$) and labor productivity shock ($\varepsilon'$) of the offspring, and the moving cost ($h$).

**Human Capital Technology.** Each household is born with his child’s ability level $z \in [\bar{z}, \tilde{z}]$. The ability level is persistent across generations and evolves stochastically according to $\pi(z'|z)$.

Human capital investment $\tilde{x}$\textsuperscript{18} increases the probability of becoming a skilled worker.

\textsuperscript{18}In the current formulation of the model, human capital is accumulated through goods input (buying books, for example). Erosa and Koreshkova (2007), on the other hand, models human capital investment as human capital services purchased at the market wage rate, which makes parental decisions less responsive to changes in taxation. If I were to follow their modeling choices, I would need to take a stand on whether human capital services are provided by skilled or unskilled workers (and the human capital investment would enter the budget constraints as $w\tilde{x}$).
Since I assume that public and private spending on education are perfect substitutes, \( \bar{x} \) refers to the total education spending (public and private). Probabilities of becoming a skilled or unskilled worker are represented by \( Q(s|z, \bar{x}) \) and \( Q(u|z, \bar{x}) \), where \( Q(\cdot|z, \bar{x}) \) is continuous in \( z \) and \( \bar{x} \). Moreover, \( Q_{x \bar{x}}(s|z, \bar{x}) \) is positive, which implies that the return to human capital investment is higher for children with higher ability.

**Production.** Production requires three inputs: physical capital, unskilled labor, and skilled labor. I denote by \( K \), capital, and \( U \) and \( S \) are the total amount of labor provided by unskilled and skilled households, respectively. The production functions represented by \( F(K, U, S) \) and capital depreciates at rate \( \delta \).

**Government Policy.** The sources of government revenue are proceeds from labor and capital income taxes. With the proceeds, the governments provide public education (\( E^i \)) and transfer back the rest of the proceeds in a lump-sum fashion (\( TR^i \)). I restrict that the transfers must be greater than or equal to zero (\( TR^i \geq 0 \)). Since with labor mobility, population size is endogenously determined, I denote by \( e^i \) and \( tr^i \), per capita public education and lump-sum transfer provided by the governments. I do not allow the governments to issue debt, and thus, they must satisfy a period-by-period budget constraint.

The labor income tax code is denoted by \( \tau(y; \bar{y}) \): An individual who earns \( y \) in an economy where average labor earning is \( \bar{y} \) faces a tax rate of \( \tau(y; \bar{y}) \). The purpose of this function is to allow for the progressive labor income tax code, present in many countries. Though I explicitly model progressive labor income taxation, for capital income taxes, I assume a flat tax rate of \( \tau_k \).\(^{19}\) Moreover, I also assume that governments levy capital income tax on returns from the domestic capital market (\( k \)), but not on returns from international bonds (\( b \)). In most countries, returns from participating in domestic and international capital markets are treated the same. However, due to no-arbitrage conditions, tax rates on international bonds must be the same in equilibrium. This would imply that

\(^{19}\)This is the approach commonly taken in the dynamic optimal taxation literature. In a quantitative exercise, Conesa, et al. (2009), among others, also allow for a progressive labor income tax code and a flat capital income tax rate.
different tax rates on capital income cannot be supported. One way of circumventing this problem is to assume that countries levy different taxes on international bond returns than they do on domestic capital income, with the restriction that tax rates on bonds are same across countries. Under the current formulation, I assume that tax rates on interest from international bonds are zero in both countries.\footnote{Mendoza and Tesar (1998) also make the same assumption (zero taxes on international bonds) in their benchmark model. In the sensitivity analysis, they conduct an experiment where they levy taxes on international bonds in both countries at the same level, but lower the capital income tax rate in the U.S. This leads to important welfare consequences due to the higher cost of borrowing from abroad for the U.S. Taxes on bonds decrease welfare in the U.S. when it decreases capital income taxes, although without the tax on bonds (benchmark), welfare increases.}

These are the restrictions on the kinds of tax instruments that governments use that I impose. Therefore, in the tax competition game, governments maximize within the tax functions that I describe, taking per capita public education expenditure $e^i$ and transfers $tr^i$ as given.

**Market Structure.** Households cannot insure against idiosyncratic labor income risk by trading insurance contracts. However, they can participate in financial markets.

Financial market structures in open-economy models can take several forms, and each has important implications for the model outcomes. If households have costless access to both domestic and foreign capital markets and pay taxes based on residence, the pre- and post-tax returns to capital in the two countries must be equalized in equilibrium. This is in marked contrast to observed differences in capital income tax rates across countries. In order to allow for the equilibrium capital income tax rates that differ across countries and for the existence of a global financial market, I follow Mendoza and Tesar (1998) who assume that households can only trade international bonds and invest in domestic capital markets. Under this market structure, after-tax returns to capital is equalized in equilibrium.\footnote{As pointed out in Gross (2012), the equilibrium condition of this financial market structure is the same as the one in which households can invest in both countries and countries levy source-based capital income taxes.} Therefore, even though capital cannot be invested in both countries, there will be reallocative effects of capital in response to changes in tax rates, not only in the home country, but also in the foreign country; this is the key mechanism necessary for the
analysis in this paper.

**Timing.** The model timing is as the following (summarized in Figure 1.1).

A household enters its current period with the state variables \((i, z, \theta, k, b, \varepsilon)\). Each individual (household) residing in country \(i\) becomes a parent with the child’s ability \(z\), his own human capital \(\theta\) (skilled or unskilled), capital goods \(k\) and international bonds \(b\) bequeathed from his parents, and productivity shock \(\varepsilon\). Households provide labor \((l)\), consume \((c)\), invest in human capital \((x)\), purchase private capital \((k')\), and trade international bonds \((b')\).

After the ability of the grandchild \((z')\), skill \((\theta')\) and labor productivity shock \((\varepsilon')\) of the child, and location preference shock \((h)\) are realized, households make migration choices.

According to the model timing, parents start their lives in one country, at which point their skill level has been determined (e.g., whether or not they have received college education) and they have a child; this would roughly correspond to the ages between 30 and 40. A recent report from the UN (Population Division, DESA, 2011), documents that the median age of migrants in developed countries is 42. Since the survey was taken in the migrant’s country, it implies that the migration decision was made before the age of 42. Thus, the data can be taken to suggest that the timing assumption is not unreasonable.

1.2.2 Competitive Equilibrium of the Economy

This section contains a detailed discussion of the household problem, the firm problem, government budget constraints, and the definition of the competitive equilibrium (given an exogenous tax system in both countries).

**Household Problem.** Each household starts his parental life in country \(i\) with offspring’s ability level \((z)\), human capital \((\theta)\), assets bequeathed from parents \((a)\), and the i.i.d. labor productivity shock \((\varepsilon)\). As mentioned above, households can bequeath assets as domestic capital \((k)\) or international bonds \((b)\). However, due to the existence of a global financial market, the returns to assets must be equalized in equilibrium. Therefore, the sufficient
statistic for the household problem is $a$, and the choice variable, $a'$. The household problem can then be stated as

$$V^i(z, \theta, a, \varepsilon) = \max_{l, c, x, a'} u(c, l)$$

$$+ \beta \sum_{z', \theta', \varepsilon', h'} \Xi(z, x, z', \theta', \varepsilon', h') \left\{ \mathbb{E}_h \max \left[ \begin{array}{c}
V^i(z', \theta', a', \varepsilon'), \\
V^j(z', \theta', a', \varepsilon') - h
\end{array} \right] \right\}$$

s.t. $c + x + a'^h = (1 - \tau(y))y + R + TR^i$,

$$y = w^i(\theta)l\varepsilon,$$

$$a' \geq 0, \quad c \geq 0, \quad 0 \leq l \leq 1$$

where $\Xi(z, x, z', \theta', \varepsilon', h') \equiv \pi(z'|z)Q(\theta'|z', x + E^i)\pi_\varepsilon(\varepsilon')f(h)$.

Each household (parents) chooses the optimal level of labor hours ($l$), consumption ($c$), human capital investment ($x$), and total bequest ($a'$) that maximizes his utility. Utility is derived from consumption and leisure today and the expected value of the child’s utility in the future. The term $\Xi(z, x, z', \theta', \varepsilon', h')$ reflects the future expectations about realization of abilities ($z'$), skill level ($\theta'$), i.i.d. labor productivity shock ($\varepsilon'$), and migration cost shock ($h$). The ability of the grandchild evolves stochastically, conditional on the ability of the child through $\pi(z'|z)$, and the skill level of the child is a function of the child’s ability ($z'$) and the sum of human capital investment by parents ($x$) and public education expenditure ($E^i$). I assume that private and public investment are perfect substitutes, and therefore, the probability of becoming a skilled worker is $Q(s|z, x + E^i)$. Moreover, the i.i.d. productivity shock ($\varepsilon'$) is taken into account for evaluating the future value function, as well as moving cost ($h$), which has a pdf of $f(h)$.

The second part of the household objective function (1.1) includes the max operator that reflects migration decisions of the households. Although all economic decisions of the parents are made prior to the decision of migration, households only migrate after all realizations ($\{z', \theta', a', \varepsilon', h\}$) have been made. I also assume that the migrants do not face
any depreciation in their productivity, following Bell (1997), who does not find any wage
differential between immigrants and natives in the long run.

Households maximize their lifetime utility subject to constraints (1.2)–(1.4). The budget
constraint of the household is represented by (1.2). The left-hand side represents possible
expenditures, consumption, human capital investment, and physical capital investment (be-
quests), and the right-hand side represents sources of income, after-tax labor income, assets,
and transfers from the government. Equation (1.3) specifies labor income of the households,
and equation (1.4) states that parents cannot borrow against the future earnings of their
child (they can only leave zero or positive levels of assets).

Because households make human capital investment choices prior to moving, having
the option to migrate can affect their human capital investment decisions. For example, in
high-tax countries, households might want to over-invest in human capital, relative to what
is optimal conditional on not migrating with probability one, in expectation of moving to
the low-tax country, whereas in low-tax countries, households might have less incentives to
do so, compared to a world without labor mobility, since the skill premium can decrease
due to an increased supply of skilled workers through immigration.

From here after, for simplicity, I denote as \( s \), the vector of state variables \( (i, z, \theta, a, \varepsilon) \),
and \( \Phi(s) \), the measure of households of type \( s \).

**Firm Problem.** Firms in country \( i \) produce goods using unskilled and skilled labor and
capital, given market wage rates and rental rate of capital

\[
\max_{K^i, U^i, S^i} \quad F(K^i, U^i, S^i) - (r^i + \delta)K - w^i_u U^i - w^i_s S^i, \tag{1.5}
\]

where \( \delta \) is the depreciation rate of capital.
This implies that in equilibrium, factor prices satisfy

\[ r^i = F_K(K^i, U^i, S^i) - \delta \]  
\[ w^i_u = F_U(K^i, U^i, S^i) \]  
\[ w^i_s = F_S(K^i, U^i, S^i). \]

**Government Budget Balance.** Governments use labor and capital income tax to provide per capita public education $e^i$ and transfer back $tr^i$ to households in a lump-sum fashion. They balance their budget every period, and thus the government budget balance condition reads as:

\[ e^i \int_s d\Phi^i(s) + tr^i \int_s d\Phi^i(s) = \int_s y^i(s)\tau^i(y^i(s))d\Phi^i(s) + r^i\tau_k K^i, \quad tr^i \geq 0. \]  

Governments cannot levy lump sum taxes, i.e., $tr^i \geq 0$. Moreover, since governments tax capital at a flat rate $\tau_k$, the sufficient statistic for obtaining the total tax proceeds from capital income is total private capital ($K^i$) in each country.

**Market Clearing.** Market clearing conditions for the markets for goods and bonds are:

\[ \sum_{i=1,2} \left( F(K^i, U^i, S^i) + (1 - \delta)K^i \right) = \sum_{i=1,2} \left( C^i + X^i + K^i + E^i \right) \]  
\[ B^1 + B^2 = 0, \]
where

$$C^i = \int_s c^i(s) d\Phi^i(s), \quad (1.12)$$

$$X^i = \int_s x^i(s) d\Phi^i(s), \quad (1.13)$$

$$A^i = \int_s a^i d\Phi^i(s)$$

$$= K^i + B^i, \quad (1.14)$$

$$U^i = \int_{z,a,\varepsilon} l^i(z,u,a,\varepsilon) d\Phi^i(z,u,a,\varepsilon), \quad (1.15)$$

$$S^i = \int_{z,a,\varepsilon} l^i(z,s,a,\varepsilon) d\Phi^i(z,s,a,\varepsilon), \quad (1.16)$$

and $K^i$ and $B^i$ are aggregate capital and bonds in country $i$. As I noted in the previous section, since the sufficient statistic in the household problem is the amount of assets bequeathed ($a$), but not its allocation in private capital ($k$) and bonds ($b$), there is an indeterminacy of the capital and bond position at the household level. However, by strict concavity of the production function, for a given equilibrium price of capital $r_i$, there is a unique level of aggregate capital such that the equilibrium price condition (1.6) is satisfied. This condition pins down the level of aggregate capital ($K^i$) in both economies, and from $A^i$, I back out aggregate bond positions using $B^i = A^i - K^i$.

Now, I define a stationary competitive equilibrium of this economy, given government policy in country $i$ \{e$^i$, tr$^i$\}, and the set of government policies, \{τ$^1$, τ$^2$\}.

**Definition 1.** Given an exogenous level of public education expenditure per capita \{e$^1$, e$^2$\}, transfers per capita \{tr$^1$, tr$^2$\}, and tax policies \{τ$^1$, τ$^2$\}, a stationary competitive equilibrium consists of value functions for households, $V^i$; policy functions for households, \{c$^i$, l$^i$, x$^i$, a$^{' i}$\}; production plans for the firms, \{K$^i$, U$^i$, S$^i$\}; prices, \{R, r$^i$, w$^u_i$, w$^s_i$\}; and measures, $\Phi^i$, in both countries $i \in \{1, 2\}$, such that:

(i) households maximize given prices and policies: value functions solve (1.1) subject to constraints (1.2)–(1.4) with \{c$^i$, l$^i$, x$^i$, a$^{' i}$\} as associated policy functions;
(ii) firms maximize given prices and policies: prices \( \{r^i, w^i_L, w^i_s\} \) satisfy (1.6)–(1.8);

(iii) government budget balance condition (1.9) is satisfied;

(iv) market clearing conditions (1.10) and (1.11) are satisfied with the associated aggregate variables defined in (1.12)–(1.16); and

(v) Law of Motion: \( \Phi^i \) is derived from the policy functions of households and probability distribution for i.i.d. labor productivity shocks.

Thus far, I have defined the stationary competitive equilibrium of the economy for given tax systems in both countries. In the next section, I will describe how I model tax competition between the two countries and define the Nash equilibrium of the tax competition game between the governments.

1.2.3 Tax Competition and Nash Equilibrium

Governments choose income tax function and a capital income tax rate that maximizes the steady-state welfare of households at the stationary equilibrium supported by the tax system, taking as given the other country’s tax rates. They decide on their time-invariant tax rates (functions) in the beginning of the period that satisfy their budget constraints, and I assume that both governments can fully commit to the tax system.

The welfare function that the governments maximize can take several forms depending on the value they put on immigrants and emigrants. In this paper, I assume that governments have a utilitarian social welfare function over the population physically residing in the country, at the time the tax system is chosen.\(^{22}\) Thus, each government’s social welfare function, given the stationary distribution of households \( \Phi^i(s) \) and the welfare of households achieved in stationary equilibrium \( V^i(s) \) is

\[
\int_s V^i(s) d\Phi^i(s). \tag{1.17}
\]

\(^{22}\)The social welfare criterion that I use only captures the lifetime utility of the residents. However, the lifetime utility of current residents includes the potential utility gain from migrating to the other country in the future.
Moreover, the tax functions must be chosen so that they satisfy the governments’ period-by-period budget constraints. The tax revenue requirements are set at the same per capita level as calculated from the calibrated equilibrium, that I describe in the next section. Rather than using a fixed (regardless of population size) amount of government expenditure, I use a per capita tax revenue requirement. It captures that the revenues necessary to provide a given amount of government services increase with population, which is important in this model, since population size is determined endogenously when labor is mobile.

In the next, I formally define a Nash equilibrium of the tax competition game.

**Definition 2.** A Nash equilibrium (non-cooperative equilibrium) of the tax competition game (given public education expenditures per capita \( \{e^1, e^2\} \) and transfers per capita \( \{tr^1, tr^2\} \) in this economy consists of a vector of government policies \( \tau^i_1 \), and associated competitive equilibrium prices \( \{R, r^i, w^i_u, w^i_s\} \), and allocations \( \{c^i, l^i, x^i, a^i\} \), and the stationary distribution of households in each country \( \Phi^i \) for \( i \in \{1, 2\} \) such that:

(i) for each country \( i \), \( \tau^i_1 \) maximizes the steady-state welfare of the economy (1.17), taking as given the tax system of the other country, \( \tau^{-i}_1 \); and

(ii) for \( \{\tau^1, \tau^2\} \), the resulting prices, allocations, and stationary distributions are a competitive equilibrium.

Since this is a dynamic general equilibrium model with heterogeneous households and incomplete markets, it is difficult to derive analytical characteristics of the effects of changing tax rates on macroeconomic outcomes and the best response of the game. However, I briefly discuss some (qualitative) effects of changing labor and capital income tax rates that shed light on their consequences.

Lowering either labor or capital income tax rates would attract capital and labor, since in equilibrium, \( (1-\tau_k)r = (1-\tau_k^*)r^* \) must hold by no-arbitrage condition in an economy with international financial market. In the case of lower capital income taxes, increased capital flows attract migrants through higher wages. On the other hand, under lower labor income taxes, an increase in migrants and the domestic labor supply attracts capital to equalize
after-tax returns to capital. Moreover, lower labor income taxes would increase incentives for households to invest in the human capital of their offspring, as well as increase incentives for labor hours. The extent to which taxes can be better employed as a tool for increasing the welfare of households, therefore, is not straightforward and must be solved numerically under a reasonable set of parameter values. That is the goal of the quantitative analysis conducted in the following section.

1.3 Calibration

I use the United Kingdom and Continental Europe (France, Germany, Italy, Sweden, and Spain, hereafter referred to as the “CE”) as countries of interest for the quantitative analysis. The free movement of labor within countries in the European Union makes the UK and CE an appropriate choice for the tax competition analysis with labor mobility. Moreover, as presented in Table 1.1, the UK has a tax system and aggregate economic outcomes that differs considerably from the CE in consideration. Therefore, considering the UK and CE is a suitable division of the countries in the EU for calibration, and for the further analysis of the impact of free labor mobility on optimal taxation.

1.3.1 Descriptive Statistics of the Macroeconomic Data in the UK and CE

Before I present in detail the calibration strategy and parameter values, I first summarize some of the relevant data statistics (in 2000) in the UK and CE in Table 1.1.24

For the CE data, I use weighted averages from France, Germany, Italy, Spain, and Sweden. Weights are obtained using the average weights of GDP and population: 0.24, 0.34, 0.23, 0.15, and 0.04 respectively for France, Germany, Italy, Spain, and Sweden.25

---

23By Article 45 Treaty on the Functioning of the European Union (ex 39 and 48), which states that: “1. Freedom of movement for workers shall be secured within the Community. 2. Such freedom of movement shall entail the abolition of any discrimination based on nationality between workers of the Member States as regards employment, remuneration and other conditions of work and employment.”
24Statistics for individual countries in Continental Europe are summarized in the Appendix (Table A.1).
25Although the total population in the Continental European countries I use is about four times larger than the population in the United Kingdom, I normalize the sizes in the two countries in the quantitative
As shown in Table 1.1, the UK and CE have comparable GDP per capita. The UK has a higher skill premium, higher labor earnings inequality, and a higher percentage of college graduates. However, overall tertiary education spending is higher in the CE, with its public spending comprising about 75 percent of the total expenditure on tertiary education.

Intergenerational persistence in schooling is obtained from Causa and Johansson (2009). They use OECD data to obtain the percentage increase in probability of an offspring obtaining college education when his or her father is college educated, compared to those whose fathers only have upper-secondary education.\footnote{They also report the same measure comparing fathers with college education and those with less than upper-secondary education. In the UK and CE, the probability premium is around 0.38 and 0.45 respectively (calculated as weighted average of France, Italy, Sweden, and Spain. German data is not available.).}

Migration-related statistics are obtained from OECD bilateral migration data. I calculate immigration stocks from the UK to the CE and vice versa, and find the percentage of college-educated immigrants among the migrants from the UK and CE. Positive selection is evident from data, as the percentage of college graduates among immigrants is 37.7 percent in the UK (36.4 percent in the CE), whereas the same statistic for the total population within the country is 25.7 percent (20.0 percent).

In the following, I first describe the public policies in the UK and CE, and discuss the functional forms and calibration strategies.

### 1.3.2 Public Policies in the UK and CE

Following Benabou (2002) and Heathcote et al. (2012), the after-tax labor income is assumed to take the following functional form:

\[
\tilde{y} = a_0 y^{1-a_1}. \tag{1.18}
\]

The after- and before-tax labor income is represented by \(\tilde{y}\) and \(y\), respectively. The parameter \(a_1\) denotes the degree of progressiveness. The average labor income tax function analysis. In tax competition, sizes can matter, and Kanbur and Keen (1993) is one of the papers that study tax competition and harmonization with countries that differ in size. They find that small countries tend to be tax havens. I discuss a possible implication of this assumption briefly in footnote 42 in section 1.4.3.
implied by the suggested tax relation is

$$\tau(y) = 1 - a_0 y^{-0.1}. \quad (1.19)$$

Note that with $a_1 = 0$, the labor income tax rate is flat at $1 - a_0$. This is the income tax formulation that reflects the labor income tax system in different countries quite well.

I use this tax function to obtain the *status quo* tax function from data of the UK and CE. Following Guvenen, et al. (2011), I use top marginal income rates and thresholds to impute tax rates at multiples of average wages. Then, using the functional form (1.19), I find estimates of $a_0$ and $a_1$ that minimize the distance between the tax rates observed from the imputed tax rates and the labor income tax function. Figure 1.2 plots the estimated labor income tax function in the UK and CE.

The corresponding tax functions are

$$\tau^{UK}(y) = 1 - 0.7792 \left( \frac{y}{\bar{y}} \right)^{-0.0692},$$

$$\tau^{CE}(y) = 1 - 0.7196 \left( \frac{y}{\bar{y}} \right)^{-0.0970},$$

where $\bar{y}$ is the average labor earnings in the economy. The smaller value of $a_0$ and a greater absolute value of $a_1$ in the CE tax function imply a higher (in levels) and a more progressive labor income tax code in the CE than in the UK.

For capital income taxes, I take effective average capital income taxes directly from Carey and Tschilinguirian (2000) who calculate capital income taxes between 1991 and 1997 using the methods of Mendoza et al. (1994). The capital income tax rates are 48 percent in the UK and 28 percent in the CE.\(^{27}\)

In the model, I allow governments to use some of their tax proceeds on providing public education. As will be clear in the next subsection, I will define the skilled labor as those

\(^{27}\)The trends in capital and labor income tax rates in the UK and CE over the last thirty years show that the UK had lower labor income tax rates and higher capital income tax rates than the CE. The trends of tax rates are reported in the Appendix. Moreover, capital and labor income tax rates of individual countries in the CE are also in the Appendix.
who have graduated from college. Therefore, I use the percentages of GDP spent on tertiary education in the UK and CE as their status quo public education policies. The UK, which has a lower labor income tax rates with less progressive income tax code, spends 0.6 percent of GDP in providing tertiary education, whereas the CE spends 0.9 percent of GDP.\textsuperscript{28} In the status quo, the CE, whose labor income tax code is more progressive, provides more public education, which can offset the disincentives created by its tax policies. The rest of the tax proceeds are assumed to be transferred back to the public in a lump-sum fashion. These lump-sum transfers are meant to represent the governments’ redistributive policies, which are not explicitly modeled in this paper.

1.3.3 Calibration Strategy

Given the labor and capital tax income systems and public education expenditure (as a percent of GDP) in the UK and CE, I find parameters to match equilibrium outcomes of the model (with capital and labor mobility) to the observed data. In doing so, I allow for heterogeneity in the relative efficiency of unskilled ($A_{i}^{u}$) and skilled ($A_{i}^{s}$) labor in production and heterogeneity in parameters in the human capital production function ($\nu^{i}, \xi^{i}$) across countries. All other parameters – utility function ($\sigma, \eta, \gamma$), ability distribution ($\rho_{z}, \sigma_{z}$), labor productivity shocks ($\sigma_{\varepsilon}$), and migration costs ($k_{u}, k_{s}, k_{wu}, k_{w}$) – are chosen to match relevant statistics in the UK. Although countries might differ in labor productivity shocks\textsuperscript{29} and migration costs\textsuperscript{30}, in this paper, I keep the model parsimonious. By doing so, I can focus more on the effects of differences in tax and education policies on the migration decision of households and optimal taxation.

While each parameter has implications for different economic outcomes, I simultaneously

\textsuperscript{28}One might argue that public education spending per student is a better indication of a country’s education policy. Since using the measure necessitates the information on the number of students and requires some normalization across countries, using the GDP measure is more straightforward. However, education spending per student is also higher in the CE than it is in the UK, and thus is qualitatively consistent with the public education measures that I use.

\textsuperscript{29}Different labor market institutions (more rigid labor market in the CE, for example) might manifest as differences in labor productivity shocks.

\textsuperscript{30}While (almost all) households in the CE learn to speak English, it is not necessarily the case the households in the UK learn to speak French or German.
calibrate the parameters to match equilibrium outcomes of the model to the data. Using these parameters, I quantitatively solve for the Nash equilibrium of the tax competition game.

A. Production. I use a constant returns to scale production function, which takes the form\(^{31}\) of

\[
Y^i = K^\alpha (A^i_u U^\rho + A^i_s S^\rho)^{\frac{1-\alpha}{\rho}}. 
\]

The labor input in production is a constant elasticity of substitution aggregate of unskilled (\(U\)) and skilled labor (\(S\)), with the implied skill premium of

\[
\frac{w^i_s}{w^i_u} = \frac{A^i_s}{A^i_u} \left( \frac{S^i}{U^i} \right)^{\rho-1}. 
\]

The parameter \(\alpha\) represents the capital share in production, while \(A_u\) and \(A_s\) are the country-specific efficiency of unskilled and skilled labor, and \(\rho\) controls the elasticity of substitution between unskilled and skilled labor (the elasticity of substitution is \(\frac{1}{1-\rho}\)).\(^{32}\)

This parametrization has several benefits when compared to a version which models only one kind of labor input. Most importantly, there is evidence of cross-country differences in the substitutability of skilled and unskilled labor.\(^{33}\) In a model with labor mobility as this one, differences in relative efficiency of two kinds of labor have implications for the tax-induced efficiency gain from reallocation of labor across coun-

---

\(^{31}\)This is a simplified version of Krusell et al. (2000): I abstract from capital-skill complementarity and the division of capital into the capital equipment and capital structures. Caselli and Coleman (2006) uses this production function for estimation of the production parameters across countries.

\(^{32}\)The production function can be rewritten as

\[
F(K, U, S) = z^i K^\alpha (\tilde{A}^i U^\rho + (1 - \tilde{A}^i) S^\rho)^{\frac{1-\alpha}{\rho}}, 
\]

where \(z^i = (A^i_u + A^i_s)^{\frac{1-\alpha}{\rho}}\), and \(\tilde{A}^i = \frac{A^i_s}{A^i_u + A^i_s}\). Thus, differences in \(A^i_u\) and \(A^i_s\) can also be interpreted as differences in total factor productivity \(z^i\), and efficiency of skilled and unskilled labor with a normalization (\(\tilde{A}^i\)).

\(^{33}\)Caselli and Coleman (2006) estimate labor efficiency parameters \(A_u\) and \(A_s\) of the CES production function using the skill premium, and find heterogeneity across countries.
tries. In quantitative analysis, as I define the skilled labor in the model to be the college-educated workers in the data, the skill premium (college premium)\textsuperscript{34} directly obtained from the production function has the data counterparts that I can use as a target in the calibration exercise. Moreover, I can use data on migration statistics for households with and without college education by mapping them directly to the model.

As is customary in the literature, I set capital share in production ($\alpha$) to be 0.33. Another production function parameter, $\rho$, which controls the elasticity of substitution between skilled and unskilled labor, is set so that the CES is around 1.4 following Katz and Murphy (1992).

Taking as given capital share $\alpha$ and $\rho$, which controls elasticity of substitution between skilled and unskilled labor, I calibrate $A_{s}^{UK}$, $A_{u}^{UK}$, $A_{s}^{CE}$, and $A_{u}^{CE}$ that represent the efficiency of skilled and unskilled labor in production within the model. Skilled labor in my model is defined as workers who have graduated from college, and unskilled labor, as those who have not. With this classification, productivities are normalized by letting $A_{s}^{UK} + A_{u}^{UK} = 1$. Then, using the college premium in the UK and CE, and the ratio of GDP per capita in the UK to the CE, values for $A_{u}^{UK}$, $A_{s}^{CE}$, and $A_{u}^{CE}$ are determined. The estimates of the college premium are taken from a series of papers published as a part of the project “Cross Sectional Facts for Macroeconomists”\textsuperscript{35}. This calibration strategy is similar to the one used by Caselli and Coleman (2006), in which they estimate the parameters of the CES production function in different countries using skill premiums, capital level, and supply of skilled and unskilled labor.

The capital depreciation rate $\delta$ is chosen to match the capital-output ratio in the UK.

\textsuperscript{34}I use skill and college premium interchangeably.

\textsuperscript{35}A summary of the project is presented in Krueger, et al (2010). I also obtain data for individual countries from Blundell and Etheridge (2010), Domeij and Floden (2010), Fuchs-Schuendeln, et al. (2010), Jappelli and Pistaferri (2010), and Pijoan-Mas and Sanchez-Marcos (2010), which summarize macroeconomic facts in Britain, Sweden, Germany, Italy, and Spain, respectively. From here on, I denote this series of papers as CSFM.
B. Utility Function. Households' life time function is

\[ u(c, l) + \beta EV' = \frac{c^{1-\sigma} - 1}{1-\sigma} - \eta \frac{l^{1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}} + \beta EV', \]

where \( c \) and \( l \) are consumption and labor hours, respectively. Under this specification, the theoretical Frisch elasticity of labor supply is \( \gamma \).

I set the risk aversion parameter \( \sigma \) to be 2, leaving three parameters to be calibrated within the model: the altruism factor \( \beta \), weight on utility of leisure \( \eta \), and the curvature of the utility of leisure \( \gamma \). Since households live for two periods, I assume a model time period of 30 years. Thus, I find \( \beta^{\frac{1}{30}} \) to match the annual interest rate of 5 percent. In order to jointly pin down the values of \( \eta \) and \( \gamma \), average hours worked in the UK and the Frisch labor supply elasticity as targets. The OECD reports that the average hours worked in the UK is around 1,715 hours annually, or 4.70 hours a day. Letting time endowment be around 13 hours a day, this is equivalent to labor hours of 0.361. Moreover, the target value of the Frisch labor supply elasticity (\( \gamma \)) is 0.3 (Browning et al. (1999)).

C. Ability and Human Capital Production. I assume that ability across generations follows a first-order Markov process with mean normalized to one. In particular, the following AR(1) process is used:

\[ \log z' = \rho_z \log z + \varepsilon, \quad \varepsilon \sim (0, \sigma_z^2), \]

which leaves two parameters, \( \rho_z \) (correlation of ability across generations) and \( \sigma_z^2 \) (variance of error term) to be calibrated, where I use Tauchen method with 5 grid points to discretize the process.

Probabilities of becoming a skilled or unskilled worker, \( Q(s|z, \tilde{x}) \) and \( Q(u|z, \tilde{x}) \), are

\[ ^{36}\text{Since the model time period is around 30 years, households might not respond to wage rates as much as macro Frisch elasticities used in business cycle studies suggest. Therefore, I use the target Frisch elasticity to those estimated and commonly used in micro literature.} \]
specified as follows:

\[ Q(s|z, \bar{x}) = \min\{\nu^i z \bar{x}^{\xi^i}, 1\} \]
\[ Q(u|z, \bar{x}) = 1 - Q(s|z, \bar{x}). \]

This functional form satisfies decreasing marginal returns in human capital investment. Additionally, ability \((z)\) and investment \((\bar{x})\) are complementary, generating intergenerational persistence in schooling. Therefore, there are two parameters for each country to be calibrated: \(\{\nu^{UK}, \nu^{CE}\}\), the technology of human capital production, and \(\{\xi^{UK}, \xi^{CE}\}\), the returns to human capital investment. I normalize \(\nu^{UK} = 1\) and find \(\{\xi^{UK}, \xi^{CE}, \nu^{CE}\}\) within the model, leaving a total of 5 parameters including the ability parameters \(\rho_z\) and \(\epsilon_z^2\).

These parameters affect the intergenerational persistence in schooling, percentage of college graduates, and percentage of GDP spent on education. In the UK, schooling persistence is 0.22. The OECD reports that the percentage of college graduates among the 25 – 64-year-old age group is 25.7 percent for the UK and about 20 percent in the CE. Moreover, the total percentages of GDP spent on tertiary education in 2000 are 0.78 percent and 1.06 percent in the UK and CE, from which 0.6 percent and 0.9 percent are public spending, and the rest, private.

One of the key aggregate outcomes that I would like to analyze in this paper is the skill distribution in different countries. As shown in the summary statistics, the share of skilled population in the CE is only about 80 percent of that in the UK. With common ability and human capital production function parameters, such big differences in skill distribution cannot be matched. Since the education system and quality of countries can differ, I allow for heterogeneity in technology \((\nu)\) and returns to human capital investment \((\xi)\) across countries. This makes it possible to match the skill distribution across countries, with which I can analyze the effects of different tax systems in both countries.
D. Labor Productivity Shocks. I assume that the i.i.d. productivity shock to labor income is drawn from a log-normal distribution:

\[ \log \varepsilon \sim N(0, \sigma_{\varepsilon}^2). \]

The value of \( \sigma_{\varepsilon}^2 \) is found to match labor earnings Gini in the UK, which is reported to be 0.45 in 2000 from Blundell and Etheridge (2010).

E. Migration. Moving cost shocks have an exponential distribution (as in Armenter and Ortega (2009)) with skill-dependent scale parameters of \( \{k_u, k_s\} \) and a minimum value of moving cost \( \{k_u, k_s\} \), i.e.,

\[ f(h) = k_i e^{-k_i(h+k_i)}, \quad i \in \{u, s\}. \]

Since this is a steady-state analysis of an open-economy with labor and capital mobility, net mobility is always zero (while gross mobility might not be). Therefore, I set the percentage of migration stock in the UK (from the CE) in its population as one of the data targets.

In order to match the percentage of college-educated migrants (positive selection is evident, as seen in Table 1.1), I assume heterogeneity in scale parameters of the exponential distribution across skill levels. Relevant targets are the population ratio between the two countries and the percentage of college-educated immigrants.

Table 1.2 summarizes the model parametrization.

1.3.4 Calibration Results

Values of the calibrated parameters and model fit are presented in Tables 1.3 and 1.4.

As emphasized before, there are two fundamental differences between the two countries: the efficiency of skilled and unskilled labor in production, and the human capital production technology. In the calibration exercise, I find that the UK has higher efficiency of skilled
labor compared to unskilled labor in production, i.e.,

\[
\frac{A^U_{SK}}{A^U_{SK}} > \frac{A^C_{SK}}{A^C_{SK}}.
\]

Moreover, the sum of labor efficiencies in the CE \((A^C_{SK} + A^C_{US})\) is 0.945, which is lower than normalized value of 1 in the UK.\(^37\) The differences in labor efficiency parameters between the UK and CE are qualitatively similar to those found in Caselli and Coleman (2006). The calibrated parameters of the human capital production technology shows that the CE has higher technology \((\nu)\) and higher returns to human capital investment \((\xi)\).

These parameters (production and human capital technology) are found jointly to match the GDP per capita ratio and skill premiums, skill distributions, and education spending in both economies. The UK has higher skill premium and higher percentage of skilled workers in general equilibrium, which is consistent with the higher relative efficiency of the skilled in production. Moreover, high technology of human capital production and returns to human capital investment are necessary to match the skill distribution in the CE, while having a low skill premium. The model is successful in matching targets along these dimensions.

Migration parameter choices also require some discussion. Since I conduct a steady-state analysis, net migration in both countries is always zero by construction. Therefore, in order to ensure a stationary distribution of population in both countries, households should migrate to and from both countries. To achieve this, I allow for the minimum value of the location preferences (including moving costs) to be negative, which implies that a household living in the UK might prefer living in the CE and vice versa. This allows households in the UK to migrate to the CE, for example, even though it is not in their best interest to do so for economic reasons. I use data on migration stock as a calibration target. In the status quo economy, migration stocks from the CE in the UK is only 1.1 percent of the

\(^37\)The parameter values imply that the production function in Equation (1.20) is represented as

\[
F^{UK}(K,U,S) = K^\alpha (0.5762U^\rho + 0.4238S^\rho)^{1-\alpha \rho};
\]

\[
F^{CE}(K,U,S) = 0.8764K^\alpha (0.6534U^\rho + 0.3466S^\rho)^{1-\alpha \rho},
\]

for the UK and CE, respectively. The overall productivity is lower in the CE (0.8764) and the unskilled efficiency in production, higher (0.6534), than in the UK.
total UK population. This leads the scale of the distribution to be very high, implying a very strong preference for households to live in their birth country. The lower scale for the skilled compared to unskilled reflects the lower cost of migration for the skilled.

In the fully calibrated economy, government revenues collected from labor and capital income taxes amount to about 30 percent of GDP in the UK and CE. According to the OECD Tax Database, total tax revenues in the UK and CE (in 2000) are 37 percent and 40 percent respectively, among which 30 percent are collected as consumption taxes (general and specific). Since consumption, capital, and labor income taxes constitute three major sources of tax revenues, labor and capital tax revenue range between 26 and 28 percent. Thus, the calibrated economy is within a reasonable range in terms of the magnitude of tax revenues relative to GDP.

1.3.5 Computation of Nash Equilibrium

Before I analyze the results in detail, I briefly describe the quantitative procedures of solving for the Nash equilibrium.

In order to find a Nash equilibrium of the tax competition game, I generate a grid of the labor income tax parameters, \( a_0 \) and \( a_1 \). Taking as given the tax system of country \( j \) \( (\{a_0^j, a_1^j, \tau_k^j\}) \), country \( i \) searches over the set of grids \( \{a_0, a_1\} \) to find the tax system that maximizes the steady-state welfare of households, while collecting a fixed amount of tax revenue per capita.

As governments’ budget constraints have to be satisfied period by period, I find capital income tax rate \( \tau_k \) that balances country \( i \)’s budget, given the labor income tax code represented by \( \{a_0, a_1\} \). This procedure is repeated iteratively for countries \( i \) and \( j \).

---

38 Total migration stock in the UK is about 11 percent of its population, with 3.5 percent coming from other EU countries. However, since I want to focus on the effect of different tax policies on migration and aggregate outcomes between the UK and the CE, 1.1 percent is the relevant data target for the purpose of this paper.

39 This is consistent with qualitative differences in estimates for moving costs in Armenter and Ortega (2009).

40 Specific tax categories include personal income tax, corporate income tax, social security contributions, payroll taxes, property taxes, and general and specific consumption taxes.

41 Notice that since this is a general equilibrium model, equilibrium allocations and prices are determined given the tax systems in both countries.
until a fixed point is found. Given the high dimensionality of the choice variables, the uniqueness of the Nash equilibrium cannot be guaranteed. However, I searched over big domains of tax parameters and tried starting from various initial tax systems, and have not found multiple equilibria. I describe this procedure in more detail in Appendix A.3.

In the following sections, I present the quantitative results based on the calibrated parameters.

1.4 Results

To more intuitively understand the forces leading to the optimal taxation results, I first start by analyzing household behaviors. Then I let each government (UK and CE) vary its tax code \((a_0, a_1, \text{ and } \tau_k)\) one by one, and analyze its aggregate effects and welfare. This exercise provides an intuitive picture of what the best response of each country is, taking as given the status quo tax system of the other. After presenting the best responses, I discuss the Nash equilibrium tax system and its aggregate outcomes. Finally, by comparing the optimal taxation in the benchmark economy (both capital and labor mobility) to the economy with only capital mobility, I study how labor mobility affects the fiscal choices of governments.

1.4.1 Understanding Household Behavior

Governments in the economy maximize the steady-state welfare of households. Therefore, understanding household behaviors is an important first step in fully analyzing the optimal taxation results.

Among the most important components in the model are the human capital investment and the migration decisions of households. Since governments in both countries provide public education, for households whose return to human capital investment is low, their optimal investment \((x)\) in the model is zero. Given the marginal complementarity between ability \((z)\) and human capital investment, these are households with low ability levels.
The incentives to invest in human capital also differ across skill group and country of residence. Figure 1.3 shows the human capital investment decisions of households living in the UK and CE across skill groups, for the highest-ability parents as a function of asset level. The skilled parents, whose labor income is higher, invest more in the human capital of the offspring than their unskilled counterparts do, since they have higher total wealth. Moreover, households living in the CE have lesser incentives to invest in human capital, since the skill premium is lower, the labor income tax rate (and progressivity) is higher, and the public education expenditure in the CE is higher.

To understand the migration decisions of households, in Figure 1.4, I plot the differences in the value of living in the CE and living in the UK, for households with different skill and ability levels, across assets. As is evident from the figure, the skilled prefer living in the UK, since the skill premium is higher and the labor income tax rate is lower than those in the CE. Moreover, parents whose children are of low ability prefer living in the CE (conditional on their own skill level), where the public education expenditure and the return to human capital investment are higher than those in the UK. However, the differences in the value of living in the UK and CE vanish as assets increase, since labor income becomes a smaller share of the total household wealth and there is a global financial market (the after-tax returns to capital are equalized across countries).

These household behaviors show that government policies, along with fundamentals in the economy, can have important implications for the aggregate outcomes, especially with international movement of labor.

1.4.2 From Status Quo to the Nash Equilibrium

In this subsection, I discuss the response of the UK (CE) government to the status quo tax system of the CE (UK). This analysis provides insight into the optimal choices of the governments.
Effects of Changes in the Tax Code

First, I consider the effects of changes in $\tau_k$, $a_0$, and $a_1$ of the UK, taking as given the CE's status quo tax system. Table 1.5 presents the effects of lowering the capital income tax rate $\tau_k$ and average labor income tax rate $a_0$ (keeping progressivity constant) by 1 percent, and the effects of reforming the labor income tax code to a proportional tax.

As is evident in Table 1.5, lowering the capital income tax rate is the most effective tool for increasing capital. Capital increases by 0.04 percent in response to a 1 percent decrease in the capital income tax rate. On the other hand, changes in the labor income tax code increase average hours worked in the economy. In the current formulation of the tax code, the governments can take two routes for reforming the tax code: either lower the average labor income tax rate (through higher $a_0$) or lower the progressivity of the tax code (through lower $a_1$). The comparison of the last two columns of Table 1.5 shows that lowering the progressivity of the labor income tax code leads to a higher increase in labor provided, as well as an increase in the number of skilled immigrants.

The last three rows in Table 1.5 present the consumption equivalent variations: the consumption increase necessary for households to be indifferent between living in the status quo economy or in the economy with the new tax regime. It is clear that the capital income tax reform performs the best in increasing the overall welfare of households; the consumption equivalent variations are the highest in this case. Another interesting aspect of the economy is that implementing a proportional labor income tax system results in a welfare loss: the insurance benefits from using a progressive labor income tax code outweigh its incentive costs. Even though the skilled households are in favor of the proportional labor income tax code (0.019), the unskilled, who comprise about 75 percent of the population in the UK, suffer welfare loss (-0.008). The heterogeneity in ability levels and idiosyncratic labor income risks cannot be insured in the market, and governments play important roles in providing insurance benefits against these risks. A progressive labor income tax code achieves this goal.
Best Responses to the Status Quo Tax System

The analysis in the previous section provides insight into what the best response of the UK is (I call this the \textit{unilateral} reform by the UK government). In the best response, the UK sets its capital income tax rate at -35 percent (subsidized), average labor income tax rate at 53 percent, and progressivity at 0.25 (value of \(a_1\)). The changes to the aggregate outcomes in the UK compared to the \textit{status quo} are presented in Table 1.6.

The dramatic decrease of the capital income tax rate in the UK (from 48 percent to -35 percent) leads to higher capital and output per capita. The return to capital decreases by 42 percent, but due to the capital income subsidy, the after-tax return to capital increases by 19 percent. Moreover, as higher capital increases the marginal product of labor, wages of both unskilled and skilled labor also increase by around 37 percent.

The results of the analogous exercise for the CE are presented in Table 1.7. The CE also lowers its capital income tax rate and increases progressivity, but the magnitudes are smaller. Since the \textit{status quo} capital income tax rate in the UK (48 percent) is higher than that in the CE (28 percent), it attracts sufficient capital at a relatively higher capital income tax rate.

The presentations of the effects of changes in the tax code and the best responses to the \textit{status quo} tax system lead us to analyze the Nash equilibrium results, which I proceed to do in the following section.

1.4.3 Nash Equilibrium of the Model

The Nash equilibrium of the tax competition game is

\[
\bar{\tau}_{y}^{UK}(y) = 0.52, \quad a_{1}^{UK} = 0.28 \quad ; \quad \bar{\tau}_{k}^{UK} = -0.32
\]

\[
\bar{\tau}_{y}^{CE}(y) = 0.50, \quad a_{1}^{CE} = 0.11 \quad ; \quad \bar{\tau}_{k}^{CE} = 0.06.
\]

The average labor income tax rates in the UK and CE are 52 percent and 50 percent, respectively. While the UK subsidizes its capital income at 32 percent, it uses a more
progressive labor income tax code than the CE does.\footnote{As I have noted in footnote 25 section 1.3.1, I normalize the country size of the CE in this analysis. The main result of Kanbur and Keen (1993), who studies tax competition between countries with different sizes, is that the small country sets a lower tax rate, since there is more to be taken advantage of from the big country. In the Nash equilibrium, the UK sets a lower capital income tax rate than the CE does, and it attracts capital, which is the more mobile factor of production (compared to labor). Thus, I expect to see a even stronger incentive for the UK to set a lower capital income tax rate, if I take into account the size of the CE. Of course, I will have to recalibrate the model to be able to analyze it more accurately, which I plan to do in the future.} Figure 1.5 plots the average labor income tax rates in the Nash equilibrium. Moreover, as a measure of progressivity, I also plot in Figure 1.6, progressivity wedges defined as

\[ 1 - \frac{1 - \tau(k \times 0.5)}{1 - \tau(0.5)} \text{ for } k = 2, 3, \ldots, 5, \]

following Guvenen et al. (2011) The interpretation is as follows: In the Nash equilibrium, a household living in the UK, whose labor earning is two times the average earning in the economy, earns 25 percent less than they would in a flat-tax system. An analogous worker earns 12 percent less, if he resides in the CE. In the status quo, the wedges are 8 and 10 percent respectively, in the UK and CE. Therefore, it is evident that the UK uses a much more progressive labor income tax code to collect tax revenues.

Before discussing the aggregate effects of implementing the Nash equilibrium policies and the impact of competition between the two countries, I provide a brief insight for why the UK sets a lower capital income tax rate than the CE does.

In the model economy, labor income risks cannot be insured in the market, i.e., financial markets are incomplete. From the government’s point of view (which maximizes the steady-state welfare of households, using a utilitarian welfare function), the most efficient way of increasing societal welfare is to increase the lifetime utility of the consumption-poor, as they are the segment of society with the highest marginal utility of consumption. Thus, as shown in Davila et al. (2012),\footnote{The condition for constrained efficiency in a simplified version of my model with three factors of production (capital, unskilled, and skilled labor) in an open economy is presented in Appendix A.4. This is an extended version of Davila et al.} the most important determinant of the optimal capital income tax rate is the wealth composition of the poor in the competitive equilibrium (without government policies). If the consumption of the poor is low since they are relatively more
wealth-poor (but have an abundant labor endowment), then social planners can increase the welfare of the poor by increasing capital, since an increase in capital would result in higher wages. On the other hand, if the poor in the economy is relatively more labor income-poor (but have high asset positions), the welfare of the poor can be increased if the return to capital is higher – an argument for positive capital income tax rate.

In the UK, with higher efficiencies in production, capital and wages are higher than in the CE in competitive equilibrium (without government policies). Moreover, the wealth inequality in the UK is higher than that in the CE. Thus, the UK government benefits more from increasing capital, which also drives up wages. This leads to the optimal capital income tax rate being relatively lower (negative – capital income is subsidized) than that in the CE, and the optimal labor income tax code, progressive.

**Aggregate Outcomes in the Nash Equilibrium**

Table 1.8 presents the changes in the aggregate outcomes in the UK and CE when both countries implement the Nash equilibrium tax systems.

In the Nash equilibrium, the UK lowers its capital income tax rate dramatically. As a consequence, output and capital per capita increases in the UK by 0.21 and 1.18. However, the CE, which under the status quo had higher capital per capita than the UK does not perform economically as well as the UK does in the Nash equilibrium. Its output and capital per capita decreases in the Nash equilibrium. However, the lower capital income tax rate in both countries, compared to the status quo leads to higher after-tax returns to capital. Therefore, even with lower average working hours, consumption per capita in both countries increases.

Another notable aspect of the economy is that, in the Nash equilibrium, there is a less skilled population. As the after-tax return to physical capital investment increases, the incentives to invest in human capital decreases. Moreover, the highly progressive labor income tax code, compared to the status quo, decreases the incentives for parents to invest in the human capital of their offspring. The lower human capital investment leads to
lower intergenerational schooling persistence in the UK, while in the CE, the schooling persistence increases. Under the Nash equilibrium policies and with its (relatively more) generous public education system, in the CE, households’ human capital investment nears zero and the schooling persistence increases slightly. On the other hand, in the UK, the decrease in the human capital investments are more pronounced for the skilled, since they have higher asset income. This leads to a lower schooling persistence in the UK. The cross-sectional inequality in both countries increases, as measured by the income Gini coefficient, because the asset-rich households decrease their working hours.

Overall, implementing the Nash equilibrium policies leads to a significant increase in lifetime welfare of households, as measured in consumption equivalent variation. While the welfare in the UK increases by around 13 percent, the welfare in the CE increases by around 10 percent. The welfare increase is derived from the higher consumption per capita and lower hours worked. The unskilled benefit more than the skilled do, by about twofold, in both countries.

**Effects of Competition**

Given the best responses of the governments discussed in section 1.4.2, the model can be used to analyze the effects of competition on optimal taxation. Comparing the Nash equilibrium tax rates with the tax rates of the unilateral reforms, I find that the governments set lower capital income tax rates under unilateral reform than they do in the Nash equilibrium. For both countries, the benefit of lowering the capital income tax rate is higher when the other country’s tax rate is high, since there is more capital that can be taken advantage of. Thus, both governments set lower capital income tax rates when they unilaterally reform their tax systems.
1.4.4 Effects of Labor Mobility on Optimal Taxation and Aggregate Outcomes

In this subsection, I study the effects of labor mobility on optimal taxation in the UK and CE, which is the main contribution of this paper.

For the analysis, I find the optimal tax code in a closed economy, an economy with only capital mobility, and an economy with both capital and labor mobility (benchmark result). In the first subsection, I compare the optimal taxation under different mobility assumptions and provide elasticities to tax rates to understand the optimal taxation results. I also discuss the aggregate outcomes under the economy with only capital mobility and with both capital and labor mobility to analyze the macroeconomic effects of labor mobility.

I find that labor mobility is crucial in determination of the optimal tax system in the UK and CE. Given the differences in the two countries, incorporating labor mobility leads to a divergence in the tax systems and reallocation of unskilled and skilled labor across countries. I discuss these issues in more detail in the following.

Optimal Taxation with versus without Labor Mobility

Tables 1.9 and 1.10 present the optimal tax codes of the UK and CE in the economies with different mobility assumptions.

First, I focus on the optimal taxation results in the UK. A noticeable finding is that the optimal capital income tax rate is lower under both labor and capital mobility in the UK. When labor is mobile, the governments can attract migrants by decreasing labor income tax rates or by increasing wages directly: Households’ economic decisions depend on the level of after-tax labor income. Thus, the governments weigh the trade-off they face in choosing labor or capital income tax rates as a source of their tax revenue. If they lower capital income tax rates, at the expense of higher labor income tax rates, they can induce higher labor hours and attract migrants through higher wages. Similarly, lower labor income tax rates achieve the same goal.

The quantitative exercise in the calibrated model, however, shows that the former (a
lower capital income tax rate) is more effective in improving economic well-being in the UK. To more formally argue this point, I present in Table 1.11 the effects of a one percent decrease in the capital income tax rate under different mobility assumptions in the UK.

In the closed economy, when the government lowers the capital income tax rate, capital increases, return to capital decreases, and wages increase. When a global financial market is formed (capital mobility), the lower capital income tax rate attracts higher capital, but its effect is smaller, as the price of capital is set in the global market: The return to capital increases, despite the higher level of capital. As a consequence, the effects on wages are also lower than under the closed economy.

In the last column of the Table 1.11 are the results in the economy with both capital and labor mobility. It is evident that the benefit of a lower capital income tax rate is higher under dual capital and labor mobility, than it is under only capital mobility. Capital increases more, and unlike in the other cases, so does average hours worked. In the UK, the benefit of a lower capital income tax rate is higher with labor mobility.

This result is driven by two facts. Firstly, capital is more mobile than labor. This is assumed in the model, and reflects reality. International migration of households is not only costly financially, but in most cases, also requires overcoming language and cultural barriers. On the other hand, financial markets in recent years have been largely globalized, and households can easily engage in international financial transactions with no cost. Secondly, capital and labor are complementary in production: An increase in one factor increases the marginal product of the other. Therefore, governments may find it more efficient to use lower capital income tax rates to attract capital and thus, indirectly, labor. The UK therefore lowers its capital income tax rate further with labor mobility, and it attracts migrants by doing so. The population ratio (UK to CE) in the Nash equilibrium is 1.03.

Another effect of labor mobility is that the optimal tax systems diverge in the UK and CE. In an economy where only capital is mobile, the capital income tax rates in both

\[44\] Although after-tax returns to capital are equalized in the global financial market, there is a cost to migrate to the other countries. This not only reflects the reality, but is necessary for the existence of a steady-state in the economy with a non-degenerate distribution of population in the two countries.
countries are around -11 percent. However, with labor mobility, while the UK lowers its capital income tax rate further, the CE lowers its labor income tax rate (higher capital income tax rate).

The divergence of the tax system in an economy with labor mobility is driven from the productivity differences in the UK and CE. As emphasized before, in the calibrated economy, the UK has higher productivity than the CE does. Thus, when labor is mobile, the UK attracts more capital and labor: The population ratio between the UK and CE (UK to CE) is greater than one. Given a relatively smaller population, the welfare-maximizing capital in the CE is smaller. Moreover, since in equilibrium, after-tax returns to capital between the UK and CE are equalized (no-arbitrage condition in the global financial market), the CE finds it optimal to use a higher capital income tax than the UK does.

In order to analyze the welfare effects of using different tax policies further, I present in Table 1.12, the effects of lowering the capital versus labor income tax rate by one percent (from the Nash equilibrium policies) in the UK and CE, respectively. The UK, which has higher overall production efficiency and relative skilled efficiency, benefits more from lowering its capital income tax rate, as measured by consumption equivalent variation. On the other hand, the consumption equivalent variation in the CE is higher when it lowers its labor income tax rate. This leads to a divergence of tax systems in the two countries.

**Comparison of Aggregate Outcomes with Different Mobility Assumption**

In Table 1.13, I compare aggregate outcomes in the economy with only capital mobility and the one with both capital and labor mobility.

As countries open up for labor mobility, the UK, which is more productive, gains as output and capital per capita increase. While the output and capital per capita in the CE decreases, with more unskilled and skilled labor provided, consumption per capita increases. Moreover, given the higher relative efficiency of the skilled in the UK, more skilled workers live in the UK, while in the CE, more unskilled workers reside. Thus, skilled and unskilled

---

45 As specified in footnote 31, the overall productivity in the CE (0.8764) is lower than that in the UK (1).
labor are reallocated across countries as labor becomes mobile.

The analysis in this section shows that labor mobility has important implications for the optimal taxation of capital and labor. In the calibrated model, the UK lowers its capital income tax rate, while the CE lowers its labor income tax rate, compared to an economy with only capital mobility. In the aggregate, the UK, which is more productive and has higher relative efficiency of the skilled, enjoys higher population rates and output and capital per capita, when labor is mobile.

1.5 Conclusions

The recent trend of increased labor mobility has raised concerns about the possibility of international labor income tax competition among policy makers. In light of this trend, this paper asks how labor mobility affects the optimal choice of governments’ tax policies in an international tax competition framework.

Based on a two-country, open-economy model, I find that labor mobility and a global financial market are important factors in determining the optimal tax systems. The application of the model to the United Kingdom and Continental European countries shows that countries set their capital income tax rates less aggressively in a competitive environment. Moreover, in the United Kingdom, the optimal taxation of capital income is lower under both capital and labor mobility, compared to an economy with only capital mobility, while the reverse is true in Continental Europe. Though some policy makers worry that international tax competition can be harmful, implementing the Nash equilibrium tax rates in both countries increases welfare by 11 percent of consumption in the status quo economy.

An interesting avenue for future research is to endogenize the choice of public education expenditures in governments’ problems. Since governments’ public education policies affect the migration decisions of skilled and unskilled households differentially, education policies might be used to attract migrants selectively.\textsuperscript{46} Moreover, the interaction between using a

\textsuperscript{46} Anderson (2005) theoretically investigates the link between taxation and educational policies in presence of migration.
progressive labor income tax code and a generous public education policy can be analyzed in an international tax competition framework.

The model framework developed in this paper is useful beyond the analysis of the United Kingdom and Continental Europe. It is also suitable for investigating the effects of increased labor mobility on the Nash equilibrium as well as aggregate outcomes. As workers become more mobile, governments face stronger constraints on labor income taxes, with higher benefits from an increased level of capital stock. A quantitative assessment of the strength of the two effects can shed some light on the future direction of optimal tax policies for governments facing a more globalized world. Moreover, the model can also be used for the analysis of migration reforms. Tightening immigration policies, for example, in the United States, can have implications for optimal taxation and aggregate economics outcomes. I leave these questions for future research.
### 1.6 Tables and Figures

Table 1.1: Summary Statistics in the United Kingdom and Continental Europe

<table>
<thead>
<tr>
<th></th>
<th>United Kingdom</th>
<th>Continental Europe</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP per capita$^a$</td>
<td>$25,255</td>
<td>$25,174</td>
</tr>
<tr>
<td>Skill premium$^a$</td>
<td>1.62</td>
<td>1.45</td>
</tr>
<tr>
<td>Intergenerational persistence in schooling$^b$</td>
<td>0.22</td>
<td>0.23</td>
</tr>
<tr>
<td>Gini coefficient of labor earning$^a$</td>
<td>0.45</td>
<td>0.36</td>
</tr>
<tr>
<td>% college graduates$^c$</td>
<td>25.7 %</td>
<td>20.0%</td>
</tr>
<tr>
<td>Tertiary education spending (% GDP)$^c$</td>
<td>0.78%</td>
<td>1.06%</td>
</tr>
<tr>
<td>Public spending</td>
<td>0.58%</td>
<td>0.90%</td>
</tr>
<tr>
<td>Private spending</td>
<td>0.20%</td>
<td>0.16%</td>
</tr>
<tr>
<td>Average hours worked$^c$</td>
<td>0.361</td>
<td>0.342</td>
</tr>
<tr>
<td>% of immigrants among pop.$^b$</td>
<td>1.10%</td>
<td>0.56%</td>
</tr>
<tr>
<td>% college graduates among immigrants$^d$</td>
<td>37.70%</td>
<td>36.41%</td>
</tr>
</tbody>
</table>

$^a$. Cross Sectional Facts for Macroeconomists (2011)

$^b$. Causa and Joahnsson (2009)

$^c$. OECD

$^d$. OECD and World Bank

*Note*: All statistics are data from 2000.
Table 1.2: Calibration Targets

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Target</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_{UK}^{u}$</td>
<td>Unskilled efficiency – UK</td>
<td>Skill premium – UK</td>
<td>1.62</td>
</tr>
<tr>
<td>$A_{CE}^{u}$</td>
<td>Unskilled efficiency – CE</td>
<td>Skill premium – CE</td>
<td>1.45</td>
</tr>
<tr>
<td>$A_{CE}^{s}$</td>
<td>Skilled efficiency – CE</td>
<td>GDP per capita ratio</td>
<td>1.00</td>
</tr>
<tr>
<td>$\alpha^*$</td>
<td>Capital share</td>
<td>-</td>
<td>0.33</td>
</tr>
<tr>
<td>$\frac{1}{\delta - \rho}$</td>
<td>Elasticity of Substitution</td>
<td>-</td>
<td>1.4</td>
</tr>
<tr>
<td>$\sigma^*$</td>
<td>Risk aversion</td>
<td>Capital-output ratio – UK</td>
<td>3.0</td>
</tr>
<tr>
<td>$\eta$</td>
<td>Weight in disutility of labor</td>
<td>Average hours worked – UK</td>
<td>0.36</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>Curvature of leisure</td>
<td>Frisch elasticity</td>
<td>0.3</td>
</tr>
<tr>
<td>$\beta \frac{1}{\pi}$</td>
<td>Discount (altruism) factor</td>
<td>Annual bond yields</td>
<td>4%</td>
</tr>
<tr>
<td>$\rho_z$</td>
<td>Persistence in ability</td>
<td>Inter-gen. sch. pers. – UK</td>
<td>0.22</td>
</tr>
<tr>
<td>$\sigma_{z}^{2}$</td>
<td>Variance in ability</td>
<td>Educ. spending – UK</td>
<td>0.78%</td>
</tr>
<tr>
<td>$\nu_{CE}$</td>
<td>HC technology</td>
<td>Educ. spending – CE</td>
<td>1.06%</td>
</tr>
<tr>
<td>$\xi_{UK}$</td>
<td>Returns to HC inv. – UK</td>
<td>% college grads – UK</td>
<td>26%</td>
</tr>
<tr>
<td>$\xi_{CE}$</td>
<td>Returns to HC inv. – CE</td>
<td>% college grads – CE</td>
<td>20%</td>
</tr>
<tr>
<td>$\sigma_{z}^{2}$</td>
<td>Var. of iid. shock</td>
<td>Labor earnings Gini – UK</td>
<td>0.45</td>
</tr>
<tr>
<td>$h_{U}^{U}$</td>
<td>Min. utility cost of moving – unskilled</td>
<td>Pop. ratio – UK to CE</td>
<td>0.94</td>
</tr>
<tr>
<td>$h_{S}^{U}$</td>
<td>Min. utility cost of moving – skilled</td>
<td>% imm. among pop. – UK</td>
<td>1.1%</td>
</tr>
<tr>
<td>$h_{U}^{V}$</td>
<td>Scale of cost distribution</td>
<td>% skilled among imm. – UK</td>
<td>38%</td>
</tr>
<tr>
<td>$h_{S}^{V}$</td>
<td>Scale of cost distribution</td>
<td>% skilled among imm. – CE</td>
<td>36%</td>
</tr>
</tbody>
</table>

* These parameters are chosen outside the model, and take standard values used in macroeconomics literature.
Table 1.3: Calibrated Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A^U)</td>
<td>Unskilled efficiency – UK</td>
<td>0.577</td>
</tr>
<tr>
<td>(A^C)</td>
<td>Unskilled efficiency – CE</td>
<td>0.618</td>
</tr>
<tr>
<td>(A^U)</td>
<td>Skilled efficiency – UK</td>
<td>0.423</td>
</tr>
<tr>
<td>(A^C)</td>
<td>Skilled efficiency – CE</td>
<td>0.329</td>
</tr>
<tr>
<td>(\delta)</td>
<td>Capital depreciation rate</td>
<td>0.049</td>
</tr>
<tr>
<td>(\rho)</td>
<td>Persistence in ability</td>
<td>0.362</td>
</tr>
<tr>
<td>(\sigma^2)</td>
<td>Variance in ability</td>
<td>0.570</td>
</tr>
<tr>
<td>(\xi^U)</td>
<td>Returns to HC investment – UK</td>
<td>0.197</td>
</tr>
<tr>
<td>(\xi^C)</td>
<td>Returns to HC investment – CE</td>
<td>0.253</td>
</tr>
<tr>
<td>(\nu^U)</td>
<td>Skill production technology – UK</td>
<td>1.000</td>
</tr>
<tr>
<td>(\nu^C)</td>
<td>Skill production technology – CE</td>
<td>1.145</td>
</tr>
<tr>
<td>(\sigma^2)</td>
<td>Variance in labor productivity shock</td>
<td>1.076</td>
</tr>
<tr>
<td>(\eta)</td>
<td>Weight in disutility of labor</td>
<td>777</td>
</tr>
<tr>
<td>(\gamma)</td>
<td>Frisch elasticity</td>
<td>0.3</td>
</tr>
<tr>
<td>(\beta)</td>
<td>Discount (altruism) factor</td>
<td>0.948</td>
</tr>
<tr>
<td>(h_U)</td>
<td>Minimum utility cost of moving – unskilled</td>
<td>89</td>
</tr>
<tr>
<td>(h_S)</td>
<td>Minimum utility cost of moving – skilled</td>
<td>59</td>
</tr>
<tr>
<td>(h_U)</td>
<td>Scale of moving cost distribution – unskilled</td>
<td>9800</td>
</tr>
<tr>
<td>(h_S)</td>
<td>Scale of moving cost distribution – skilled</td>
<td>3200</td>
</tr>
</tbody>
</table>
Table 1.4: Model Fit

<table>
<thead>
<tr>
<th>Moments Data Model</th>
<th>Data</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skill premium – UK</td>
<td>1.62</td>
<td>1.62</td>
</tr>
<tr>
<td>Skill premium – CE</td>
<td>1.45</td>
<td>1.44</td>
</tr>
<tr>
<td>GDP per capita ratio</td>
<td>1.00</td>
<td>0.96</td>
</tr>
<tr>
<td>Capital-output ratio</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Annual bond yields</td>
<td>1.04</td>
<td>1.04</td>
</tr>
<tr>
<td>Intergenerational persistence in schooling – UK</td>
<td>0.22</td>
<td>0.20</td>
</tr>
<tr>
<td>Gini coefficient of labor earning – UK</td>
<td>0.45</td>
<td>0.45</td>
</tr>
<tr>
<td>% college graduates – UK</td>
<td>25.7%</td>
<td>25.2%</td>
</tr>
<tr>
<td>% college graduates – CE</td>
<td>20.0%</td>
<td>20.1%</td>
</tr>
<tr>
<td>Tertiary education spending (% GDP) – UK</td>
<td>0.78%</td>
<td>0.74%</td>
</tr>
<tr>
<td>Tertiary education spending (% GDP) – CE</td>
<td>1.06%</td>
<td>1.05%</td>
</tr>
<tr>
<td>Average hours worked – UK</td>
<td>0.36</td>
<td>0.35</td>
</tr>
<tr>
<td>Population ratio – UK to CE</td>
<td>0.94</td>
<td>0.97</td>
</tr>
<tr>
<td>% of immigrants among Pop. – UK</td>
<td>1.1%</td>
<td>1.1%</td>
</tr>
<tr>
<td>% college graduates among immigrants – UK</td>
<td>38%</td>
<td>34%</td>
</tr>
<tr>
<td>% college graduates among immigrants – CE</td>
<td>36%</td>
<td>41%</td>
</tr>
</tbody>
</table>

Table 1.5: Effects of Changes in the Tax Code – United Kingdom

<table>
<thead>
<tr>
<th>Variable</th>
<th>$\tau_k^{UK}$ ↓</th>
<th>$a_0^{UK}(y)$ ↑</th>
<th>$a_1^{UK}(y)$ ↓</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capital</td>
<td>0.038</td>
<td>0.006</td>
<td>0.015</td>
</tr>
<tr>
<td>Hours worked</td>
<td>0.000</td>
<td>0.002</td>
<td>0.005</td>
</tr>
<tr>
<td>Skilled immigrants</td>
<td>-0.001</td>
<td>0.000</td>
<td>-0.002</td>
</tr>
<tr>
<td>Consumption equivalent variation</td>
<td>0.004</td>
<td>0.001</td>
<td>-0.003</td>
</tr>
<tr>
<td>Unskilled</td>
<td>0.003</td>
<td>0.000</td>
<td>-0.003</td>
</tr>
<tr>
<td>Skilled</td>
<td>0.007</td>
<td>0.004</td>
<td>0.019</td>
</tr>
</tbody>
</table>

Note: Starting from the status quo tax system in the United Kingdom, I vary each component in the tax system (capital income tax rate, average labor income tax rate, and progressivity of the labor income tax code) one by one. Since labor income tax function is

$$\tau(y; \bar{y}) = 1 - a_0 \left( \frac{y}{\bar{y}} \right)^{-a_1},$$

increasing $a_0$ is equivalent to lowering the average labor income tax rate, while decreasing $a_1$ represents a less progressive labor income tax code. All variables represent the changes of the aggregate outcomes from the status quo tax system in the United Kingdom, taking as given the status quo tax system in the Continental Europe.
Table 1.6: Unilateral Reform of the United Kingdom and Aggregate Outcomes

<table>
<thead>
<tr>
<th>Tax code</th>
<th>Status Quo</th>
<th>Unilateral</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>UK</td>
<td>CE</td>
</tr>
<tr>
<td>Capital income tax</td>
<td>0.48</td>
<td>0.28</td>
</tr>
<tr>
<td>Average labor income tax</td>
<td>0.24</td>
<td>0.31</td>
</tr>
<tr>
<td>Progressivity</td>
<td>0.07</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Aggregate outcome

<table>
<thead>
<tr>
<th>Status Quo</th>
<th>Unilateral</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>UK</td>
</tr>
<tr>
<td>Output per capita</td>
<td>0.27</td>
</tr>
<tr>
<td>Capital per capita</td>
<td>1.42</td>
</tr>
<tr>
<td>Average hours worked</td>
<td>-0.08</td>
</tr>
<tr>
<td>Return to capital</td>
<td>-0.42</td>
</tr>
<tr>
<td>After-tax return to capital</td>
<td>0.19</td>
</tr>
<tr>
<td>Wages, unskilled</td>
<td>0.37</td>
</tr>
<tr>
<td>Wages, skilled</td>
<td>0.38</td>
</tr>
</tbody>
</table>

Note: Aggregate variables are changes from the status quo tax system.

Table 1.7: Unilateral Reform of Continental Europe and Aggregate Outcomes

<table>
<thead>
<tr>
<th>Tax code</th>
<th>Status Quo</th>
<th>Unilateral</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>UK</td>
<td>CE</td>
</tr>
<tr>
<td>Capital income tax</td>
<td>0.48</td>
<td>0.28</td>
</tr>
<tr>
<td>Average labor income tax</td>
<td>0.24</td>
<td>0.31</td>
</tr>
<tr>
<td>Progressivity</td>
<td>0.07</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Aggregate outcome

<table>
<thead>
<tr>
<th>Status Quo</th>
<th>Unilateral</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>UK</td>
</tr>
<tr>
<td>Output per capita</td>
<td>0.08</td>
</tr>
<tr>
<td>Capital per capita</td>
<td>0.35</td>
</tr>
<tr>
<td>Average hours worked</td>
<td>-0.04</td>
</tr>
<tr>
<td>Return to capital</td>
<td>-0.21</td>
</tr>
<tr>
<td>After-tax return to capital</td>
<td>0.12</td>
</tr>
<tr>
<td>Wages, unskilled</td>
<td>0.12</td>
</tr>
<tr>
<td>Wages, skilled</td>
<td>0.12</td>
</tr>
</tbody>
</table>

Note: Aggregate variables are changes from the status quo tax system.
### Table 1.8: Aggregate Outcomes in the Nash Equilibrium

<table>
<thead>
<tr>
<th>Variable</th>
<th>Nash Equilibrium</th>
<th>United Kingdom</th>
<th>Continental Europe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output per capita</td>
<td>0.21</td>
<td>-0.11</td>
<td></td>
</tr>
<tr>
<td>Capital per capita</td>
<td>1.18</td>
<td>-0.17</td>
<td></td>
</tr>
<tr>
<td>Average hours worked</td>
<td>-0.12</td>
<td>-0.13</td>
<td></td>
</tr>
<tr>
<td>Consumption per capita</td>
<td>0.04</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>Return to capital</td>
<td>-0.40</td>
<td>-0.02</td>
<td></td>
</tr>
<tr>
<td>After-tax return to capital</td>
<td>0.21</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td>% skilled</td>
<td>-0.02</td>
<td>-0.01</td>
<td></td>
</tr>
<tr>
<td>Labor income Gini coefficient</td>
<td>0.04</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>Schooling persistence</td>
<td>-0.14</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Population</td>
<td>0.04</td>
<td>-0.04</td>
<td></td>
</tr>
<tr>
<td>Skilled immigrants</td>
<td>-0.03</td>
<td>-0.03</td>
<td></td>
</tr>
<tr>
<td>Consumption equivalent variation</td>
<td>0.13</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>Unskilled</td>
<td>0.14</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>Skilled</td>
<td>0.07</td>
<td>0.06</td>
<td></td>
</tr>
</tbody>
</table>

*Note: All variables represent changes from the status quo tax system.*
Table 1.9: Effects of Labor Mobility on Optimal Taxation in the United Kingdom

<table>
<thead>
<tr>
<th>Optimal Tax Code</th>
<th>Closed</th>
<th>Capital Mobility</th>
<th>Capital &amp; Labor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capital income tax rate</td>
<td>-0.26</td>
<td>-0.12</td>
<td>-0.32</td>
</tr>
<tr>
<td>Average labor income tax rate</td>
<td>0.51</td>
<td>0.49</td>
<td>0.52</td>
</tr>
<tr>
<td>Progressivity</td>
<td>0.30</td>
<td>0.25</td>
<td>0.28</td>
</tr>
</tbody>
</table>

*Note: The tax codes are based on the optimal taxation in a closed economy, in an economy with only capital mobility, and in an economy with both capital and labor mobility.*

Table 1.10: Effects of Labor Mobility on Optimal Taxation in the Continental Europe

<table>
<thead>
<tr>
<th>Optimal Tax Code</th>
<th>Closed</th>
<th>Capital Mobility</th>
<th>Capital &amp; Labor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capital income tax rate</td>
<td>-0.15</td>
<td>-0.11</td>
<td>0.06</td>
</tr>
<tr>
<td>Average labor income tax rate</td>
<td>0.35</td>
<td>0.51</td>
<td>0.50</td>
</tr>
<tr>
<td>Progressivity</td>
<td>0.28</td>
<td>0.25</td>
<td>0.11</td>
</tr>
</tbody>
</table>

*Note: The tax codes are based on the optimal taxation in a closed economy, in an economy with only capital mobility, and in an economy with both capital and labor mobility.*

Table 1.11: Effects of Changes to the Capital Income Tax Rate in the United Kingdom under Different Mobility Assumptions

<table>
<thead>
<tr>
<th>Variable</th>
<th>Closed</th>
<th>Capital Mobility</th>
<th>Capital &amp; Labor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capital</td>
<td>0.012</td>
<td>0.006</td>
<td>0.012</td>
</tr>
<tr>
<td>Return to capital</td>
<td>-0.003</td>
<td>0.003</td>
<td>0.000</td>
</tr>
<tr>
<td>Average hours worked</td>
<td>-0.002</td>
<td>-0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>Wages, unskilled</td>
<td>0.005</td>
<td>0.003</td>
<td>0.004</td>
</tr>
<tr>
<td>Wages, skilled</td>
<td>0.005</td>
<td>0.002</td>
<td>0.002</td>
</tr>
</tbody>
</table>

*Note: The variables represent the changes in aggregate outcomes under the optimal tax code and outcomes under a capital income tax rate that is one percent lower than the optimal capital income tax rate (fixing the labor income tax code).*
Table 1.12: Effects of Changes to Tax Codes in the United Kingdom and Continental Europe at Nash Equilibrium

<table>
<thead>
<tr>
<th>Variable</th>
<th>United Kingdom</th>
<th>Continental Europe</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\tau_y^{UK}$</td>
<td>$\tau_k^{UK}$</td>
</tr>
<tr>
<td>Capital</td>
<td>0.001</td>
<td>0.012</td>
</tr>
<tr>
<td>Consumption</td>
<td>0.007</td>
<td>0.006</td>
</tr>
<tr>
<td>Consumption equivalent variation</td>
<td>0.003</td>
<td>0.006</td>
</tr>
</tbody>
</table>

*Note:* The variables represent the changes from aggregate outcomes under the optimal tax code and outcomes under a labor (capital) income tax rate that is one percent lower than the optimal capital income tax rate, fixing the labor (capital) income tax code.

Table 1.13: Effects of Labor Mobility on Aggregate Outcomes

<table>
<thead>
<tr>
<th>Variable</th>
<th>United Kingdom</th>
<th>Continental Europe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output per capita</td>
<td>0.079</td>
<td>-0.074</td>
</tr>
<tr>
<td>Capital per capita</td>
<td>0.268</td>
<td>-0.219</td>
</tr>
<tr>
<td>Consumption per capita</td>
<td>-0.010</td>
<td>0.057</td>
</tr>
<tr>
<td>Unskilled labor</td>
<td>-0.001</td>
<td>0.002</td>
</tr>
<tr>
<td>Skilled labor</td>
<td>0.002</td>
<td>0.005</td>
</tr>
<tr>
<td>% skilled</td>
<td>0.001</td>
<td>-0.002</td>
</tr>
<tr>
<td>Labor income Gini coefficient</td>
<td>-0.002</td>
<td>0.013</td>
</tr>
<tr>
<td>Population</td>
<td>0.014</td>
<td>-0.014</td>
</tr>
</tbody>
</table>

*Note:* All variables represent changes from the optimal tax system with only capital mobility.
Parents in country $i$ with $(z, \theta, k, b, \varepsilon)$ work $l$ hours, consume $c$, invest in human capital of child $x$, bequeath $k', b'$, and make migration decisions (move if $V^i(\cdot) < V^j(\cdot) - h$).

Figure 1.2: Average Labor Income Tax Rates in the Continental Europe and United Kingdom

Note: The United Kingdom has a less progressive labor income tax code, with a lower average labor income tax rate, when compared to the Continental Europe.
Figure 1.3: Human Capital Investment Decisions of Households

Note: Households living in the United Kingdom have higher incentives to invest in the human capital of their offspring conditional on skill level, thanks to the country’s high skill premium and low labor income tax rate. Moreover, public education spending in the United Kingdom is lower, leaving more room for investment in children. On the other hand, investment by households in the Continental Europe is lower, since public education expenditures are high and the market incentives are lower (low skill premium and high labor income tax rates). The skilled parents invest more (conditional on assets) than the unskilled parents, since they have higher labor income.
Figure 1.4: Differences in the Value of Living in the United Kingdom versus Continental Europe

Note: The y-axis represents the differences in the value of living in the Continental Europe and living in the United Kingdom. A positive value implies that a household of a given characteristic prefers living in the Continental Europe. Overall, the skilled labor force prefers the United Kingdom, while households whose offspring is of low ability prefer living in the Continental Europe. The differences between living in the United Kingdom and Continental Europe vanish as assets increase.

Figure 1.5: Average Labor Income Tax Rates in the Nash Equilibrium

Note: In the Nash equilibrium, the United Kingdom uses a more progressive labor income tax code than the Continental Europe does.
Figure 1.6: Progressivity Wedge in the *Status Quo* and in the Nash Equilibrium

*Note:* Following Guvenen et al. (2011), progressivity wedge is defined as $1 - \frac{1 - \tau(k \times 0.5)}{1 - \tau(0.5)}$ for $k = 2, 3, ... 5$. The interpretation is as follows: In the Nash equilibrium, a household living in the United Kingdom, whose labor earning is two times the average earning in the economy, earns 25 percent less than in a flat-tax system. An analogous worker earns 12 percent less, if he resides in the Continental Europe. In the *status quo*, the wedges are 8 and 10 percent respectively, in the United Kingdom and Continental Europe.
Chapter 2

Analyzing the Effects of Insuring Health Risks:
On the Tradeoff between Short-Run Insurance Benefits vs. Long-Run Incentive Costs

This chapter is co-authored with Harold L. Cole and Dirk Krueger.

Summary

This paper constructs a dynamic model of health insurance to evaluate the short- and long run effects of policies that prevent firms from conditioning wages on health conditions of their workers, and that prevent health insurance companies from charging individuals with adverse health conditions higher insurance premia. Our study is motivated by recent US legislation that has tightened regulations on wage discrimination against workers with poorer health status (Americans with Disability Act of 2009, ADA, and ADA Amendments Act of 2008, ADAAA) and that will prohibit health insurance companies from charging different premiums for workers of different health status starting in 2014 (Patient Protection and Affordable Care Act, PPACA). In the model, a trade-off arises between the static gains from better insurance against poor health induced by these policies and their adverse dynamic incentive effects on household efforts to lead a healthy life. Using household panel
data from the PSID we estimate and calibrate the model and then use it to evaluate the static and dynamic consequences of no-wage discrimination and no-prior conditions laws for the evolution of the cross-sectional health and consumption distribution of a cohort of households, as well as ex-ante lifetime utility of a typical member of this cohort. In our quantitative analysis we find that although a combination of both policies is effective in providing full consumption insurance period by period, it is suboptimal to introduce both policies jointly since such policy innovation induces a more rapid deterioration of the cohort health distribution over time. This is due to the fact that combination of both laws severely undermines the incentives to lead healthier lives. The resulting negative effects on health outcomes in society more than offset the static gains from better consumption insurance so that expected discounted lifetime utility is lower under both policies, relative to only implementing wage nondiscrimination legislation.

2.1 Introduction

In this paper we study the impact of social insurance policies aimed at reducing households’ exposure to health related risk during their working life. These risks come through higher medical costs, higher medical premiums and lower earnings. Historically there have been major insurance efforts aimed at the elderly through Medicaid and Social Security, and the poor through Medicare and income support programs like Welfare and Food Stamps. Recently the extent of these programs and the scope of the different groups they impact on has been greatly expanded. On the health insurance front, HIPPA in 1966 and the Patient Protection and Affordable Care Act in 2010 sought to increase access to health care and to prevent health insurance being differentially priced based upon pre-existing conditions. On the income front, the 1990 Americans with Disabilities Act and its Amendment in 2009 sought to restrict the ability of employers to employ and compensate workers differentially based upon health related reasons.

In order to analyze the impact of these policies we construct a dynamic model of health insurance with heterogeneous households. As in Grossman (1972), health for these house-
holds is a state variable. A household’s health state helps to determine both their productivity at work and the likelihood that they will be subject to adverse health shocks. Our model features the two-way interaction between health and income that has been emphasized in the literature. Our model of health shocks includes temporary health shocks that impact on productivity and can be offset by medical expenditures (as in Dey and Flinn 2005), and catastrophic health shocks which require nondiscretionary health expenditures to avoid death. Health status in our model is persistent and evolves stochastically. This evolution is affected by the household’s efforts to maintain their health which results in a moral hazard problem as health related insurance reduces households’ incentives to maintain their health. We explicitly model the choice of medical expenditure and thereby endogenously determine the health insurance policy and how it responds both to the household’s state in terms of health status, age and education.

The focus of our analysis is how the distributions of health status, earnings and health insurance costs will evolve under different policy choices and the impact of these choices on welfare. We consider several different policy regimes. The first is a complete insurance benchmark in which the social planner can dictate both the health insurance contract, the effort made to maintain health and the extent of redistributive transfers that provide full insurance against all health related shocks. The second is pure competition in which workers enter into one-period employment and insurance contracts. Competition leads these contracts to partially insure the worker against within period temporary health shocks, but not against his initial health status and the transition of this status. The second is a version of the no-prior conditions restriction on health insurance in which health insurance companies compete to offer one-period health insurance contracts in which they cannot differentially charge based upon the worker’s health status. The third is a version of the no-discrimination restrictions on employment in which firms cannot differentially hire or pay workers based upon their health status. In the fourth version we consider the impact of both the no-prior conditions and the no-discrimination restrictions jointly.

We study both the static and the dynamic impact of these policies. One of the key
aspects of the dynamic analysis is the impact these policies have on individuals’ incentives to maintain their health and the feedback this creates between the health distribution of the population and the costs of health insurance and productivity of the workforce.

We evaluate the quantitative impacts of the different policies on consumption insurance, incentives and aggregate outcomes, and, ultimately, welfare. To do so, we first estimate and calibrate the model using PSID data to match key aggregate statistics on labor earnings, medical expenditures and observed physical exercise levels. We then use the parameterized version of the model as a laboratory to evaluate different policy scenarios. Our results show that a combination of wage non-discrimination law and no prior conditions law provides full insurance against health risks and restores the first-best consumption insurance allocation in the short run, but leads to a severe deterioration of incentives and thus the population health distribution in the long run. Quantitatively evaluating the welfare consequences of this trade-off we find that even though both policies improve upon the laissez-faire equilibrium, implementing them jointly is suboptimal, relative to introducing a wage nondiscrimination in isolation.

2.1.1 Institutional Background

The U.S. has a long history of policy initiatives in relation to health risk. Implicitly Welfare programs, which date back to the 1930s and were greatly expanded by the Great Society in the 1960s, insure workers against a variety of shocks, implicitly including health related shocks insofar they affect earnings. Since 1965 Medicare has sought to provide health insurance to the elderly and the disabled. Medicaid has sought to provide health insurance to the poor since the 1990s. The last two decades legislation in the U.S. was passed that limits the ability of employers to condition wages on the health conditions of employees, and to discriminate against applicants with prior health conditions when filling vacant positions.
Wage Based Discrimination

In 1990 Congress enacted the Americans with Disabilities Act (ADA) to ensure that the disabled have equal access to employment opportunities. At this point a disability was interpreted as an impairment that prevents or severely restricts an individual from doing activities that are of central importance to one’s daily life. In 2009 the ADA Amendments Act (ADAAA) went into effect. This act rejected the strict interpretation of the ADA, broadening the notion of a disability. This included prohibiting the consideration of measures that reduce or mitigate the impact of a disability in determining whether someone is disabled. It also allowed people who are discriminated against on the basis of a perceived disability to pursue a claim on the basis of the ADA regardless of whether the perceived disability limits or is perceived to limit a major life activity. The ADAAA excludes from the definition of a disability those temporary or minor impairments. Under the ADAAA people can be disabled even if their disability is episodic or in remission. For example people whose cancer is remission or whose diabetes is controlled by medication, or whose seizures are prevented by medication, or who can function at a high level with learning disabilities are all disabled under the act.

Before the ADA job seekers could be asked about their medical conditions and were often required to submit to a medical exam. The act prohibited certain inquiries and conducting a medical exam before making an employment offer. However, the job could still be conditioned upon successful completion of a medical exam.

The ADA permits an employer to establish job-related qualifications on the basis of business necessity. However, business necessity is limited to essential functions of the job. So impairments that would only occasionally interfere with the employee’s ability to perform tasks cannot be included on this list. A job function is essential if the job exists to perform

---

1 The ADA sets the federal minimum standard of protection. States may have a more stringent level.
2 Under the ADAAA major life activities now include: caring for oneself, performing manual tasks, seeing, hearing, eating, sleeping, walking, standing, lifting, bending, speaking, breathing, learning, reading, concentrating, thinking, communicating, working, as well as major bodily functions.
3 For example, the Equal Employment Opportunity Commission (EEOC) has ruled that an employee may be asked "how many days were you absent from work?", but not "how many days were you sick?".
4 For example, an employer cannot require a driver’s license for a clerking job because it would occasionally be useful to have that employee run errands. Also qualification cannot be such that a reasonable
that function or if the limited number of employees available at the firm requires that the task must be performed by this worker. Furthermore, a core requirement of the ADA is the obligation of the employer to make a reasonable accommodation to qualified disabled people.\(^5\)

**Insurance Cost and Exclusion Discrimination**

In 1996, Congress passed the Health Insurance Portability and Accountability Act (HIPAA) which placed limits on the extent to which insurance companies could exclude people or deny coverage based upon pre-existing conditions. Although insurance companies were allowed exclusions periods for coverage of pre-existing conditions, these exclusion periods were reduced by the extent of prior insurance. In particular, if an individual had at least a full year of prior health insurance and she enrolled in a new plan with a break of less than 63 days, she could not be denied coverage. However, insurers were still allowed to charge higher premiums based upon initial conditions, limit coverage and set lifetime limits on benefits.\(^6\) There is evidence that many patients with pre-existing conditions ended up either being denied coverage,\(^7\) or having their access to benefits limited.\(^8\)

The Patient Protection and Affordable Care Act of 2010 further extended protection against pre-existing conditions. Beginning in 2010 children below the age of 19 could not be excluded from their parents’ health insurance policy or denied treatment for pre-existing conditions. Beginning in 2014 this restriction will apply to adults as well. Moreover, insurance companies will no longer be able to use health status to determine eligibility, benefits or premia. In addition, insurers will be prevented from limiting lifetime or annual benefits or from taking away coverage because of an application mistake.\(^9\)

---

\(^5\) These accommodations include: a) making existing facilities accessible and usable b) job restructuring c) part-time or modified work schedules d) reassigning a disabled employee to a vacant position e) acquiring or modifying equipment or devices f) providing qualified readers or interpreters.

\(^6\) See http://www.healthcare.gov/center/reports/preexisting.html

\(^7\) See Kass et al. (2007).

\(^8\) See Sommers (2006).

\(^9\) See again http://www.healthcare.gov/center/reports/preexisting.html
Summary

It is our interpretation of these legislative changes that, relative to 20 years ago, it is much more difficult now for employers to condition wages on the health status of their (potential) employees and preferentially hire workers with better health. In addition, current and pending legislation will make it increasingly difficult to condition the acceptance into, and insurance premia of health insurance plans on prior health conditions.

The purpose of the remainder of this paper is to analyze the aggregate and distributional consequences of these two legislative innovations in the short and in the long run, with specific focus on their interactions.

2.1.2 Related Literature

Our paper incorporates health as a productive factor, and studies the effect of labor and health insurance market policies on its evolution. The dynamics of health transitions (but not current health status) in our model is affected by costly effort choices, which empirically we proxy by physical exercise. Colman and Dave (2013) provide empirical support for this modeling choice when they find that, controlling for observables and accounting for unobserved heterogeneity, physical activity reduces the risk of heart disease, and that the effect of past physical activity has a larger impact on current health status than current physical activity.

In the model we allow for a two-way interaction between health shocks and earnings through worker productivity. We model medical expenditures which mitigate the impact of these health shocks. There have been a number of studies that empirically estimate the effect of health on wages. These papers (see the summary in Currie and Madrian, 1999) generally find that poor health decreases wages, both directly and indirectly through a decrease in hours worked. The effect of a health shock on wages ranges from 1% to as high as 15%. Many studies consistently find that the effects on hours worked is greater than that on wages. Specifically relevant for us is Cawley (2004). In addition, there is a substantial empirical literature that estimates the effects of health inputs (and primarily, nutrition) on
health outcomes and wages in developing countries. Weil (2007) surveys this literature and uses its findings to determine the impact of country-wide health differences on cross-country income variation.

Similarly to what we do for working age individuals, Pijoan-Mas and Rios-Rull (2012), use HRS data on self-report health status to estimate a health transition function from age 50 onwards. They find that there is an important dependence in this transition function on socioeconomic status (most importantly education), and that this dependence is quantitatively crucial for explaining longevity differentials by socioeconomic groups. As we do Hai (2012) and Prados (2012) model the interaction between health and earnings over the life cycle, but focus on the implications of their models for wage-, earnings- and health insurance inequality.10

A relatively small literature examines the incentive linkages between health insurance and health status. Bhattacharya et al. (2009) use evidence from a Rand health insurance experiment, which featured randomized assignment to health insurance contracts, to show that access to health insurance leads to increases in body mass and obesity. They argue that this comes from the fact that insurance, especially through its pooling effect, insulates people from the impact of their excess weight on their medical expenditure costs. Consistent with this, they find the impact of being health-insured is larger for public insurance programs than in private ones in which the health insurance premium is more likely to reflect the individuals’ body mass.11

This paper contributes to the broad literature that examines the macroeconomic and distributional implications of health, health insurance and health care policy reform. Important related contributions include Grossman (1972), Ehrlich and Becker (1972), Ehrlich and Chuma (1990), French and Jones (2004), Hall and Jones (2007), Jeske, and Kitao (2009), Jung and Tran (2010), Attanasio, Kitao and Violante (2011), Ales, Hosseini and Jones (2012), Halliday, He and Zhang (2012), Hansen, Hsu and Lee (2012), Kopecky and

10Both papers also study the impact of compulsory health insurance legislation.
11Kowalski (2012) empirically investigates, in the context of a static model, the trade-off between insurance and the moral hazard effects of health insurance provision on medical spending, finding that the latter outweigh the former.
Koreshkova (2012), Laun (2012), Ozkan (2012), Pashchenko and Porapakkam (2012) and Scholz and Seshadri (2012). Brügemann and Manovskii (2010), while endogenizing health, study the macroeconomic effects of the employer-sponsored health insurance system that is unique to the US labor market. Concretely, they determine the effect of PPACA on health insurance coverage, but do not study the incentive effects of the regulation that we formalize in our model.

Several papers investigate the impact of regulation designed to limit the direct effect of health on both health insurance costs and on wages. Short and Lair (1994) examine how health status interacts with insurance choices. Madrian (1994) studies the lock-in effect of employer provided health care. Dey and Flinn (2005) estimate a model of health insurance with search, matching and bargaining and argue that employer provided health care insurance leads to reasonably efficient outcomes.

Related to our study of wage non-discrimination laws is the literature that studies the effect of the ADA legislation of 1990 on employment, wages and labor hours of the disabled (see DeLeire (2001) and Acemoglu and Angrist (2001), for example). Most find that it has decreased the employment of the disabled. DeLeire (2001) quantifies the effect of ADA on wages of disabled workers and reports that the negative effect of poor health on the earnings of the disabled fell by 11.3% due to ADA.

Finally, a recent literature examines the impact of health on savings and portfolio choice in life cycle models that share elements with our framework. These include Yogo (2009), Edwards (2008) and Hugonnier et al. (2012). The latter study jointly portfolio of health and other asset choices. In their model health increases productivity (labor income) and decreases occurrence of morbidity and mortality shock arrival rates (as they do in our model). The paper argues that in order to explain the correlation between financial and health status, these should be modelled jointly.

---

12 A mostly empirical literature studies the impact of the 2006 health care reform in Massachusetts. See e.g. Miller (2011).
2.2 The Model

Time $t = 0, 1, 2, \ldots, T$ is discrete and finite and the economy is populated by a cohort of a continuum of individuals of mass 1. Since we are modeling a given cohort of individuals we will use time and the age of households interchangeably. We think of $T$ as the end of working life of the age cohort under study.

2.2.1 Endowments and Preferences

Households are endowed with one unit of time which they supply inelastically to the market. They are also endowed with an initial level of health $h$ and we denote by $H = \{h_1, \ldots, h_N\}$ the finite set of possible health levels. Households value current consumption $c$ and dislike the effort $e$ that helps maintain their health. We will assume that their preferences are additively separable over time, and that they discount the future at time discount factor $\beta$. We will also assume that preferences are separable between consumption and effort, and that households value consumption according to the common period utility function $u(c)$ and value effort according to the period disutility function $q(e)$.

We will denote the probability distribution over the health status $h$ at the beginning of period $t$ by $\Phi_t(h)$, and denote by $\Phi_0(h)$ the initial distribution over this characteristic.

**Assumption 1.** The utility function $u$ is twice differentiable, strictly increasing and strictly concave. $q$ is twice differentiable, strictly increasing, strictly convex, with $q(0) = q'(0) = 0$ and $\lim_{e \to \infty} q'(e) = \infty$.

2.2.2 Technology

Health Technology

Let $\varepsilon$ denote the current health shock.\footnote{In the quantitative analysis we will introduce a second, fully insured (by assumption) health shock to provide a more accurate map between our model and the health expenditure data.} In every period households with current health $h$ remain healthy (that is, $\varepsilon = 0$) with probability $g(h)$. With probability $1 - g(h)$ the
household draws a health shock $\varepsilon \in (0, \bar{\varepsilon}]$ which is distributed according to the probability density function $f(\varepsilon)$.

**Assumption 2.** $f$ is continuous and $g$ is twice differentiable with $g(h) \in [0, 1]$, and $g'(h) > 0, g''(h) < 0$ for all $h \in H$.

An individual’s health status evolves stochastically over time, according to the Markov transition function $Q(h', h; e)$, where $e \geq 0$ is the level of exercise by the individual. We impose the following assumption on the Markov transition function $Q$

**Assumption 3.** If $e' > e$ then $Q(h', h; e)$ first order stochastically dominates $Q(h', h; e')$.

**Production Technology**

A individual with health status $h$ and current health shock $\varepsilon$ that consumes health expenditures $x$ produces $F(h, \varepsilon - x)$ units of output.

**Assumption 4.** $F$ is continuously differentiable in both arguments, increasing in $h$, and satisfies $F(h, y) = F(h, 0)$ for all $y \leq 0$, and $F_2(h, y) < 0$ as well as $F_2(h, \bar{\varepsilon}) < -1$. Finally $F_{22}(h, y) < 0$ for all $y > 0$ and $F_{12}(h, y) \geq 0$.

The left panel of figure A.2 displays the production function $F(h, \cdot)$, for two different levels of the current health shock. Holding health status $h$ constant, output is decreasing in the uncured portion of the health shock $\varepsilon - x$, and the decline is more rapid for lower levels of health ($h^* < h$). The right panel of figure 2.1 displays the production function as function of health expenditures $x$, for a fixed level of the shock $\varepsilon$, and shows that expenditures $x$ exceeding the health shock $\varepsilon$ leave output $F(h, \varepsilon - x)$ unaffected (and thus are suboptimal). Furthermore, a reduction of the shock $\varepsilon$ to a lower level, $\varepsilon^*$, shifts the point at which health expenditures $x$ become ineffective to the left.

The assumptions on the production function $F$ imply that health expenditures can offset the impact of a health shock on productivity, but not raise an individual’s productivity above what it would be if there had been no shock. In addition, the last assumption on $F$ that $F_{12} \geq 0$ implies that the negative impact of a given net health shock $y$ is lower the healthier
a person is.\footnote{This is also the approach taken by Hugonnier et al. (2012) and Ehrlich and Chuma (1990).} The assumption $F_2(h, \tilde{\varepsilon}) < -1$ insures that, if hit by the worst health shock the cost of treating this health shock, at the margin, is smaller than the positive impact on productivity (output) this treatment has.

2.2.3 Time Line of Events

In the current period the timing of events is as follows

1. Households enter the period with current health status $h$.

2. Households choose $e$.

3. Firms offer wage $w(h)$ and health insurance contracts $\{x(\varepsilon, h), P(h)\}$\footnote{Since we restrict attention to static contracts, whether firm offers contracts before or after the effort is undertaken does not matter.} to households with health status $h$ which these households accept.

4. The health shock $\varepsilon$ is drawn according to the distributions $g, f$.

5. Resources on health $x = x(\varepsilon, h)$ are spent.

6. Production and consumption takes place.

7. The new health status $h'$ of a household is drawn according to the health transition function $Q$.

2.2.4 Market Structure without Government

There are a large number of production firms that in each period compete for workers. Firms observe the health status of a worker $h$ and then, prior to the realization of the health shocks, compete for workers of type $h$ by offering a wage $w(h)$ that pools the risk of the health shocks and bundle the wage with an associated health insurance contract (specifying health expenditures $x(\varepsilon, h)$ and an insurance premium $P(h)$) that breaks even. Perfect competition for workers of type $h$ requires that the combined wage and health
insurance contract maximize period utility of the household, subject to the firm breaking even.\textsuperscript{16}

In the absence of government intervention a firm specializing on workers of health type $h$ therefore offers a wage $w^{CE}(h)$ (where $CE$ stands for competitive equilibrium) and health insurance contract $(x^{CE}(\epsilon, h), P^{CE}(h))$ that solves

$$U^{CE}(h) = \max_{w(h),x(\epsilon,h),P(h)} u(w(h) - P(h))$$ (2.1)

s.t.

$$P(h) = g(h)x(0, h) + (1 - g(h)) \int_0^{\epsilon} f(\epsilon)x(\epsilon, h)d\epsilon$$ (2.2)

$$w(h) = g(h)F(h, -x(0, h)) + (1 - g(h)) \int_0^{\epsilon} f(\epsilon)F(h, \epsilon - x(\epsilon, h))d\epsilon$$ (2.3)

Note that by bundling wages and health insurance the firm provides efficient insurance against health shocks $\epsilon$, and the only source of risk remaining in the competitive equilibrium is health status risk associated with $h$. This risk stems both from the dependence of wages $w(h)$ as well as health insurance premia $P(h)$ on $h$ in the competitive equilibrium, and these are exactly the sources of consumption risk that government policies preventing wage discrimination and prohibiting prior health conditions to affect insurance premia are designed to tackle.

### 2.2.5 Government Policies

We now describe in turn how we operationalize, within the context of our model, a policy that outlaws health insurance premia to be conditioned on prior health conditions $h$, and a policy that limits the extent to which firms can pay workers of varying health $h$ differential wages.

\textsuperscript{16}Note that instead of assuming that firms completely specialize by hiring only a specific health type of workers $h$ we could alternatively consider a market structure in which all firms are representative in terms of hiring workers of health types according to the population distribution and pay workers of different health $h$ differential wages according to the schedule $w^{CE}(h)$. In other words health variation in wages and variation in hired health types $h$ are perfect substitutes at the level of the individual firm in terms of supporting the competitive equilibrium allocation.
No Prior Conditions Law

Under this law health insurance companies are assumed to be constrained in terms of their pricing, their insurance schedule offers and their applicant acceptance criteria. The purpose of these constraints is to prevent the companies from differentially pricing insurance based upon health status.\textsuperscript{17} To be completely successful, these constraints must lead to a pooling equilibrium in which all individuals are insured at the same price. The best such regulation in addition assures that the equilibrium health insurance schedule \( x(\varepsilon, h) \), given the constraints, is efficient. We now describe the regulations sufficient to achieve this goal.

The first constraint on health insurers is that a company must specify the total number of contracts that it wishes to issue, it must charge a fixed price independent of health status, and accept applications in their order of application up to the sales limit of the company. In this way, the insurance company cannot examine applications first and then decide whether or not to offer the applicant a health insurance contract.

The second constraint regulates the health expenditure schedule. If the no-prior conditions law is to have any bite the government needs to prevent the emergence of a separating equilibrium in which the health insurance companies (or the production firms in case they offer health insurance contracts) use the health expenditure schedule \( x(\varepsilon, h) \) to effectively select the desired health types, given that they are barred from conditioning the health insurance premium \( P \) on \( h \) directly. Therefore, to achieve any sort of pooling in the health insurance market requires the government to regulate the health expenditure schedule \( x(\varepsilon, h) \). To give the legislation the best chance of being successful we will assume that the government \textit{regulates the health expenditure schedule} \( x(\varepsilon, h) \) \textit{efficiently}. For the same reason, since risk pooling is limited if some household types \( h \) choose not to buy insurance, we assume that all individuals are \textit{forced} to buy insurance.

Given this structure of regulation and a cross-sectional distribution of workers by health type, \( \Phi \), the health insurance premium \( P \) charged by competitive firms (or competitive

\textsuperscript{17}Consistent with this restricted purpose, we will assume that the government cannot use health insurance to offset underlying differences in productivity coming from, say, education. This will prove important in the quantitative section.
insurance companies, who offer health insurance in the model), given the set of regulations spelled out above, is determined by

\[ P = \sum_h \left[ g(h)x(0, h) + (1 - g(h)) \int f(\varepsilon)x(\varepsilon, h)d\varepsilon \right] \Phi(h) \]  

(2.4)

where \( x(\varepsilon, h) \) is the expenditure schedule regulated by the government. This schedule is chosen to maximize

\[ \sum_h u(w(h) - P)\Phi(h) \]

with wages \( w(h) \) determined by (2.3).

**No Wage Discrimination Law**

The objective of the government is to prevent workers with a lower health status \( h \), and hence lower productivity, being paid less. As with the no prior conditions law, the purpose of this legislation is to help insure workers against their health status risk. However, if a production firm is penalized for paying workers with low health status \( h \) low wages, but not for preferentially hiring workers with a favorable health status (high \( h \)), then a firm can effectively circumvent the wage nondiscrimination law. Therefore, to be effective such a law must penalize both wage discrimination and hiring discrimination by health status.

Limiting wage dispersion with respect to gross wages \( w(h) \) via legislation necessitates regulation of the health insurance market as well, in order to prevent the insurance gains from decreasing wage dispersion being undone through the adjustment of employer-provided health insurance. For example, the firm could also offer health insurance and overcharge low productivity workers and undercharge high productivity workers for this insurance, effectively undermining the illegal wage discrimination. This suggests that the government will need to limit the extent to which the cost of a worker’s health insurance contracts deviates from its actuarially fair value. However, this will not be sufficient to make this policy effective.

Since the productivity of a worker depends upon the extent of his health insurance,
workers whose expected productivity is below their wage will face pressure to increase their productivity through increased spending on health (and hence better health insurance coverage) while those whose productivity is above their wage will have an incentive to lower their health insurance purchases. To prevent these distortions in the health insurance market and thereby achieve better consumption insurance across $h$ types, policy makers will need to regulate the health insurance directly as well. The moderate version of health insurance regulation would be to ensure that each policy was individually optimal and actuarially fair. The most extreme version of regulation would be to combine no-wage discrimination legislation with no-prior conditions legislation and thereby achieve the static first-best, full insurance outcome. In this case health insurance would be socially efficient and actuarially fair on average (that is, across the insured population).

We will analyze both cases. It will turn out that limiting wage dispersion with respect to net wages, $w(h) - P(h)$, avoids the negative incentive effects on the health insurance market. The policy of combining both no-wage discrimination and no-prior conditions can therefore be implemented through a policy of limiting net wage dispersion. The impact of the nondiscrimination law will, unfortunately, be sensitive to the way in which the law is implemented, and in particular, to the form of punishment used. If the limitation in wage variation is achieved through a policy that penalizes the firms for discriminating, then these costs are realized in equilibrium, reducing overall efficiency in the economy. If, however, the limitation on wage variation is achieved either through the threat of punishment (e.g. through grim trigger strategies in repeated interactions between firms and the government) or through the delegation of hiring in a union hiring hall type arrangement, then costs from the wage nondiscrimination law will not be realized in equilibrium.\footnote{The delegation method is similar to the structure we assumed in the insurance market since insurance companies were restricted to serving their customers on a first-come-first-serve basis. This assumption to us seems more problematic in the labor market because of the idiosyncratic nature of the benefits to the worker-firm match.}

Since we wish to give the no wage discrimination law the best shot of being successful, in the main text we focus on the version of the policy in which no costs from the policy are realized in equilibrium, leaving the analysis of the alternative case to appendix B.2.2.
and B.2.3. In either case we only tackle the extreme versions of these policies in which there is no wage discrimination (rather than limited wage discrimination) in equilibrium for reasons of analytic tractability. Under the policy, the firm takes as given thresholds on the size of the gap in wages or employment shares that will trigger the punishment. Assume that the wage penalty will be imposed if the maximum wage gap within the firm exceeds the threshold $\varepsilon_w$. Since type $h = 0$ will receive the lowest wage in equilibrium, to avoid the penalty a firm has to offer a wage schedule that satisfies:

$$\max_h |w(h) - w(0)| \leq \varepsilon_w.$$  

Letting $n(h)$ denote the number of workers of type $h$ hired by the firm, assume that the hiring penalty will be imposed if the employment share of type $h$ deviates from the population average by more than $\delta$, and hence

$$\left| \frac{n(h)}{\sum_h n(h)} - \frac{\Phi(h)}{\sum_h \Phi(h)} \right| \leq \delta.$$  

We will assume that the punishment is sufficiently dire that the firm will never choose to violate these thresholds.

We analyze the more general case in appendix B.2.1, but here focus on the limiting case in which the thresholds $\varepsilon_w$ and $\delta$ converge to zero. In this case, the firm will simply take as given the economy-wide wage $w^*$ at which it can hire a representative worker. We assume that the government regulates the insurance market determining the extent of coverage by health type, $x(e, h)$, subject to the requirement that the offered health insurance contracts exactly break even, either health type by health type (in the absence of a no prior conditions law) or in expectation across health types (in the presence of the no prior conditions law).

Perfect competition drives down equilibrium profits of firms to zero which determines the equilibrium wage rate as

$$w^* = \sum_h \left\{ g(h) F(h, -x(0, h)) + (1 - g(h)) \int_0^{\varepsilon} f(\varepsilon) \left[ F(h, \varepsilon - x(\varepsilon, h)) \right] d\varepsilon \right\} \Phi(h)$$  

(2.5)
The insurance premium charged to the household is

\[ P(h) = g(h)x(0, h) + (1 - g(h)) \int f(\varepsilon)x(\varepsilon, h)d\varepsilon \]  \hspace{1cm} (2.6)

in the absence of a no-prior conditions law and

\[ P = \sum_h \left[ g(h)x(0, h) + (1 - g(h)) \int f(\varepsilon)x(\varepsilon, h)d\varepsilon \right] \Phi(h) \]  \hspace{1cm} (2.7)

in its presence. Household consumption is given by

\[ c(h) = w^* - P(h) \] or

\[ c = w^* - P \]

depending on whether a no prior conditions law is in place or not.

Given a cross-sectional health distribution \( \Phi \) the efficiently regulated health insurance contract \( x(\varepsilon, h) \) is the solution to

\[ \max_x \sum_h u(w^* - P(h))\Phi(h) \]

subject to (2.5) and (2.6) if the no-prior conditions restriction is not imposed on health insurance, and subject to (2.7) instead of (2.6) if the no-prior conditions restriction is present.

We now turn to the analysis of the model, starting with a static version in which by construction the choice of effort is not distorted in equilibrium. We will show that in this case the competitive equilibrium implements an efficient allocation of health expenditures, but fails to provide efficient consumption insurance against prior health conditions, that is against cross-sectional variation in \( h \). We then argue that a combination of a strict wage non-discrimination law and a no prior conditions law in addition results in efficient consumption insurance in the competitive equilibrium, restoring full efficiency of allocations in the regulated market economy.
2.3 Analysis of the Static Model

We now turn to the analysis of the static version of our model, and we will characterize both efficient and equilibrium allocations (in the absence and presence of the nondiscrimination policies). The purpose of this analysis is two-fold. First, it will result in the characterization of the optimal and equilibrium health insurance contract, a key ingredient for our dynamic model. Second, the analysis will demonstrate that in the short run (that is statically) the combination of both policies is ideally suited to provide full consumption insurance in the regulated market equilibrium, and thus restores full efficiency of the market outcome. The static benefits of these policies are then traded off against the adverse dynamic consequences on the health distribution, as our analysis of the dynamic model will uncover in the next section.

2.3.1 Social Planner Problem

Given an initial cross-sectional distribution over health status in the population \( \Phi(h) \) the social planner maximizes utilitarian social welfare. The social planner problem is therefore given by

\[
U^{SP}(\Phi) = \max_{e(h),x(\varepsilon,h),c(\varepsilon,h) \geq 0} \sum_h \left\{ -q(e(h)) + g(h)u(c(0,h)) + (1 - g(h)) \int f(\varepsilon)u(c(\varepsilon,h))d\varepsilon \right\} \Phi(h)
\]

subject to

\[
\sum_h \left\{ g(h)c(0,h) + (1 - g(h)) \int f(\varepsilon)c(\varepsilon,h)d\varepsilon + g(h)x(0,h) + (1 - g(h)) \int f(\varepsilon)x(\varepsilon,h)d\varepsilon \right\} \Phi(h) \leq \sum_h \left\{ g(h)F(h,-x(0,h)) + (1 - g(h)) \int f(\varepsilon)F(h,\varepsilon-x(\varepsilon,h))d\varepsilon \right\} \Phi(h)
\]

We summarize the optimal solution to the static social planner problem in the following proposition, whose proof follows directly from the first order conditions and assumption 4 (see Appendix B.1).
Proposition 1. The solution to the social planner problem 
\{c^{SP}(\varepsilon, h), x^{SP}(\varepsilon, h), e^{SP}(h)\}_{h \in H} is given by

\[
e^{SP}(h) = 0
\]
\[
c^{SP}(\varepsilon, h) = c^{SP}
\]
\[
x^{SP}(\varepsilon, h) = \max \left[0, \varepsilon - \bar{\varepsilon}^{SP}(h)\right]
\]
where the cutoffs satisfy

\[-F_{2}(h, \bar{\varepsilon}^{SP}(h)) = 1, \tag{2.8}\]

and the first best consumption level is given by

\[
c^{SP} = \sum_{h} \left[ g(h)F(h, 0) + (1 - g(h)) \int_{0}^{\bar{\varepsilon}} f(\varepsilon) \left[ F(h, \varepsilon - x^{SP}(\varepsilon, h)) - x^{SP}(\varepsilon, h) \right] d\varepsilon \right] \Phi(h) \tag{2.9}\]

The optimal cutoff \{\bar{\varepsilon}^{SP}(h)\} is increasing in \( h \), strictly so if \( F_{12}(h, y) > 0 \).

The social planner finds it optimal to not have the household exercise (given that there are no dynamic benefits from doing so in the static model) and to provide full consumption insurance against adverse health shocks \( \varepsilon \), but also against bad prior health conditions as consumption is constant in \( h \).

The optimal level of health expenditure and its implications on production is graphically presented in Figure 2.4. As shown in the previous proposition, optimal medical expenditures take a simple cutoff rule: small health shocks \( \varepsilon < \bar{\varepsilon}^{SP}(h) \) are not treated at all, but all larger shocks are fully treated up to the threshold \( \bar{\varepsilon}^{SP}(h) \). These optimal medical expenditures are displayed in Figure 2.4(b) for two different initial levels of health \( h_1 < h_2 \): below the \( h \)-specific threshold \( \bar{\varepsilon}^{SP}(h) \) health expenditures are zero, and then rise one for one with the health shock \( \varepsilon \). The determination of the threshold itself is displayed in Figure 2.4(a). It shows that under the assumption that the impact of health shocks on productivity is less severe for healthy households \( (F_{12}(h, y) > 0) \), reflected as a “more concave” curve for \( h_1 \).
than for $h_2$ in Figure 2.4(a)), then the social planner finds it optimal to “insure” healthier households less, in the sense of undoing less of the negative health shocks $\epsilon$ through medical treatment $x(\epsilon,h)$. This is reflected in a lower threshold (more insurance) for $h_1$ than for $h_2$, that is $\bar{\epsilon}^{SP}(h_2) < \bar{\epsilon}^{SP}(h_1)$. The optimal health expenditure policy function leads to a net of-health-treatment production function $F(h,\epsilon - x^{SP}(\epsilon,h))$ as shown in Figure 2.4(c).

### 2.3.2 Competitive Equilibrium

As in the social planner problem there is no incentive for households to exercise in the static model, and thus $e(h) = 0$. As described in section 2.2.4 the equilibrium wage and health insurance contract solves

\begin{equation}
U^{CE}(h) = \max_{w(h),x(\epsilon,h),P(h)} u (w(h) - P(h)) \tag{2.10}
\end{equation}

s.t.

\begin{align}
P(h) &= g(h)x(0,h) + (1 - g(h)) \int_{0}^{\bar{\epsilon}} f(\epsilon)x(\epsilon,h)d\epsilon \tag{2.11} \\
w(h) &= g(h)F(h,-x(0,h)) + (1 - g(h)) \int_{0}^{\bar{\epsilon}} f(\epsilon)F(h,\epsilon - x(\epsilon,h))d\epsilon \tag{2.12}
\end{align}

The following proposition characterizes the solution to this problem:

**Proposition 2.** The unique equilibrium health insurance contract and associated consumption are given by

\begin{align}
x^{CE}(\epsilon,h) &= \max [0,\epsilon - \bar{\epsilon}^{CE}(h)] \tag{2.13} \\
c^{CE}(\epsilon,h) &= c^{CE}(h) = w^{CE}(h) - P^{CE}(h) \tag{2.14} \\
P^{CE}(h) &= (1 - g(h)) \int_{\bar{\epsilon}^{CE}(h)}^{\bar{\epsilon}} f(\epsilon) [\epsilon - \bar{\epsilon}^{CE}(h)] d\epsilon \tag{2.15} \\
w^{CE}(h) &= g(h)F(h,0) + (1 - g(h)) \int_{0}^{\bar{\epsilon}} f(\epsilon)F(h,\epsilon - x(\epsilon,h))d\epsilon \tag{2.16}
\end{align}

and the cutoff satisfies

\begin{equation}
-F_2(h,\bar{\epsilon}^{CE}(h)) = 1 \tag{2.17}
\end{equation}
Proof: See Appendix

We immediately obtain the following

**Corollary 3.** The competitive equilibrium implements the socially efficient health expenditure allocation since \( \bar{\varepsilon}^{CE}(h) = \bar{\varepsilon}^{SP}(h) \) for all \( h \in H \).

**Corollary 4.** The cutoff \( \bar{\varepsilon}^{CE}(h) \) is increasing in \( h \), strictly so if \( F_{12}(h,y) > 0 \).

While it follows trivially from our assumptions that the worker’s net pay, \( w(h) - P(h) \), is increasing in \( h \), it is not necessarily true that his gross wage, \( w(h) \), is increasing in \( h \) as well since optimal health expenditures are decreasing in health status. We analyze the behavior of gross wages \( w(h) \) with respect to health status further in Appendix B.3, where we provide a sufficient condition for the gross wage schedule to be monotonically increasing in \( h \).

In any case, the previous results show that in the static case the only source of inefficiency of the competitive equilibrium comes from the inefficient lack of consumption insurance against adverse prior health conditions \( h \). This can be seen by noting that

\[
\begin{align*}
\Phi(h) &= \Phi(h) \\
= & \sum_h \left[ (w^{CE}(h) - P^{CE}(h)) \Phi(h) \right] = \sum_h \Phi(h) \Phi(h)
\end{align*}
\]

In contrast to what will be the case in the dynamic model, effort trivially is not distorted in the equilibrium, relative to the allocation the social planner implements (since in both cases \( e^{SP} = e^{CE} = 0 \)). Furthermore the equilibrium allocation of health expenditures is efficient, due to the fact that the firm bundles the determination of wages and the provision of health insurance, and thus internalizes the positive effects of health spending \( x(\varepsilon, h) \) on worker productivity.

Given these results it is plausible to expect, within the context of the static model, that policies preventing competitive equilibrium wages \( w^{CE}(h) \) to depend on health status (a wage non-discrimination law) and insurance premia \( P^{CE}(h) \) to depend on health status
(a no prior conditions law) will restore full efficiency of the policy-regulated competitive equilibrium by providing full consumption insurance. We will show next that this is indeed the case, providing a normative justification for the two policy interventions within the static version of our model.

2.3.3 Competitive Equilibrium with a No Prior Condition Law

As discussed above, in order to effectively implement a no prior conditions law the government has to regulate the health insurance provision done by firms or insurance companies. Given a population health distribution $\Phi$ the regulatory authority solves the problem:

$$U_{NP}^{\Phi}(\Phi) = \max_{x,\varepsilon} \sum_h u(w(h) - P)\Phi(h)$$ (2.18) s.t. 

$$P = \sum_h \left[ g(h)x(0,h) + (1 - g(h)) \int f(\varepsilon)x(\varepsilon,h) d\varepsilon \right] \Phi(h)$$ (2.19) 

$$w(h) = g(h)F(h,-x(0,h)) + (1 - g(h)) \int_0^\varepsilon f(\varepsilon)F(h,\varepsilon - x(\varepsilon,h)) d\varepsilon$$ (2.20)

The next proposition characterizes the resulting regulated equilibrium allocation

**Proposition 5.** The equilibrium health expenditures under a no-prior condition law satisfies, for each $\tilde{h} \in H$

$$x^{NP}(\varepsilon, \tilde{h}) = \max[0, \varepsilon - \tilde{\varepsilon}^{NP}(\tilde{h})]$$

with cutoffs uniquely determined by

$$-F_2(\tilde{h}, \tilde{\varepsilon}^{NP}(\tilde{h})) = \frac{\sum_h u'(w^{NP}(h) - P^{NP})\Phi(h)}{u'(w(\tilde{h}) - P^{NP})}. $$

The equilibrium wage, for each $\tilde{h}$, is given by

$$w^{NP}(\tilde{h}) = g(\tilde{h})F(\tilde{h}, 0) + (1 - g(\tilde{h})) \int_0^\varepsilon f(\varepsilon)[F(\tilde{h}, \varepsilon - x^{NP}(\varepsilon, \tilde{h}))] d\varepsilon$$
and the health insurance premium is determined as

\[ P^{NP} = \sum_h \left[ g(h)x^{NP}(0, h) + (1 - g(h)) \int f(\varepsilon)x^{NP}(\varepsilon, h) d\varepsilon \right] \Phi(h). \]

Moreover, the optimal cutoffs are increasing in health status.

**Proof:** See Appendix.

Note that the health expenditure levels are no longer efficient as the government provides partial consumption insurance against initial health status when choosing the cutoff levels \( \bar{\varepsilon}^{NP}(h) \), in the absence of direct insurance against low wages induced by bad health. In fact, as shown in the next proposition, it is efficient to over-insure households with bad health status and under-insure those with good health status, relative to the first-best.

**Proposition 6.** Let \( \tilde{h} \) be the health status whose marginal utility of consumption is equal to the population average, i.e. for \( \tilde{h} \),

\[ -F_2(\tilde{h}, \bar{\varepsilon}(\tilde{h})) = \frac{\sum_h u'(w(h) - P)\Phi(h)}{u'(w(\tilde{h}) - P)} = 1 \]  

holds.\(^{19}\) Then,

\[ \bar{\varepsilon}^{NP}(h) < \bar{\varepsilon}^{SP}(h), \quad \text{for } h < \tilde{h} \]

\[ \bar{\varepsilon}^{NP}(h) = \bar{\varepsilon}^{SP}(h), \quad \text{for } h = \tilde{h} \]

\[ \bar{\varepsilon}^{NP}(h) > \bar{\varepsilon}^{SP}(h), \quad \text{for } h > \tilde{h}, \]

The cutoffs \( \bar{\varepsilon}(h) \) are strictly monotonically increasing in health status \( h \).

**Proof:** See Appendix.

This feature of the optimal health expenditure with a no prior conditions law also indicates that mandatory participation in the health insurance contract is an important part of government regulation, since in the allocation described above healthy households

\(^{19}\)For the purpose of the proposition it does not matter whether \( \tilde{h} \in H \) or not.
cross-subsidize the unhealthy in terms of insurance premia and they are given a less generous health expenditure plan (higher thresholds) than the unhealthy.

### 2.3.4 Competitive Equilibrium with a No Wage Discrimination Law

The equilibrium with a no wage discrimination law is determined by the solution to the program:

\[
U^{ND}(\Phi) = \max_{x(\varepsilon,h)} \sum_h u(w - P(h))\Phi(h) \\
\text{s.t.} \\
P(h) = \left[g(h)x_0(h) + (1 - g(h)) \int f(\varepsilon)x(\varepsilon,h)\,d\varepsilon\right] \\
w = \sum_h \left\{g(h)F(h,-x_0(h)) + (1 - g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon)F(h,\varepsilon - x(\varepsilon,h))\,d\varepsilon\right\} \Phi(h)
\]

**Proposition 7.** The equilibrium health expenditures under a no-wage discrimination law alone satisfies, for each \(\tilde{h} \in H\)

\[
x^{ND}(\varepsilon,\tilde{h}) = \max \left[0, \varepsilon - \tilde{\varepsilon}^{ND}(\tilde{h})\right]
\]

with cutoffs determined by

\[-F_2(\tilde{h},\tilde{\varepsilon}^{ND}(\tilde{h})) = \frac{u'(w^{ND} - P(\tilde{h}))}{\sum_h u'(w^{ND} - P(h))\Phi(h)}.
\]

The equilibrium wage is given by

\[
w^{ND} = \sum_h \left[g(h)\left[F(h,0)\right] + (1 - g(h)) \int_0^{\tilde{\varepsilon}} f(\varepsilon)\left[F(h,\varepsilon - x^{ND}(\varepsilon,h))\right] \,d\varepsilon\right] \Phi(h)
\]

and the health insurance premium is given by, for each \(\tilde{h}\),

\[
P^{ND}(\tilde{h}) = \left[g(\tilde{h})x^{ND}(0,\tilde{h}) + (1 - g(\tilde{h})) \int f(\varepsilon)x^{ND}(\varepsilon,\tilde{h})\,d\varepsilon\right].
\]
Proof: Follows directly from the first order conditions of the program (2.22).

Unlike in the no prior conditions case, we cannot establish monotonicity in the cutoffs $\bar{\epsilon}^{ND}(\tilde{h})$. Note that under a no prior conditions law the regulatory authority partially insures consumption of the unhealthy by allocating higher medical expenditure to them. Under a no wage discrimination law instead, there are two opposing forces, preventing us from establishing monotonicity in cutoffs $\bar{\epsilon}^{ND}(h)$ across health groups $h$. On one hand, a one unit increase in medical expenditure $P(h)$ is more costly to the unhealthy since marginal utility of consumption is higher for this group. On the other hand, production efficiency calls for higher medical expenditure for the unhealthy, given our assumption of $F_{12} \geq 0$ (as was the case for the no prior conditions law). Thus the cutoffs $\bar{\epsilon}^{ND}(h)$ need not be monotone in $h$.

2.3.5 Competitive Equilibrium with Both Policies

Finally, combining both a no-wage discrimination law and a no-prior conditions legislation restores efficiency of the regulated equilibrium since both policies in conjunction provide full consumption insurance against bad health realizations $h$. This is the content of the next.

Corollary 8. The unique competitive equilibrium allocation in the presence of both a no wage discrimination and a no prior conditions law implements the socially efficient allocation in the static model.

Proof: The equilibrium is the solution to

$$\max_{x(\varepsilon,h)} \sum_{h} u(w^* - P)\Phi(h)$$

s.t.

$$P = \sum_{h} \left[ g(h)x(0,h) + (1 - g(h)) \int f(\varepsilon)x(\varepsilon,h) d\varepsilon \right]\Phi(h)$$

$$w^* = \sum_{h} \left\{ g(h)F(h,-x(0,h)) + (1 - g(h)) \int^{\varepsilon}_{0} f(\varepsilon)F(h,\varepsilon - x(\varepsilon,h)) d\varepsilon \right\}\Phi(h).$$

The result then follows trivially from the fact that this maximization problem is equivalent
to the social planner problem analyzed above. The no prior conditions law equalizes health insurance premia $P$ across health types, the no wage discrimination law implements a common wage $w^*$ across health types, and the (assumed) efficient regulation of the health insurance market assures that the health expenditure schedule is efficient as well.

### 2.3.6 Summary of the Analysis of the Static Model

The competitive equilibrium implements the efficient health expenditure allocation but does not insure households against initial health conditions. Both a no-prior conditions law and a no-wage discrimination law provide partial, but not complete, consumption insurance against this risk, without distorting the effort level. The health expenditure schedule is distorted when each policy is implemented in isolation, relative to the social optimum, as the government provides additional partial consumption insurance through health expenditures. Only both laws in conjunction implement a fully efficient health expenditure schedule and full consumption insurance against initial health conditions $h$, and therefore restore the first best allocation in the static model. Enacting both policies jointly is thus fully successful in what they are designed to achieve in a static world (partially due to the fact that additional government regulation severely restricted the options of firms to circumvent the government policies).

### 2.4 Analysis of the Dynamic Model

We now study a dynamic version of our economy. Both in terms of casting the problem, as well as in terms of its computation we make use of the fact that there is no aggregate risk (due to the continuum of agents cum law of large numbers assumption). Therefore the sequence of cross-sectional health distributions $\{\Phi_t\}_{t=0}^T$ is a deterministic sequence. Furthermore, conditional on a distribution $\Phi_t$ today the health distribution tomorrow is completely determined by the effort choice $e_t(h)$ of households\(^{20}\) (or the social planner), so

\[^{20}\text{We assert here that the optimal effort in period } t \text{ is only a function of the current individual health status } h. \text{ We will discuss below the assumptions required to make this assertion correct.}\]
that we can write
\[ \Phi_{t+1} = H(\Phi_t; e_t(.)) \] (2.23)
where the time-invariant function \( H \) is in turn completely determined by the Markov transition function \( Q(h'; h, e) \). The initial distribution \( \Phi_0 \) is an initial condition and exogenously given.

Under each policy, given a sequence of aggregate distributions \( \{\Phi_t\}_{t=0}^T \) we can solve an appropriate dynamic maximization problem of an individual household for the sequence of optimal effort decisions \( \{e_t(h)\in H\}_{t=0}^T \) which in turn imply a new sequence of aggregate distributions via (2.23). Our computational algorithm for solving competitive equilibria then amounts to iterating on the sequences \( \{\Phi_t, e_t\} \). Within each period the timing of events follows exactly that of the static problem in the previous section.

2.4.1 Social Planner Problem

The dynamic problem of the social planner is to solve
\[
V(\Phi_0) = \max_{\{e_t(h)\}} \sum_{t=0}^T \beta^t \left\{ U^{SP}(\Phi_t) - \sum_h q(e_t(h))\Phi_t(h) \right\}
\]
where \( \{\Phi_{t+1}\} \) is determined by equation (2.23) and
\[
U^{SP}(\Phi) = \max_{x(\varepsilon,h),c(\varepsilon,h)} \sum_h \left\{ g(h)u(c(0,h)) + (1-g(h)) \int f(\varepsilon)u(c(\varepsilon,h))d\varepsilon \right\} \Phi(h)
\]
is the solution to the static social planner problem characterized in section 2.3.1:
\[
x^{SP}(\varepsilon,h) = \max \left[ 0, \varepsilon - \bar{\varepsilon}^{SP}(h) \right]
\]
with cutoffs defined by
\[
- F_2(h, \bar{\varepsilon}^{SP}(h)) = 1 \] (2.24)
and consumption of each household given by

\[ c^{SP}(\Phi) = \sum_{h} \left[ g(h)F(h,0) + (1 - g(h)) \int_{\varepsilon} f(\varepsilon) \left[ F(h,\varepsilon - x^{SP}(\varepsilon,h)) - x^{SP}(\varepsilon,h) \right] d\varepsilon \right] \Phi(h). \]

We now want to characterize the optimal effort choice by the social planner, the key dynamic decision in our model both in the planner problem and the competitive equilibrium. In contrast to households in the competitive equilibrium, the social planner fully takes into account the effect of effort choices today on the aggregate health distribution and thus aggregate consumption tomorrow.

A semi-recursive formulation of the problem is useful to characterize the optimal effort choice, but also to explain the computational algorithm for the social planner problem. For a given cross-sectional distribution \( \Phi_t \) at the beginning of period \( t \) the social planner solves:

\[
V_t(\Phi_t) = u(c_t) + \max_{e_t(h) \in H} \left\{ - \sum_h q(e_t(h))\Phi_t(h) + \beta V_{t+1}(\Phi_{t+1}) \right\}
\]

s.t. \( c_t = c^{SP}(\Phi_t) \)

\[
\Phi_{t+1}(h') = \sum_h Q(h';h,e_t(h))\Phi(h)
\]

(2.25)

In appendix B.4 we discuss how we solve this problem numerically, iterating on sequences \( \{c_t, e_t(h), \Phi_t(h)\}_{t=0}^T \) from the terminal condition \( V_T(\Phi_T) = u(c_T) \). To characterize the optimal effort choice, for an arbitrary time period \( t \) we obtain the first order condition:

\[
q'(e_t(h))\Phi_t(h) = \beta \sum_{h'} \frac{\partial V_{t+1}(\Phi_{t+1})}{\partial \Phi_{t+1}(h')} \cdot \frac{\partial \Phi_{t+1}(h')}{\partial e_t(h)}
\]

\[
= \beta \sum_{h'} \frac{\partial V_{t+1}(\Phi_{t+1})}{\partial \Phi_{t+1}(h')} \cdot \frac{\partial Q(h';h,e_t(h))}{\partial e_t(h)} \Phi_t(h),
\]

This simplifies to

\[
q'(e_t(h)) = \beta \sum_{h'} \frac{\partial V_{t+1}(\Phi_{t+1})}{\partial \Phi_{t+1}(h')} \cdot \frac{\partial Q(h';h,e_t(h))}{\partial e_t(h)}.
\]

(2.26)
Thus the marginal cost of extra effort \( q'(e_t(h)) \) is equated to the marginal benefit, the latter being given by the benefit that effort has on the health distribution tomorrow, \( \frac{\partial Q(h', h, e_t(h))}{\partial e_t(h)} \), times the benefit of a better health distribution \( \frac{\partial V_{t+1}(\Phi_{t+1})}{\partial \Phi_{t+1}(h')} \) from tomorrow on. By assumption 1, \( q'(0) = 0 \), and assumption 3 guarantees that the right hand side of equation (2.26) is strictly positive. Therefore the social planner finds it optimal to make every household exert positive effort to lead a healthy life: \( e_t(h) > 0 \) for all \( t \) and all \( h \in H \).

From the envelope theorem the benefit of a better health distribution is given by:

\[
\frac{\partial V_t(\Phi_t)}{\partial \Phi_t(h)} = u'(c_t) \cdot \Psi(h) - q(e_t(h)) + \beta \sum_{h'} \frac{\partial V_{t+1}(\Phi_{t+1})}{\partial \Phi_{t+1}(h')} \cdot Q(h'; h, e_t(h)).
\] (2.27)

Here \( \Psi(h) \) denotes the expected output, net of health expenditures, that an individual of health status \( h \) delivers to the social planner.\(^{21}\)

### 2.4.2 Competitive Equilibrium without Policy

In our model, since absent wage and health insurance policies households do not interact in any way, we can solve the dynamic programming problem of each household independently of the rest of society. The only state variables of the household are her current health \( h \) and age \( t \), and the dynamic program reads as:

\[
v_t(h) = U^{CE}(h) + \max_{e_t(h)} \left\{ -q(e_t(h)) + \beta \sum_{h'} Q(h'; h, e_t(h))v_{t+1}(h') \right\}
\] (2.28)

\(^{21}\)Note that

\[
\Psi(h) = \left[ g(h)F(h, 0) + (1 - g(h)) \int_\varepsilon f(\varepsilon) \left[ F(h, \varepsilon - x^{SP}(\varepsilon, h)) - x^{SP}(\varepsilon, h) \right] d\varepsilon \right]
\]

is exclusively determined by the optimal cut-off rule \( \varepsilon^{SP}(h) \) for health expenditures, which is independent of \( c_t \) or \( \Phi_t \).
where

\[
U_{CE}(h) = \max_{x(\varepsilon,h),w(h),P(h)} u(w(h) - P(h))
\]
\[
s.t.
\]
\[
w(h) = g(h)F(h,-x(0,h)) + (1-g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon)F(h,\varepsilon - x(\varepsilon,h))d\varepsilon
\]
\[
P(h) = g(h)x(0,h) + (1-g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon)x(\varepsilon,h)d\varepsilon
\]
is the solution to the static equilibrium problem in section 2.3.2, which was given by:

\[
x_{CE}(\varepsilon,h) = \max [0,\varepsilon - \varepsilon_{CE}(h)]
\]
\[
c_{CE}(h) = w_{CE}(h) - P_{CE}(h)
\]
\[
P_{CE}(h) = (1-g(h)) \int_{\varepsilon_{CE}(h)}^{\bar{\varepsilon}} f(\varepsilon) [\varepsilon - \varepsilon_{CE}(h)] d\varepsilon
\]
\[
w_{CE}(h) = g(h)F(h,0) + (1-g(h)) \int_0^{\varepsilon_{CE}} f(\varepsilon)F(h,\varepsilon - x(\varepsilon,h))d\varepsilon
\]

with cutoff:

\[-F_2(h,\varepsilon_{CE}(h)) = 1\]

Note again that the provision of health insurance is socially efficient in the competitive equilibrium.

In contrast to the social planner problem, and in contrast to what will be the case in a competitive equilibrium with a no-wage discrimination law or a no-prior conditions law, in the unregulated competitive equilibrium there is no interaction between the maximization problems of individual households. Thus the dynamic household maximization problem can be solved independent of the evolution of the cross-sectional health distribution. It is a simple dynamic programming problem with terminal value function

\[v_T(h) = U_{CE}(h)\]

and can be solved by straightforward backward iteration.
Given the solution \( \{e_t(h)\} \) of the household dynamic programming problem and given an initial distribution \( \Phi_0 \) the dynamics of the health distribution is then determined by the aggregate law of motion (2.23). The optimal choice \( e_t(h) \) solves the first order condition

\[
q'(e_t(h)) = \beta \sum_{h'} \frac{\partial Q(h'; h, e_t(h))}{\partial e_t(h)} v_{t+1}(h')
\]

(2.29)

Note that at time \( t \) when the decision \( e_t(h) \) is taken the function \( v_{t+1}(.) \) is known. Furthermore, given knowledge of \( v_{t+1} \) and the optimal \( e_t \) the period \( t \) value function \( v_t \) is determined by (2.28). As in the social planner problem, by assumptions 1 and 3 effort \( e_t(h) \) is positive for all \( t \) and \( h \).

### 2.4.3 Competitive Equilibrium with a No Prior Condition Law

As discussed above, we assume that the government in every period \( t \) takes as given the health distribution \( \Phi_t \) and enforces the no prior condition law and regulates health insurance contracts efficiently, as in the static analysis of section 2.3.3. We now make explicit that the solution of the static government regulation problem (2.18)-(2.20) is a function of the cross-sectional health distribution,

\[
x^{NP}(\varepsilon, \tilde{h}; \Phi_t) = \max[0, \varepsilon - \tilde{\varepsilon}^{NP}(\tilde{h}; \Phi_t)]
\]

(2.30)

with cutoffs for each \( \tilde{h} \in H \) determined by

\[
-F_2(\tilde{h}, \tilde{\varepsilon}^{NP}(\tilde{h}; \Phi_t)) u'(w^{NP}(\tilde{h}; \Phi_t) - P^{NP}(\Phi_t)) = \sum_h u'(w^{NP}(h; \Phi_t) - P^{NP}(\Phi_t)) \Phi_t(h)
\]

\[
:= Eu'(\Phi_t)
\]

(2.31)
and

\[ w^{NP}(h; \Phi_t) = g(h)F(h, 0) + (1 - g(h)) \int f(\varepsilon)[F(h, \varepsilon - x^{NP}(\varepsilon, h; \Phi_t))]d\varepsilon \]  
(2.32)

\[ P^{NP}(\Phi_t) = \sum_h \left[ g(h)x^{NP}(0, h; \Phi_t) \Phi_t(h) + (1 - g(h)) \int f(\varepsilon)x^{NP}(\varepsilon, h; \Phi) d\varepsilon \right] \]  
(2.33)

In order for the household to solve her dynamic programming problem she only needs to know the sequence of wages and health insurance premia \( \{w_t(h), P_t\} \), but not necessarily the sequence of distributions that led to it. Given such a sequence the dynamic programming problem of the household then reads as

\[ v_t(h) = u(w_t(h) - P_t) + \max_{e_t(h)} \left\{ -q(e_t(h)) + \beta \sum_{h'} Q(h'; h, e_t(h))v_{t+1}(h') \right\} \]  
(2.34)

with terminal condition \( v_T(h) = u(w_T(h) - P_T) \). As before the optimality condition reads as

\[ q'(e_t(h)) = \beta \sum_{h'} \frac{\partial Q(h'; h, e_t(h))}{\partial e_t(h)} v_{t+1}(h'). \]  
(2.35)

and thus equates the marginal cost of providing effort, \( q'(e) \) with the marginal benefit of an improved health distribution tomorrow. Although equation (2.35) looks identical to equation (2.29) from the unregulated equilibrium, the determination of the value functions that appear on the right hand side of both equations is not (compare the first terms on the right hand sides of equations (2.28) and (2.34)). The difference in these equations highlights the extra consumption insurance induced by the no-prior conditions law, in that with this policy the health insurance premium does not vary with \( h \). This extra consumption insurance, ceteris paribus, reduces the variation of \( v_{t+1} \) in \( h' \) and thus limits the incentives to exert effort in order to achieve a (stochastically) higher health level tomorrow. In appendix B.5 we describe a computational algorithm to solve the dynamic model with a no-prior conditions law.
2.4.4 Competitive Equilibrium with a No Wage Discrimination Law

The main difference to the previous section is that now the static health insurance contract and premium are given by health spending

\[ x^{ND}(\varepsilon, \tilde{h}; \Phi_t) = \max\{0, \varepsilon - \tilde{\varepsilon}^{ND}(\tilde{h}; \Phi_t)\} \]  

(2.36)

with cutoffs for each \( \tilde{h} \in H \) determined by

\[ -F_2(h, \tilde{\varepsilon}^{ND}(h))Eu'_t = u'(w^{ND}(\Phi_t) - P^{ND}(h, \Phi_t)) \]  

(2.37)

where

\[ Eu'_t := \sum_h u'(w^{ND}(\Phi_t) - P^{ND}(h, \Phi_t))\Phi_t(h). \]  

(2.38)

The equilibrium wage is given by

\[ w^{ND}(\Phi_t) = \sum_h \left\{ g(h)F(h, 0) + (1 - g(h)) \int f(\varepsilon)[F(h, \varepsilon - x^{ND}(\varepsilon, h; \Phi_t))]d\varepsilon \right\}\Phi_t(h). \]  

(2.39)

The equilibrium health insurance premium depends on whether a no prior conditions law is in place or not: Without such policy the premia are given as

\[ P^{ND}(h; \Phi_t) = P^{ND}(h) = (1 - g(h)) \int f(\varepsilon)x^{ND}(\varepsilon, h)d\varepsilon \]  

(2.40)

whereas with both policies in place the premium is determined by\(^{22}\)

\[ P^{Both}(\Phi_t) = \sum_h \left[ (1 - g(h)) \int f(\varepsilon)x^{Both}(\varepsilon, h)d\varepsilon \right]\Phi_t(h) \]  

(2.41)

For a given sequence of wages \( \{w_t, P_t(h)\} \) the dynamic problem of the household reads

\(^{22}\)Wages still take the form as in (2.39), but with \( x^{Both}(\varepsilon, h) \) replacing \( x^{ND}(\varepsilon, h) \). Recall from the static analysis that \( x^{Both}(\varepsilon, h) = x^{SP}(\varepsilon, h) \), that is, the medical expenditure schedule is socially efficient.
as before:

\[ v_t(h) = u(w_t - P_t(h)) + \max_{e_t(h)} \left\{ -q(e_t(h)) + \beta \sum_{h'} Q(h'; h, e_t(h))v_{t+1}(h') \right\} \]  

(2.42)

and the terminal condition \( v_T(h) = u(w_T - P_T(h)) \), first order conditions and updating of the value function for this version of the model are exactly the same, mutatis mutandis, as under the previous policy. In appendix B.5 we discuss the algorithm to solve this version of the model.

2.4.5 Competitive Equilibrium with Both Laws

If both policies are in place simultaneously, we can give a full analytical characterization of the equilibrium without resorting to any numerical solution procedure. We do so in the next

**Proposition 9.** Suppose there is a no wage discrimination and a no prior condition law in place simultaneously. Then

\[ e_t(h) = 0 \] for all \( h \), and all \( t \).

The provision of health insurance is socially efficient. From the initial distribution \( \Phi_0 \) the health distribution in society evolves according to (2.23) with \( e_t(h) \equiv 0 \).

The proof is by straightforward backward induction and is given in Appendix B.1. In the presence of both policies there are no incentives, either through wages or health insurance premia, to exert effort to lead a healthy life. Since effort is costly, households won’t provide any such effort in the regulated dynamic competitive equilibrium. Thus in the absence of any direct utility benefits of better health the combination of both policies leads to a complete collapse in incentives, with the associated adverse long run consequences for the distribution of health in society.

Equipped with these theoretical results and the numerical algorithms to solve the various versions of our model we now map our model to cross-sectional health and exercise data
from the PSID to quantify the effects of government regulations on the evolution of the cross-sectional health distribution, as well as aggregate production, consumption and health expenditures.

2.5 Bringing the Model to the Data

2.5.1 Augmenting the Model

The model described so far only included the necessary elements to highlight the key static insurance-dynamic incentive trade-off we want to emphasize. However, to insure that the model can capture the significant heterogeneity in health, exercise and health expenditure data observed in micro data we now augment it in four aspects. We want to stress, however, that none of the qualitative results derived so far rely on the absence of these elements, which is why we abstracted from them in our theoretical analysis.

First, in the data some households have health expenditures in a given year from catastrophic illnesses that exceed their labor earnings. In the model, the only benefit of spending resources on health is to offset the negative productivity consequences of the adverse health shocks $\varepsilon$. Thus it is never optimal to incur health expenditures that exceed the value of a worker’s production in a given period. In order to capture these large medical expenditures in data and arrive at realistic magnitudes of health insurance premia we introduce a second health shock. This exogenous shock $z$ stands in for a catastrophic health expenditure shock, and when households receive the $z$-shock, they *have to* spend $z$; otherwise, they die (or equivalently, incur a prohibitively large utility cost). Households in the augmented model are assumed to either not receive any health shock, face either a $z$-shock, or an $\varepsilon$-shock, but not both. We denote by $\mu_z(h)$ the mean of the health expenditure shock $z$, conditional on initial health $h$, and by $\kappa(h)$ the probability of receiving a positive $z$-shock. Households that received a $z$-shock can still work, but at a reduced productivity $\rho < 1$ relative to healthy workers. As described in more detail in appendix B.6.1, the $z$-shock merely scales up health insurance premia by $\mu_z(h)$ and introduces additional health-related
wage risk (since $z$-shocks come with a loss of $1 - \rho$ of labor productivity).

Second, in our model so far all variation in wages was due to either health ($h$ and $\varepsilon - x$) or age $t$. When bringing the model to the data we permit earnings in the model to also depend on the education $educ$ of a household, and consequently specify the production function as $F(t, educ, h, \varepsilon - x)$. Given this extension we have to take a stance on how households of different education levels interact in equilibrium under each policy. Since our objective is to highlight the insurance aspect of both policies with respect to health-related consumption risks we assume that even in the presence of a wage discrimination law individuals with higher education can be paid more, and that health insurance companies can charge differential premia to individuals with heterogeneous education levels even in the presence of a no-prior conditions law.

Third, for the model to have a change of generating the observed heterogeneity in exercise levels of individuals that are identical in terms of their age, health and education levels we introduce preference shocks to the disutility from effort. Instead of being given by $q(e)$, as in the theoretical analysis so far, the cost of exerting effort is now assumed to be given as $\gamma q(e)$, where $\gamma \in \Gamma$ is an individual-specific preference shock that is drawn from the finite set $\Gamma$ at the beginning of life and remains constant during the individual’s life cycle. Note that since $\gamma$ only affects the disutility of effort which is separable from the utility of consumption, the analysis of the static model in section 2.3 remains completely unchanged (and so do the optimal health insurance contracts and health expenditure allocations). In the analysis of the dynamic model, since $\gamma$ is a permanent shock, all expressions involving $q(.)$ turn into $\gamma q(.)$ but the analysis is otherwise unaltered. Under the maintained assumption that wages and insurance premia are allowed to differ across different $\gamma$-groups even in the presence of the laws (an assumption that parallels the one made in the previous paragraph) there is no interaction between the different $(\gamma, educ)$ types and equilibrium allocations under all

\footnote{It does not matter whether firms/health insurance companies observe a worker’s preference parameter $\gamma$ since they engage only in short-term contracts and since $h$ is observable ($\gamma$ only affects effort and firms as well as health insurance companies do not care how the individual’s health evolves due to the restriction of attention to short-term contracts).}
policies can be solved for each \((\gamma, educ)\) pair separately.\(^{24}\) These assumptions again highlight the role of \((\gamma, educ)\)-heterogeneity modeled here: it is not the focal point of our insurance vs. incentives analysis, but rather allows us to capture some of the heterogeneity in outcomes in the data and thus avoids attributing all of this observed heterogeneity to health differences. Ignoring these other sources of heterogeneity would quantitatively overstate likely both the insurance benefits as well as the incentive costs of the policies we analyze in this paper. Consistent with the introduction of preference and skill (education) heterogeneity the initial distribution over household types is now denoted by \(\Phi_0(h, \gamma, educ)\) and will be determined from the data (but exploiting predictions of the structural model).

The last, and perhaps most significant departure from the theoretical model is that we now endow the household with a health-dependent continuation utility \(v_{T+1}(h)\) from retirement. The theoretical model implicitly assumed that this continuation utility was identically equal to zero, independent of the health status at retirement. The vector \(v_{T+1}(h)\) will be determined as part of our structural model estimation. Endowing individuals with nontrivial continuation utility at retirement avoids the counterfactual prediction of the model that effort is zero in the last period of working life, \(T\). This assumption also introduces a direct utility benefit from better health (albeit one that materializes at retirement) and thus avoids the complete collapse of incentives to provide effort under both policies (that is, proposition 9 no longer applies).

In the rest of this section, we use the so extended version of our model to estimate parameters to match PSID data on health, expenditure and exercise in 1999. In the main body of the paper, we describe the procedure we follow in a condensed form, relegating the detailed data description and estimation procedures to the Appendix B.6. Once the model is parameterized and its reasonable fit of the data established, in section 2.6 we then use it to analyze the positive and normative short- and long-run consequences of introducing non-discrimination legislation.

\(^{24}\)In order to obtain a meaningful welfare comparison with socially optimal allocations we also solve the social planner problem separately for each \((\gamma, educ)\) combination, therefore ruling out ex ante social insurance against bad initial \((\gamma, educ)\) draws.
2.5.2 Parameter Estimation and Calibration

The determination of the model parameters proceeds in three steps. First, we fix a small subset of parameters exogenously. Second, parts of the model parameters can be estimated from the PSID data directly. These include the parameters governing the health transition function $Q(h'|h, e)$, the probabilities $(g(h), \kappa(h))$ of receiving the $\varepsilon$ and $z$ health shocks, as well as the productivity effect of the $z$-shocks given by $\rho$. Third, (and given the parameters obtained in step 1 and 2) the remaining parameters (mainly those governing the production function $F$, the $\varepsilon$-shock distribution $f(\varepsilon)$ and preferences) are then determined through a method of moments estimation of the model with PSID wage, health and effort data. We now describe these three steps in greater detail.

A Priori Chosen Parameters

First, we choose one model period to be six years, a compromise between assuring that effort has a noticeable effect on health transitions (which requires a sufficiently long time period) and reasonable sample sizes for estimation (which speaks for short time periods). We then select two preference parameters a priori. Consistent with values commonly used in the quantitative macroeconomics literature we choose a risk aversion parameter of $\sigma = 2$ and a time discount factor of $\beta = 0.96$ per annum.

Parameters Estimated Directly from the Data

In a second step we estimate part of the model parameters directly from the data, without having to rely on the equilibrium of the model.

Health Transition Function $Q(h'|h, e)$ The PSID includes measures of light and heavy exercise levels\(^{25}\) starting in 1999 which we use to estimate health transition functions. We denote by $e^l$ and $e^h$ the frequency of light and heavy exercise levels, and assume the following

\(^{25}\)Number of times an individual carries out light physical activity (walking, dancing, gardening, golfing, bowling, etc.) and heavy physical activity (heavy housework, aerobics, running, swimming, or bicycling).
parametric functional form for the health transition function:

\[
Q(h'; h, e, e_h) = \begin{cases} 
(1 + \pi(h, e, e_h)\alpha_i(h))G(h, h'), & \text{if } h' = h + i, i \in \{1, 2\} \\
(1 + \pi(h, e, e_h))G(h, h'), & \text{if } h' = h, h > 1 \\
G(h, h'), & \text{if } h' = h - 1, h > 1 \\
\frac{1 - \sum_{h' \geq h} Q(h'; h, e, e_h)}{\sum_{h' < h} G(h, h')}G(h, h'), & \text{if } h' = h, h = 1 \\
\end{cases}
\]

where

\[
\pi(h, e, e_h) = \phi(h)(\delta e + (1 - \delta) e_h)^{\lambda(h)}.
\]

Since light and heavy physical exercise can have different effects on health transition, we give weight \(\delta\) on light exercise, and \((1 - \delta)\) on heavy exercise. We think of \(\delta e + (1 - \delta) e_h\) as the composite exercise level \(e\) used in the theoretical analysis of our model.

**Health Shock Probabilities** \(g(h)\) and \(\kappa(h)\)  In our model, \(g(h)\) represents the probability of not receiving any shock, and \(\kappa(h)\) is the probability of facing a \(z\)-shock. Since we assume that households do not receive both an \(\varepsilon\)-shock and a \(z\)-shock in the same period, the probability of facing an \(\varepsilon\)-shock is given by \(1 - g(h) - \kappa(h)\). From PSID, we first construct the probabilities of having a \(z\)-shock and an \(\varepsilon\)-shock. We define households that have received a \(z\)-shock as those who were diagnosed with cancer, a heart attack, or a heart disease\(^{26}\) and those who spent more on medical expenditures than their current income when hit with a health shock. Households with all other health shocks or those who missed work due to an illness are categorized as having received an \(\varepsilon\)-shock.

\(^{26}\)These three diseases lead to the most mean medical expenditures, relative to other health conditions reported in the data.
Impact $\rho$ of a $\varepsilon$-shock on Productivity  Using the criterion for determining $\varepsilon$ and $\varepsilon$-shocks specified above, we use mean earnings of those with a $\varepsilon$-shock relative to those without any health shock to directly estimate $\rho$.

Parameters Calibrated within the Model

In a final step we now use our model to find parameters governing the production function, the $\varepsilon$- and $\varepsilon$-shock distribution, the distribution of preference parameters for exercise, and the terminal value function $v_{T+1}(h')$. The structure of our model allows us to calibrate the parameters in two separate steps. The first part of the estimation consists of finding parameters for the production function and distribution of health shocks, and only involves the static part of the model from section 2.3. This is the case since realized wages and health expenditures in the model are determined in the static part and are independent of effort decisions and the associated health evolution in the dynamic part of the model. In a second step we then employ the dynamic part of the model to estimate the preference distribution for exercise and the terminal value of health.\textsuperscript{27}

Production Function and Health Status  We assume the following parametric form for the production technology:

$$F(t, educ, h, \varepsilon - x) = A(t, educ)h + \frac{(k - (\varepsilon - x))\phi(a,educ)}{h\xi(a,educ)}, \quad 0 < \phi(\cdot), \xi(\cdot) < 1, A(\cdot) > 0.$$  

The production function captures two effects of health on production: the direct effect (first term) and the indirect effect which induces the marginal benefit of health expenditures $x$ to decline with better health (that is $-F_{12} < 0$). The term $A(t, educ)$ allows for heterogeneity in age and education of the effect of health on production and thus wages. Here age can take seven values, $t \in \{1, 2, ..., 7\}$ and we classify individuals into

\textsuperscript{27} Even though we describe the parameters and calibration targets of the different model elements in separate subsections below for expositional clarity, the parameters for production function and health shock distributions are calibrated \textit{jointly}, using the targets in these sections. Similarly, the parameters for exercise preference distribution and marginal value of health at terminal date are calibrated \textit{jointly}, using the observations in both subsections.
two education groups, those that have graduated from high school and those that have not: \( educ \in \{\text{less than High School, High School Grad}\} \). We also allow for differences in marginal effects of medical expenditures on production across education and two broad age groups through parameters \( \phi(a, educ) \) and \( \xi(a, educ) \), where \( a \in \{\text{Young, Old}\} \). We define \textit{Young} as those individuals between the ages of 24 and 41 and the rest as \textit{Old}. This age classification divides our sample roughly in half. We represent the functions \( A(t, educ), \phi(a, educ) \) and \( \xi(a, educ) \) by a full set of age and education dummies.

Since in the unregulated equilibrium the production of individuals (after health expenditures have been made) equals their labor earnings, we use data on labor earnings of households with different health status \( \left( \frac{w(h_2)}{w(h_1)}, \frac{w(h_3)}{w(h_1)}, \frac{w(h_4)}{w(h_1)} \right) \) as well as relative average earnings of the \textit{Young} and the \textit{Old} to pin down the health levels \( \{h_1, h_2, h_3, h_4\} \) in the model.\(^{28}\) Moreover, since \( A(t, educ) \) captures the effects of age \( t \) and education \( educ \) on labor earnings we use conditional (on age and education) earnings to pin down the 14 (7 × 2) parameters \( A(t, educ) \).

In order to determine the values of the dummies representing \( \phi(\cdot) \) and \( \xi(\cdot) \) we recognize that in the model they determine the expenditure cutoffs for the \( \varepsilon \)-shock, as a function of individual health status. Thus we use medical expenditure data to estimate these parameters. More specifically the four parameters representing \( \phi(a, educ) \) are determined to fit the percentage of labor earnings spent on medical expenditure (averaged over \( h \)) for each \( (a, educ) \)-group and the four parameters representing \( \xi(a, educ) \) are chosen to match the percentage of labor earnings spent on medical expenditures (averaged over \( (a, educ) \) groups) for each level \( h \in H \) of household health.\(^{29}\)

**Distribution of Health Shocks** In order to estimate the parameters governing the distribution of health shocks \( \varepsilon \) we exploit the theoretical result from section 2.3 that medical expenditures on these shocks is linear in the shock: \( x^*(\varepsilon, h) = \max\{0, \varepsilon - \bar{\varepsilon}(h)\} \).

\(^{28}\)The categories \{Excellent, Very Good, Good, Fair\} used in the data itself have no cardinal interpretation.

\(^{29}\)Since there is more variation in the data for labor earnings than by health spending by age we decided to use a finer age grouping when estimating \( A(t, educ) \) using wage data than when estimating \( \xi(a, educ) \) and \( \phi(a, educ) \) using health (expenditure) data.
the distribution of medical expenditures $x$ coincides with that of the shocks themselves, above the endogenous health-specific threshold $\bar{\varepsilon}(h)$. French and Jones (2004) argue that the cross-sectional distribution of health care costs can best be fitted by a log-normal distribution (truncated at the upper tail). We therefore assume that the health shocks $\varepsilon$ follow a truncated log-normal distribution:

$$f(\varepsilon; \mu_\varepsilon, \sigma_\varepsilon, \bar{\varepsilon}, \bar{\varepsilon}) = \frac{1}{\varepsilon \sigma_\varepsilon} \phi \left( \frac{\ln \varepsilon - \mu_\varepsilon}{\sigma_\varepsilon} \right) \Phi \left( \frac{\ln \bar{\varepsilon} - \mu_\varepsilon}{\sigma_\varepsilon} \right) - \Phi \left( \frac{\ln \varepsilon - \mu_\varepsilon}{\sigma_\varepsilon} \right)$$

where $\phi$ and $\Phi$ are standard normal pdf and cdf. We then choose the mean and standard deviation $(\mu_\varepsilon, \sigma_\varepsilon)$ of the shocks such that the endogenously determined mean and standard deviation of medical expenditures in the model matches the mean and standard deviation of health expenditures for those with $\varepsilon$-shocks from the data.

For the catastrophic health shock $z$, apart from the probability of receiving it (which was determined in section 2.5.2), only the mean expenditures $\mu_z(h)$ matter. We use the percentage of labor income spent on catastrophic medical expenditures, conditional on health status $h$, to determine these.

**Distribution of Exercise Preference Parameters**  With estimates of the production function and health shock distributions in hand we now calibrate the preference for exercise distribution, using the dynamic part of the model. We assume that the effort utility cost function takes the form

$$\gamma q(e) = \gamma \left[ \frac{1}{1 - e} - (1 + e) \right].$$

The functional form for $q$ guarantees that $q''(e) > 0$, that $q(0) = q'(0) = 0$ and that $\lim_{e \to 1} q'(e) = \infty$. We assume that for each education group the preference shock $\gamma$ can take two (education-specific) values, $\gamma \in \{\gamma_1(\text{educ}), \gamma_2(\text{educ})\}$. We treat these values (4 in total) as parameters. The initial joint distribution $\Phi_0$ over types $(h, \text{educ}, \gamma)$ is then determined by the eight numbers $\Phi_0(\gamma_1|\text{educ}, h)$ that give the fraction of low cost ($\gamma_1$) individuals for each

---

30They use HRS and AHEAD data. Health care costs include health insurance premia, drug costs and costs for hospital, nursing home care, doctor visits, dental visits and outpatient care.
of the eight \((educ, h)\)-combinations. Thus we have to a total of 12 parameters determining preference heterogeneity in the model. We choose the initial distribution \(\Phi_0(\gamma|educ, h)\) so that model effort levels match mean effort levels in period 1 (ages 24-29), conditional on health (4 targets) and conditional on education (2 targets), and mean effort levels in period 7 (ages 60-65), conditional on education (2 targets) in the data. To pin down the four values \(\gamma(educ)\), we use the aggregate mean and standard deviation of effort in period 1, and the measure of households with fair and excellent health in the last period, \(t = 7\).

Marginal Value of Health at Terminal Date  As discussed above, absent direct benefits from better health upon retirement households in the model have no incentive to exert effort, whereas in the data we still see a significant amount of exercise for those of ages 60 to 65. By introducing a terminal and health dependent continuation utility \(v_{T+1}(h)\) this problem can be rectified. Given the structure of the model and the parametric form of the health transition function \(Q(h'|h, e)\) only the differences in the continuation values

\[
\Delta_i = v_{T+1}(h_i) - v_{T+1}(h_{i-1}), \text{ for } i = 2, 3, 4
\]

matter for the choice of optimal effort in the last period \(T\). We choose the \(\Delta_2, \Delta_3, \Delta_4\) such that the model reproduces the health-contingent average effort levels of the 60 to 65 year olds, for \(h_2, h_3, h_4\).

The data targets and associated model parameters are summarized in Tables B.7 and B.8. The estimated parameter values are reported in Table B.9, together with their performance in matching the empirical calibration targets.

2.5.3 Model Fit

Our model is fairly richly parameterized (especially along the production function/labor earnings dimension). It is therefore not surprising that it fits life cycle earnings profiles well. We have also targeted effort levels for very young and very old households (the latter by health status), but have not used data on \(h\)-specific effort levels (apart from at the final
pre-retirement age) in the estimation. How well the model captures the age-effort dynamics is therefore an important “test” of the model. Figures 2.5 (for mean effort) and B.13- B.16 in appendix B.7.1 (for effort by health status) plot the evolution of effort (exercise) over the life cycle both in the data and in the model. The dotted lines show the one-standard deviation confidence bands. From Figure 2.5 we see that our model fits the average exercise level over the life cycle very well, and Figures B.13- B.16 show the same to be true for effort conditional on *Very Good* and *Excellent* health. For households with *Fair* and *Good* health the model fit is not quite as good as that for the *Very Good* and the *Excellent* health groups, but still within the one-standard deviation confidence bands (which are arguably quite wide though, on account of smaller samples once conditioning both on age and health).31

### 2.6 Results of the Policy Experiments: Insurance, Incentives and Welfare

After having established that the model provides a good approximation to the data for the late 1990’s and early 2000’s in the absence of non-discrimination policies, we now use it to answer the main counterfactual question of this paper, namely, what are the effects of introducing these policies (one at a time and in conjunction) on aggregate health, consumption and effort, their distribution, and ultimately, on social welfare.

The primary benefit of the non-discrimination policies is to provide consumption insurance against bad health, resulting in lower wages and higher insurance premia in the competitive equilibrium. However, these policies weaken incentives to exert effort to lead a healthy life, and thus worsen the long run distribution of health, aggregate productivity and thus consumption. In the next two subsections, we present the key quantitative indicators measuring this trade-off: first, the insurance benefits of policies, and second, the adverse incentive effects on aggregate production and health. Then, in subsection 2.6.3,

---

31For *Fair* and *Good* health, our model predicts higher exercise level between the ages of 30 and 54 than in the data. This is partly due to a composition effect: in the second period of life, many workers with low disutility for exercise have fair health and exercise a lot, leading to an increase in the average exercise level for the fair health group. One mechanical way of rectifying this problem would be to let the values the taste parameter $\gamma$ can take on vary with age, reflecting differences in taste for exercise at different stages of life.
we display the welfare consequences of our policy reforms. In the main text we focus on weighted averages of the aggregate variables and welfare measures across workers of different \((\text{educ}, \gamma)\)-types, and document the disaggregated results (which are qualitatively, and to a great extent, quantitatively similar to the averaged numbers) in appendix B.7.2.

### 2.6.1 Insurance Benefits of Policies

Turning first to the consumption insurance benefits of both policies, we observe from figure 2.6 that the combination of both policies is indeed effective in providing perfect consumption insurance. As in the social planner problem, within-group consumption dispersion, as measured by the coefficient of variation, is zero for all periods over the life cycle if both a no-prior conditions law and a no-wage discrimination law are in place (the lines for the social planner solution and the equilibrium under both policies lie on top of one another and are identically equal to zero).\(^{32}\) This is of course what the theoretical analysis in sections 2.3 and 2.4 predicted. Also notice from figure 2.6 that a wage non-discrimination law alone goes a long way towards providing effective consumption insurance, since the effect of differences in health levels on wage dispersion is significantly larger than the corresponding dispersion in health insurance premia. Thus, although a no-prior conditions law in isolation provides some consumption insurance and reduces within-group consumption dispersion by about 30\%, relative to the unregulated equilibrium, the remaining health-induced consumption risk remains significant.

Another measure of the insurance benefits provided by the non-discrimination policies is the level of cross-subsidization or implicit transfers: workers do not necessarily pay their own competitive (actuarially fair) price of the health insurance premium or/and they are not fully compensated for their productivity. Under no-prior conditions policy, as established theoretically in Proposition 6, the healthy workers subsidize the premium of the unhealthy. Similarly, wages of the unhealthy workers are subsidized by the healthy, productive workers.

---

\(^{32}\)Due to the presence of heterogeneity in education levels and preferences the economy as a whole displays non-trivial consumption dispersion even in the presence of both policies (as it does in the solution of the restricted social planner problem).
under the no-wage discrimination policy. Moreover, under both policies, there is cross-
subsidization in both health insurance premia and wages.

Figures 2.7 and 2.8 plot the degree of cross-subsidization over the life cycle, both for
households with excellent and those with fair health, and Table B.11 in appendix B.7.2
summarizes the transfers for all health groups. The plots for the health insurance premium
measures the differences between the actuarially fair health insurance premium a particular
health type household would have to pay and the actual premium paid in the presence of
either a no-prior conditions policy or the presence of both policies. Similarly, the wage plots
display the difference between the productivity of the worker (and thus her wage in the
unregulated equilibrium) and the wage received under a no-wage discrimination policy and
in the presence of both policies. Negative numbers imply that the worker is paying a higher
premium, or is paid lower wage than in a competitive equilibrium without government
intervention. Thus such a worker, in the presence of government policies, has to transfer
resources to workers of different (lower) health types. Reversely, positive numbers imply
that a worker is being subsidized, i.e., she is paying a lower premium and is paid higher
wage.

We observe from Figure 2.7 that the workers with excellent health significantly cross-
subsidize the other workers, both in terms of cross-subsidies in health insurance premia as
well as in terms of wage transfers. To interpret the numbers quantitatively, note that average
consumption of the excellent group is 1.04 when young and 1.75 when aged 42-47. Thus the
wage transfers delivered by this group amount to 12 – 14% of average consumption when
young and close to 30% in prime working age (note that the share of workers in excellent
health in the population has shrunk at that age, relative to when this cohort of workers was
younger). From figure 2.7 we also observe that the implicit transfers induced by a no-prior
conditions law are still significant (they amount to 3-7% of consumption for young workers
of excellent health, and 4-10% when middle-aged), but quantitatively smaller than those
implied by wage-nondiscrimination legislation.

Figure 2.8 displays the same plots for households of fair health. These households are
the primary recipients of the transfers from workers with excellent health,\textsuperscript{33} and for this group (which is small early in the life cycle but grows over time) the transfers are massive. In terms of their average competitive equilibrium consumption, the implicit health insurance premium subsidies amount to a massive 37-60\% and the wage transfers amount to a staggering 65-75\% of pre-policy average consumption of this group. Although these transfers shrink (as a fraction of pre-policy consumption) over the life cycle as the share of households with fair health increases and that with excellent health declines, they continue to account for a significant part of consumption for households of fair health. These numbers indicate that the insurance benefits from both policies, and specifically from the wage nondiscrimination law, will be substantial.

An interesting property of the subsidies is that the level of subsidization implied by a given policy is higher when only one of the non-discrimination laws is enacted, relative to when both policies are present. This is especially true for the no-prior conditions law and is due to the fact that the government insures the workers with bad health through an inefficient level of medical expenditure.

Thus far, we have discussed the insurance benefits of the non-discrimination policies. In the next subsection, we analyze the aggregate dynamic effects of the policies on production and the health distribution.

\textbf{2.6.2 Adverse Incentive Effects on Aggregate Production and Health}

The associated incentive costs from each policy are inversely proportional to their consumption insurance benefits, as figure 2.9 shows. In this figure we plot the average exerted effort over the life cycle, in the socially optimal and the equilibrium allocations under the various policy scenarios. In a nutshell, effort is highest in the solution to the social plan-

\textsuperscript{33}Table B.11 in the appendix shows that households with \textit{very good} health are also called upon to deliver transfers, albeit of much smaller magnitude, and workers with \textit{good} health are on the receiving side of (small) transfers. As the cohort ages the share of households in these different health groups shifts, and towards the end of the life cycle the now larger group of households with fair health receives subsidies from all other households, at least with respect to health insurance premia.
ner problem, positive under all policies,\textsuperscript{34} but substantially lower in the presence of the non-discrimination laws.

More precisely, two important observations emerge from figure 2.9. First, the policies that provide the most significant consumption insurance benefits also lead to the most significant reductions in incentives to lead a healthy life. It is the very dispersion of consumption due to health differences, stemming from health-dependent wages and insurance premia that induce workers to provide effort in the first place, and thus the policies that reduce that consumption dispersion the most come with the sharpest reduction in incentives. Whereas a no-prior conditions law alone leads to only a modest reduction of effort, with a wage nondiscrimination law in place the amount of exercise household find optimal to carry out shrinks more significantly. Finally, if both policies are implemented simultaneously the only benefit from exercise is a better distribution of post-retirement continuation utility, and thus effort plummets strongly, relative to the competitive equilibrium.

The second observation we make from figure 2.9 is that the impact of the policies on effort is most significant at young and middle ages, whereas towards retirement effort levels under all polices converge. This is owed to the fact that the direct utility benefits from better health materialize at retirement and are independent of the nondiscrimination laws (but heavily discounted by our impatient households), whereas the productivity and health insurance premium costs from worse health accrue through the entire working life and are strongly affected by the different policies.\textsuperscript{35}

Given the dynamics of effort over the life cycle (and a policy invariant initial health distribution), the evolution of the health distribution is exclusively determined by the health transition function $Q(h' ; h, e)$. Figure 2.10 which displays average health in the economy under the various policy scenarios is then a direct consequence of the effort dynamics from Figure 2.9. It shows that health deteriorates under all policies as a cohort ages, but more

\textsuperscript{34}Recall that, relative to the theoretical analysis, we have introduced a terminal value of health which induces not only effort in the last period even under both policies, but through the continuation values in the dynamic programming problem, positive effort in all periods. How quantitatively important this effect is for younger households depends significantly on the time discount factor $\beta$.

\textsuperscript{35}In fact, absent the terminal (and policy invariant) direct benefits from better health the differences in effort levels across policies remain fairly constant over the life cycle.
rapidly if a no-prior conditions law and especially if a wage nondiscrimination law is in place. As with effort, the conjunction of both policies has the most severe impact on public health.

Figure 2.12 demonstrates that the decline of health levels over the life cycle also induce higher expenditures on health (insurance) later in life. The level of these expenditures (and thus their relative magnitudes across different policies) are determined by two factors, a) the health distribution (which evolves differently under alternative policy scenarios) and b) the equilibrium health expenditures, which are fully determined by the thresholds $\tilde{\epsilon}(h)$ from the static analysis of the model and that vary across policies. The evolution of health is summarized by figure 2.10, and figure 2.11 displays the health dependent thresholds $\tilde{\epsilon}(h)$ for the youngest households. Recall from section 2.3 that the thresholds $\tilde{\epsilon}(h)$ under the unregulated competitive equilibrium and the equilibrium with both policies are socially efficient and thus the three graphs completely overlap. Also observe that, relative to the efficient allocation (=unregulated equilibrium) under the no-prior conditions law workers with low health are strongly over-insured (they have lower thresholds, $\tilde{\epsilon}^{NP}(h_i) > \tilde{\epsilon}^{SP}(h_i)$ for $i = 1, 2$) and workers with very good and excellent health are slightly under-insured. This was the content of Proposition 6, and it is quantitatively responsible for the finding that health expenditures are highest under this policy. The reverse is true under a no-wage discrimination law: low health types are under-insured and high types are over-insured, relative to the social optimum, but quantitatively these differences are minor.

Finally, figures 2.13 and 2.14 display aggregate production and aggregate consumption over the life cycle. Since the productivity of each worker depends on her health and on the non-treated fraction of her health shock, aggregate output is lower, ceteris paribus, under policy configurations that lead to a worse health distribution and that leave a larger share of health shocks $\varepsilon$ untreated. From figure 2.13 we observe that the deterioration of health under a policy environment that includes a wage nondiscrimination policy is especially severe, in line with the findings from figure 2.10. Interestingly, the more generous health

---

36 The figures are qualitatively similar for older cohorts.
insurance (for those of fair and good health) under a no-prior conditions law alone leads to output that even exceeds that in the unregulated equilibrium, despite the fact that the health distribution under that policy is (moderately) worse. But health expenditures of course command resources that take away from private consumption, and as figure 2.14 shows, resulting aggregate consumption over the life cycle under this policy is substantively identical to that under the wage discrimination law (and the consumption allocation is more risky under the no-prior conditions legislation). Relative to the unregulated equilibrium both policies thus entail a significant loss of average consumption in society (in one case, because less is produced, in the other case because more resources are spent on productivity enhancing health goods); the same is even more true if both policies are introduced jointly.

Overall, the effect on aggregate effort, health, production and thus consumption suggests a quantitatively important trade-off between consumption insurance and incentives. Within the spectrum of all policies, the unregulated equilibrium provides strong incentives at the expense of risky consumption, whereas a policy mix that includes both policies provides full insurance at the expense of a deterioration of the health distribution. The effects of the no-prior conditions law on both consumption insurance and incentives are modest, relative to the unregulated equilibrium. In contrast, implementing a no wage discrimination law or both policies insures away most of the consumption risk, but significantly reduces (although does not eliminate completely) the incentives to exert effort to lead a healthy life, especially early in the life cycle. In the next subsection we will now document how these two quantitatively sizable but countervailing effects translate into welfare consequences from hypothetical policy reforms.

### 2.6.3 Welfare Implications

In this section we quantify the welfare impact of the policy innovations studied in this paper. For a fixed initial distribution $\Phi_0(h)$ over health status, denote by $W(c, e)$ the expected

---

37 Recall that we carry out our analysis for each $(educ, \gamma)$-type separately and report averages across these types. Thus in what follows $\Phi_0$ suppresses the (policy-independent) dependence of the initial distribution on $(educ, \gamma)$. 

---
lifetime utility of a cohort member (where expectations are taken prior to the initial draw $h$ of health) from an arbitrary allocation of consumption and effort over the life cycle.\textsuperscript{38} Our consumption-equivalent measure of the welfare consequences of a policy reform is given by

$$W(c^{CE}(1 + CEV^i), e^{CE}) = W(c^i, e^i)$$

where $i \in \{SP, NP, NW, Both\}$ denotes the policy scenario under consideration. Thus $CEV^i$ is the percentage reduction of consumption in the competitive equilibrium consumption allocation required to make households indifferent (ex ante) between the competitive equilibrium allocation\textsuperscript{39} and that arising under policy regime $i$.

In order to emphasize the importance of the dynamic analysis in assessing the normative consequences of different policies we also report the welfare implications of the same policy reforms in the static version of the model in section 2.3. Similar to the dynamic consequences we compute the static consumption-equivalent loss (relative to the competitive equilibrium) as

$$U(c^{CE}(1 + SCEV^i)) = U(c^i)$$

where $U(c)$ is the expected utility from the period 0 consumption allocation\textsuperscript{40}, under the cross-sectional distribution $\Phi_0$, and thus is determined by the static version of the model.\textsuperscript{41}

\textsuperscript{38}That is, using the notation from section 2.4, for the socially optimal allocation

$$W(c^{SP}, e^{SP}) = V(\Phi_0)$$

and for equilibrium allocations, under policy $i$,

$$W(c^i, e^i) = \int v^i_0(h)d\Phi_0.$$ 

\textsuperscript{39}Recall that even the social planner problem is solved for each specific $(\gamma, educ)$ group separately and thus also does not permit ex-ante insurance against unfavorable $(\gamma, educ)$-draws. We consider this restricted social planner problem because we view the results are better comparable to the competitive equilibrium allocations.

\textsuperscript{40}In the static version of the model effort is identically equal to zero in the social planner problem and in the equilibrium under all policy specifications, and therefore disutility from effort is irrelevant in the static version of the model.

\textsuperscript{41}Thus, using the notation from section 2.3

$$U(c^{CE}) = U^{CE}(\Phi_0) \text{ for } i \in \{SP, NP, NW, Both\}$$
Therefore $SCEV^i$ provides a clean measure of the static gains from better consumption insurance induced by the policies against which the dynamic adverse incentive effects have to be traded off.

The static welfare consequences reported in the first column of Table 2.1 that isolate the consumption insurance benefits of the policies under consideration are consistent with the consumption dispersion displayed in Figure 2.6. Perfect consumption insurance, as implemented in the solution to the social planner problem and also achieved if both policies are implemented jointly, are worth close to 6% of unregulated equilibrium consumption. Each policy in isolation delivers a substantial share of these gains, with the no wage discrimination law being more effective than the no-prior conditions law.

Turning now to the main object of interest, the dynamic welfare consequences (column 2 of Table 2.1) paint a somewhat different picture. Consistent with the static analysis, both policies improve on the laissez-faire equilibrium, and the welfare gains are substantial, ranging from 6% to 9.5% of lifetime consumption. The sources of these welfare gains are improved consumption insurance (as in the static model) and reduced effort (which bears utility costs), which outweigh the reduction in average consumption these policies entail (recall Figure 2.14). Furthermore, as in the static model a wage nondiscrimination law dominates a no-prior conditions law. In light of Figures 2.14 and 2.6 this does not come as a surprise: both policies imply virtually the same aggregate consumption dynamics, but the no-prior conditions law provides substantially less consumption insurance.

But what we really want to stress is that there are crucial differences to the static analysis. First and foremost, it is not optimal to introduce a no-prior conditions law once a wage non-discrimination law is already in place. The latter policy already provides effective (albeit not complete) consumption insurance, and the further reduction of incentives and associated mean consumption implied by the no prior conditions law makes a combination of both policies suboptimal. The associated welfare losses of pushing social insurance too far

\[ U(c^{CE}) = \int U^{CE}(h) d\Phi_0. \]

108
amount to about 1.3% of lifetime consumption.\footnote{It should be stressed that these conclusions follow under the maintained assumption that a wage nondiscrimination law is indeed fully successful in curbing health-related wage variation, and does so completely costlessly.} Finally we see that in contrast to the static case the best policy combination (a wage nondiscrimination law alone) does not come close to providing welfare as high as the social optimum: the gap between these two scenarios turns out to about 7% of lifetime consumption. This gap is due to inefficiently little consumption insurance, inefficiently low aggregate consumption and an inefficient health expenditure allocation (see again Figure 2.11), although the latter effect is quantitatively modest. This effect is however quantitatively crucial in explaining why the no-prior conditions law in isolations fares worse than the wage nondiscrimination policies (and a combination of both policies, which restores efficiency in health expenditures, recall proposition 8).

The welfare consequences reported in Table 2.1 were measured under the veil of ignorance, before workers learn their initial health level. They mask very substantial heterogeneity in how workers feel about these policies once their initial health status in period 0 has been revealed. Given the transfers across health types displayed in Figures 2.7 and 2.8 and the persistence of health status this is hardly surprising. Table 2.2 quantifies this heterogeneity by reporting dynamic consumption-equivalent variation measures, computed exactly as before, but now computed after the initial health status has been materialized. Broadly speaking, the lower a worker’s initial health status, the more she favors policies providing consumption insurance. For the middle two health groups the ranking of policies coincides with that in the second column of Table 2.1; households with excellent health prefer only the no prior conditions law (and thus only very moderate implicit transfers) to the unregulated equilibrium, whereas young households with fair health would support the simultaneous introduction of both policies. The differences in the preference for different policy scenarios across different $h$-households are quantitatively very large: whereas fair-health types would be willing to pay 54% of laissez faire lifetime consumption to see both policies introduced, households of excellent health would be prepared to give up 4.5% of lifetime consumption to prevent exactly this policy innovation.
2.7 Conclusion

In this paper, we studied the effect of labor and health insurance market regulations on evolution of health and production, as well as welfare. We showed that both a no-wage discrimination law (an intervention in the labor market), in combination with a no-prior conditions law (an intervention in the health insurance market) provides effective consumption insurance against health shocks, holding the aggregate health distribution in society constant. However, the dynamic incentive costs and their impact on health and medical expenditures of both policies, if implemented jointly, are large. Even though both policies improve upon the laissez-faire equilibrium, implementing them jointly is suboptimal (relative to introducing a wage nondiscrimination in isolation). We therefore conclude that a complete policy analysis of health insurance reforms on one side and labor market (non-discrimination policy) reforms cannot be conducted separately, since their interaction might prove less favorable despite welfare gains from each policy separately.

These conclusions rest in part on our assumption that both policies can be implemented optimally at no direct overhead cost. To us, this assumption seems potentially more problematic for the no-wage discrimination policy than the no-prior conditions policy because match-specificity between a worker and a firm appears to be more important than between a worker and a health insurance company. One can likely implement the no-prior conditions policy through the health insurance exchanges proposed by Obama Care in which a government agency links those seeking health insurance to health insurance providers and thereby overcomes, at low cost, the incentives of the health insurance companies to cherry-pick their clients. However, a similar institution (e.g. something akin to a union hall type institution), is likely to demand higher costs, given the specificity in most worker-firm matches. In addition, the average output produced by a worker-firm pair is much larger than the expenses involved in health insurance (both in our model as well as in the data).\textsuperscript{43}

Finally, our analysis of health insurance and incentives over the working life has ignored

\footnote{\textsuperscript{43}To put these potential costs in perspective, from our quantitative results it follows that if as little as 3\% of production was consumed in implementing the no-wage discrimination policy (and the no prior conditions policy is cost-free), then it is the latter policy that would constitute the ex ante preferred policy option.}
several potentially important avenues through which health and consumption risk affect welfare. First, the benefits of health in our model are confined to higher labor productivity, and thus we model the investment motives into health explicitly. It has abstracted from an explicit modeling of the benefits better health has on survival risk, although the positive effect of health $h$ on the continuation utility after retirement partially captures this effect in our model, albeit in a fairly reduced form. Similarly, better health might have a direct effect on flow utility during working life.\footnote{As we argue in appendix B.8 at least in one extension of the model introducing a direct flow utility benefit from better health leaves our analysis qualitatively unchanged.} Finally, in our analysis labor income risk directly translates into consumption risk, in the absence of household private saving. We conjecture that the introduction of self-insurance via precautionary saving against this income risk further weakens the argument in favor of the policies studied in this paper. Future work has to uncover whether such an extension of the model also affects, quantitatively or even qualitatively, our conclusions about the relative desirability of these policies.
2.8 Tables and Figures

Figure 2.2: Production Function $F(h, \varepsilon - x)$ for fixed $\varepsilon$

Figure 2.3: Timing of the Model
$\varepsilon_t$ drawn according to $g(h_t)$ and $f(\varepsilon_t)$

$t$

$h_t$

Households choose $\varepsilon_t$

Firms offer wage $w(h_t)$ and HI contract $\{x(\varepsilon_t, h_t), P(h_t)\}$

$t+1$

$h_{t+1}$

determined by $Q(h_{t+1}, h_t; \varepsilon_t)$

Households produce $F(h_t, \varepsilon_t - x_t(\varepsilon_t))$ and consume $c_t(h_t)$

$x_t(\varepsilon_t, h_t)$ spent
Figure 2.4: Optimal Health Expenditure and Production

Figure 2.5: Average Effort in Model and Data
Figure 2.6: Consumption Dispersion

Figure 2.7: Subsidy: Excellent

Figure 2.8: Subsidy: Fair

Figure 2.9: Effort

Figure 2.10: Average Health
Figure 2.11: Cutoffs

Figure 2.12: Health Spending

Figure 2.13: Production

Figure 2.14: Consumption
Table 2.1: Aggregate Welfare Comparisons

<table>
<thead>
<tr>
<th></th>
<th>Static $CEV^i$</th>
<th>Dynamic $CEV^i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social Planner</td>
<td>5.6527</td>
<td>16.4799</td>
</tr>
<tr>
<td>Competitive Equilibrium</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>No Prior Conditions Law</td>
<td>4.1593</td>
<td>6.9782</td>
</tr>
<tr>
<td>No Wage Discrimination Law</td>
<td>5.3486</td>
<td>9.5399</td>
</tr>
<tr>
<td>Both Policies</td>
<td>5.6527</td>
<td>8.1656</td>
</tr>
</tbody>
</table>

Table 2.2: Welfare Comparison in the Dynamic Economy Conditional on Health

<table>
<thead>
<tr>
<th></th>
<th>Fair</th>
<th>Good</th>
<th>Very Good</th>
<th>Excellent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social Planner</td>
<td>56.5681</td>
<td>13.7796</td>
<td>14.4002</td>
<td>10.5597</td>
</tr>
<tr>
<td>Competitive Equilibrium</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>No Prior Conditions Law</td>
<td>36.3452</td>
<td>7.9579</td>
<td>4.2954</td>
<td>0.5892</td>
</tr>
<tr>
<td>No Wage Discrimination Law</td>
<td>45.8741</td>
<td>14.4826</td>
<td>6.6942</td>
<td>-1.8221</td>
</tr>
<tr>
<td>Both Policies</td>
<td>54.2835</td>
<td>13.2129</td>
<td>5.0420</td>
<td>-4.4532</td>
</tr>
</tbody>
</table>
A.1 Long-Term Changes in Tax Rates

Since the analysis conducted in the paper is a steady-state analysis, the aggregate statistics in 2000 might be the outcome of government policies in previous years. In this section, I document long-term changes in tax rates in the UK and CE presented in Figures in Mendoza and Tesar (2005) and an OECD report (2012).63

There is a consistent trend of a higher capital income tax rate (Figure A.1) and lower labor income tax rate (Figure A.2) in the United Kingdom compared to France, Italy, and Germany. Moreover, top combined statutory personal income tax rates (an indication of the degree of progressiveness in labor income tax code) shows similar patterns in 1980s as it does in 2000 across countries (Figure A.3).

A.2 Summary Statistics of Continental European Countries

In the quantitative exercises, I analyze the tax competition between the UK and Continental European countries comprising France, Italy, Germany, Spain, and Sweden. These are the countries in the European Union with comparable GDP per capita so that I could focus more on the economic incentives for households to move between the two regions.

63Figure A.1 and A.2 are from Mendoza and Tesar (2005), and Figure A.3 from OECD.
Moreover, the selected five countries have relatively higher labor (more progressive) income tax rates, and lower capital income tax rates than the UK. In getting the composite measures of macroeconomic outcomes and tax policies in the CE, I use weighted averages of these countries, where weights are obtained from the GDP and population of individual countries.

**Aggregate Statistics** Table A.1 presents the summary statistics for five Continental European countries that I analyze in the main part of the paper. Most of data are from IMF, OECD, and a series of papers under the project Cross Sectional Facts for Economists published in 2011.\(^{64}\) The intergenerational persistence in schooling measures are taken from...

\(^{64}\)Summary of the project is presented in Krueger, et al (2010). Moreover, I obtain data for individual countries from Blundell and Etheridge (2010), Domeij and Floden (2010), Fuchs-Schuendeln, et al. (2010), Jappelli and Pistaferri (2010), and Pijoan-Mas and Sanchez-Marcos (2010) which summarize macroeconomic facts in Britain, Sweden, Germany, Italy, and Spain, respectively.
Causa and Johansson (2009) who use 2005 OECD data to obtain a percentage increase in probability of a child going to college when his father is a college graduate rather than upper-secondary educated.

<table>
<thead>
<tr>
<th>GDP per capita(^a)</th>
<th>France</th>
<th>Germany</th>
<th>Italy</th>
<th>Spain</th>
<th>Sweden</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skill Premium(^b)</td>
<td>1.38</td>
<td>1.51</td>
<td>1.48</td>
<td>1.61</td>
<td></td>
</tr>
<tr>
<td>Intergen. Schooling Persistence(^c)</td>
<td>0.18</td>
<td>0.30</td>
<td>0.21</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>Gini of Labor Earnings(^b)</td>
<td>0.35</td>
<td>0.34</td>
<td>0.40</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td>% College Graduates(^d)</td>
<td>22 %</td>
<td>24%</td>
<td>9 %</td>
<td>22.6%</td>
<td>30.1%</td>
</tr>
<tr>
<td>Tertiary Educ. Exp.(^d)</td>
<td>1.32%</td>
<td>1.08%</td>
<td>0.83%</td>
<td>0.89%</td>
<td>1.27%</td>
</tr>
<tr>
<td>Public Spending</td>
<td>1.13%</td>
<td>0.94%</td>
<td>0.70%</td>
<td>0.72%</td>
<td>1.12%</td>
</tr>
<tr>
<td>Private Spending</td>
<td>0.18%</td>
<td>0.15%</td>
<td>0.13%</td>
<td>0.17%</td>
<td>0.15%</td>
</tr>
<tr>
<td>Average Hours Worked(^d)</td>
<td>0.333</td>
<td>0.307</td>
<td>0.388</td>
<td>0.363</td>
<td>0.341</td>
</tr>
<tr>
<td>Relative Weights</td>
<td>0.24</td>
<td>0.34</td>
<td>0.23</td>
<td>0.15</td>
<td>0.04</td>
</tr>
</tbody>
</table>

\(^a\) IMF  
\(^b\) Cross Sectional Facts for Economists (2011)  
\(^c\) Causa and Johansson (2009)  
\(^d\) OECD

**Labor Income Taxation**  Figure A.4 plot the labor income tax functions

\[
\tau(y) = 1 - a_0 \left( \frac{y}{\bar{y}} \right)^{-a_1}
\]

of the individual Continental European countries which are imputed using top marginal income tax rates and thresholds at multiples of average labor earnings \((\bar{y})\), following the procedure described in Guvenen et al (2011). These countries have progressive labor income tax schedules with Germany having the highest labor income tax rates and France, the most progressive.

**Capital Income Taxation**  In Table A.2, capital income tax rates in individual Continental European countries are reported. There are several ways of calculating capital income tax rates. Since different capital incomes are taxed at different rates, it is difficult to find a single measure of capital income tax rates. Here, I use tax rates reported in Carey and Tchilinguirian (2000) who apply the methodology presented in Mendoza et al. (1994)
to find average effective tax rates between the years of 1991 and 1997.

Mendoza et al. calculates effective tax rates by dividing sum of corporate income and personal capital income taxes by net operating surplus. These rates are not uncontroversial, but they are often used in macroeconomics literature.

Table A.2: Capital Income Tax Rates in Continental European Countries

<table>
<thead>
<tr>
<th>Country</th>
<th>Capital Income Tax Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>France</td>
<td>26.8</td>
</tr>
<tr>
<td>Germany</td>
<td>25.1</td>
</tr>
<tr>
<td>Italy</td>
<td>33.1</td>
</tr>
<tr>
<td>Spain</td>
<td>21.5</td>
</tr>
<tr>
<td>Sweden</td>
<td>52.7</td>
</tr>
</tbody>
</table>

A.3 Computation of the Nash Equilibrium

I describe the computation procedure of finding the Nash equilibrium tax systems. The governments in this model maximize the steady-state welfare of households by choosing labor income tax function parameters, $a_0$ and $a_1$, and a capital income tax rate $\tau_k$. Let $a_0 \in \{a_{01}, ..., a_{0m}\}$ and $a_1 \in \{a_{11}, ..., a_{1n}\}$. Since I need to find the best responses taking as given the other country’s tax rates, I iteratively find the tax parameters that maximize steady-state welfare of households in each country and satisfy the government budget balance.
1. Fix Country 2’s tax system \( \{a_2^0, a_2^1, \tau_2^k\} \).

2. Fix Country 1’s labor income tax parameters, \( \{a_1^0, a_1^1\} \).

3. Guess a capital income tax rate for Country 1, \( \tau_1^k \).

4. Guess an after-tax return to capital \( R \).

5. Guess aggregate moments in the economy: the wages \( w^i(\theta) \), average labor earnings \( \bar{y}^i \) (necessary since labor income tax code is a function of average earnings in the economy), and population ratio (since \textit{per capita} budget constraint has to be satisfied).

6. Solve household problems (both countries).

7. Aggregate and check if the guesses for wages, average labor earnings, and population size coincide.

8. If they do, check if the bond market clears.

   - If not, go back to Step 4, and update guesses for wages, average labor earnings, and population ratio.

9. If the bond market clears, check if the government budget constraint is satisfied.

   - If the bond market does not clear, go back to Step 3, and update the guess for the after-tax return to capital.

10. If the government budget constraint is satisfied, then go back to Step 2, and repeat the procedure for the next set of labor income tax parameters in Country 1.

    - If not, go back to Step 3, and make a new guess for the capital income tax rate.

11. Find the level of \( \{a_1^0, a_1^1\} \) and the budget-balancing \( \tau_1^k \) that maximize the steady-state welfare of households in Country 1 (Best Response).
12. Taking as given the tax system of Country 1, repeat the procedure (Steps 1 through 11) for Country 2.

13. Iterate until the tax systems arrive at a fixed point.

A.4 Constrained Efficiency in a Three-Factor, Open-Economy Model

Here, I consider a simplified version of my model to provide an explanation for why in the Nash equilibrium, the UK sets a lower capital income tax rate than the CE does.

I simplify my model to a two-period analysis to provide a condition for constrained efficiency which shed light on why $\tau_{k}^{*\text{UK}} < \tau_{k}^{*\text{CE}}$.

Assume that:

- the only sources of heterogeneity are skill level and initial wealth, $(\theta, a)$, with their initial distribution being $\Phi(\theta, a)$;
- labor is given exogenously with its endowment, either $\varepsilon_1$ or $\varepsilon_2$, occurring with probabilities $\pi_1$ and $\pi_2 = 1 - \pi_1$, respectively; and
- the social planner makes savings decision for households $a'(\theta, a)$.

With a three-factor production function and two kinds of labor (the unskilled and skilled labor), as used in the benchmark model, since the production function is homogeneous of degree one,

$$f_{kk}K + f_{uk}U + f_{sk}S = 0$$

is satisfied.

Here, households are either unskilled or skilled. Then, the planner chooses $a'(\theta, a)$ that solves

$$\max_{a'(\theta, a) \in [0, a]} \int_{(\theta, a)} \left[ u(a - a'(\theta, a)) + \beta \sum_{i=1,2} \pi_i u(f_k(K, U, S)a'(\theta, a) + f_\theta(K, U, S)\varepsilon_i) \right] \Phi(d(\theta, a))$$
where

\[
K = \int_{(\theta,a)} a(\theta,a) \Phi(d(\theta,a)),
\]

\[
U = \int_{(U,a)} (\pi \varepsilon_1 + (1 - \pi) \varepsilon_2) \Phi(d(U,a)),
\]

\[
S = \int_{(S,a)} (\pi \varepsilon_1 + (1 - \pi) \varepsilon_2) \Phi(d(S,a)).
\]

The first order condition of the social planner’s problem is:

\[
u'(c(\theta,a)) = \beta \left( \sum_{i=1,2} \pi_i u'(c_i(\theta,a)) \right) f_k(K,U,S) + \Delta,
\]

where

\[
\Delta = \beta \int_{(\theta,a)} \sum_{i=1,2} \pi_i u'(\cdot) \left( f_k(\cdot) a(\theta,a) + f_\theta(\cdot) \varepsilon_i \right) \left[ f_{kk}(\cdot) a(\theta,a) + f_{k\theta}(\cdot) \varepsilon_i \right] \Phi(d(\theta,a))
\]

by \( f_{kk}K + f_{uk}U + f_{sk}S = 0 \), and with some algebra,

\[
\Delta = \beta \int_{(\theta,a)} \sum_{i=1,2} \pi_i u'(\cdot) \left( f_k(\cdot) a(\theta,a) + f_\theta(\cdot) \varepsilon_i \right) \left[ f_{kk}(\cdot) a(\theta,a) + f_{k\theta}(\cdot) \varepsilon_i \right] \Phi(d(\theta,a))
\]

The relative labor income terms for each skill level (\( \varepsilon_i U \) and \( \varepsilon_i \)) are weighted by their relative importance (efficiency) in production (\( \frac{AU^\rho}{AU^\rho + (1 - A)S^\rho} \) and \( \frac{(1 - A)S^\rho}{AU^\rho + (1 - A)S^\rho} \)).

Under the calibrated parameters, the UK has a higher relative efficiency of the skilled in production, and a more dispersed wealth distribution. Thus, the consumption-poor in the UK is relatively more wealth-poor than they are labor income-poor. This is a force that leads the UK to prefer higher capital (therefore, higher wages), more than the CE does.
Appendix B

Appendix to Chapter 2

B.1 Proofs of Propositions

Proposition 1

Proof: Since exercise does not carry any benefits in the static model, trivially $e^{SP} = 0$. Attaching Lagrange multiplier $\mu \geq 0$ to the resource constraint, the first order condition with respect to consumption $c(\varepsilon)$ is

$$u'(c(\varepsilon, h)) = \lambda$$

and thus $c^{SP}(\varepsilon, h) = c^{SP}$ for all $\varepsilon \in E$ and $h \in H$. Thus, not surprisingly, the social planner provides full consumption insurance to households. The optimal health expenditure allocation maximizes this consumption

$$c^{SP} = \max_{x(\varepsilon, h)} \sum_h \left\{ g(h) [F(h, -x(0, h)) - x(0, h)] ight. \right. \left. \left. + (1 - g(h)) \int f(\varepsilon) [F(h, \varepsilon - x(\varepsilon, h)) - x(\varepsilon, h)] d\varepsilon \right\} \Phi(h)$$

Denoting by $\mu(\varepsilon, h) \geq 0$ the Lagrange multiplier on the constraint $x(\varepsilon, h) \geq 0$, the first order condition with respect to $x(\varepsilon, h)$ reads as

$$-F_2(h, \varepsilon - x(\varepsilon, h)) + \mu(\varepsilon, h) = 1$$
Fix $h \in H$. By assumption 4 $F_{22}(h, y) < 0$ and thus either $x(\varepsilon, h) = 0$ or $x(\varepsilon, h) > 0$ satisfying

$$-F_2(h, \varepsilon - x(\varepsilon, h)) = 1$$

for all $\varepsilon$. Thus off corners $\varepsilon - x(\varepsilon, h) = \hat{\varepsilon}^{SP}(h)$ where the threshold satisfies

$$-F_2(h, \hat{\varepsilon}^{SP}(h)) = 1. \tag{B.1}$$

Consequently

$$x^{SP}(\varepsilon, h) = \max \left[ 0, \varepsilon - \hat{\varepsilon}^{SP}(h) \right].$$

The fact that $\hat{\varepsilon}^{SP}(h)$ is increasing in $h$, strictly so if $F_{12}(h, y) > 0$, follows directly from assumption 4 and (B.1).

**Proposition 2**

**Proof:** Attaching Lagrange multiplier $\mu(h)$ to equation (2.11) and $\lambda(h)$ to equations (2.12) the first order conditions read as

$$u'(w(h) - P(h)) = \lambda(h) = -\mu(h) \tag{B.2}$$

$$\lambda(h)F_2(h, -x(0, h)) \leq \mu(h) \tag{B.3}$$

$$= \text{ if } x(0, h) > 0$$

$$\lambda(h)F_2(h, \varepsilon - x(\varepsilon, h)) \leq \mu(h) \tag{B.4}$$

$$= \text{ if } x(\varepsilon, h) > 0$$

Thus off corners we have

$$F_2(h, \hat{\varepsilon} - x(\hat{\varepsilon}, h)) = F_2(h, \varepsilon - x(\varepsilon, h)) = K \tag{B.5}$$
for some constant \( K \). Thus off corners \( \varepsilon - x(\varepsilon, h) \) is constant in \( \varepsilon \) and thus medical expenditures satisfy the cutoff rule

\[
x^{CE}(\varepsilon, h) = \max \left[ 0, \varepsilon - \bar{\varepsilon}^{CE}(h) \right].
\]  

(B.6)

Plugging (B.6) into (B.4) and evaluating it at \( \varepsilon = \bar{\varepsilon}^{CE}(h) \) yields

\[
\lambda(h)F_2(h, \bar{\varepsilon}^{CE}(h)) = \mu(h).
\]  

(B.7)

Using this result in the second part of (B.2) delivers the characterization of the equilibrium cutoff levels

\[
F_2(h, \bar{\varepsilon}^{CE}(h)) = -1 \text{ for all } h \in H
\]

which are unique, given the assumptions imposed on \( F \). Wages, consumption and health insurance premia then trivially follow from (2.11) and (2.12).

**Proposition 5**

**Proof:** Let Lagrange multipliers to equations (2.19) and (2.20) be \( \mu \) and \( \lambda(h) \), respectively. Then, the first order conditions are:

\[
\sum_h u'(w(h) - P)\Phi(h) = \mu
\]

\[
u'(w(h) - P)\Phi(h) = \lambda(h)
\]

\[
(1 - g(h))f(\varepsilon)[-F_2(h, \varepsilon - x(\varepsilon, h))]\lambda(h) \leq \mu(1 - g(h))f(\varepsilon)\Phi(h)
\]

\[
= \text{ if } x(\varepsilon, h) > 0
\]

\[
g(h)[-F_2(h, -x(0, h))]\lambda(h) \leq \mu g(h)\Phi(h)
\]

\[
= \text{ if } x(0, h) > 0
\]

Thus, off-corners we have

\[
F_2(h, \varepsilon - x(\varepsilon, h)) = F_2(h, \hat{\varepsilon} - x(\hat{\varepsilon}, h)) = K
\]
for some constant $K$ and the cutoff rule is determined by

$$u'(w(h) - P)[-F_2(h, \bar{\varepsilon}^{NP}(h))] = \sum_h u'(w(h) - P)\Phi(h). \quad (B.8)$$

Moreover, let us take the derivative of (B.8) with respect to $h$.

$$u''(w(h) - P)\frac{\partial w(h)}{\partial h}F_2 + u'(w(h) - P)\left\{F_{12} + F_{22}\frac{\partial\bar{\varepsilon}^{NP}(h)}{\partial h}\right\} = 0$$

$$u''(w(h) - P)\frac{\partial\bar{\varepsilon}^{NP}(h)}{\partial h}\frac{\partial w(h)}{\partial \bar{\varepsilon}^{NP}(h)}F_2 + u'(w(h) - P)\left\{F_{12} + F_{22}\frac{\partial\bar{\varepsilon}^{NP}(h)}{\partial h}\right\} = 0$$

$$\Rightarrow \frac{\partial\bar{\varepsilon}^{NP}(h)}{\partial h}\left\{u''(w(h) - P)F_2\frac{\partial w(h)}{\partial \bar{\varepsilon}^{NP}(h)} + u'(w(h) - P)F_{22}\right\} = -u'(w(h) - P)F_{12}$$

Note that as $\bar{\varepsilon}$ increases $w(h)$ decreases, since $F(h, \varepsilon - x(\varepsilon, h))$ is decreasing for $\varepsilon < \bar{\varepsilon}$, and constant for $\varepsilon \geq \bar{\varepsilon}$. Thus, we have

$$\frac{\partial\bar{\varepsilon}^{NP}(h)}{\partial h} > 0.$$

\[\blacksquare\]

**Proposition 6**

**Proof:** From (2.21), we immediately obtain

$$-F_2(h, \bar{\varepsilon}^{NP}(h)) = \frac{\sum u'(w(h) - P)\Phi(h)}{u'(w(h) - P)} \begin{cases} < 1 & \bar{\varepsilon}^{NP}(h) < \bar{\varepsilon}^{SP}(h) \\ = 1 & \Rightarrow \bar{\varepsilon}^{NP}(h) = \bar{\varepsilon}^{SP}(h) \\ > 1 & \bar{\varepsilon}^{NP}(h) > \bar{\varepsilon}^{SP}(h) \end{cases}$$

as $-F_2(h, \bar{\varepsilon}^{SP}(h)) = 1$.

Let us take $h_L < \bar{h} < h_H$, and suppose

$$-F_2(h_L, \bar{\varepsilon}^{NP}(h_L)) > 1 > -F_2(h_H, \bar{\varepsilon}^{NP}(h_H)),$$  \quad (B.9)
i.e.

\[ \varepsilon^{NP}(h_H) < \varepsilon^{SP}(h_H) \Rightarrow w^{NP}(h_H) > w^{SP}(h_H) \]
\[ \varepsilon^{NP}(h_L) > \varepsilon^{SP}(h_L) \Rightarrow w^{NP}(h_L) < w^{SP}(h_L), \]

where \( w^{SP}(h) = g(h)F(h, 0) + (1 - g(h)) \int f(\varepsilon)F(h, \varepsilon - x(\varepsilon, h))d\varepsilon. \) Then, we have

\[ u'^{NP}(c(h_H) - P) < u'^{SP}(c(h_H) - P) < u'^{SP}(c(h_L) - P) < u'^{NP}(c(h_L) - P), \]

where the second inequality follows from (B.12). This result, in combination with (B.9) implies

\[ u'^{NP}(c(h_L) - P)[-F_2(h_L, \varepsilon^{NP}(h_L))] > u'^{NP}(c(h_H) - P)[-F_2(h_H, \varepsilon^{NP}(h_H))], \]

a contradiction to (2.21).

**Proposition 9**

**Proof:** Is by backward induction. Trivially \( e_T(h) = 0 \). In period \( T \), since both policies are in place, the wage and health insurance premium of every household is independent of \( h \). Thus

\[ v_T(h) = u(w_T - P_T) = v_T \]

and therefore the terminal value function is independent of \( h \). Now suppose for a given time period \( t \) the value function \( v_{t+1} \) is independent of \( h \). Then from the first order condition with respect to \( e_t(h) \) we have

\[ q'(e_t(h)) = \beta v_{t+1} \sum_{h'} \frac{\partial Q(h'; h, e)}{\partial e} \]

But since for every \( e \) and every \( h \), \( Q(h'; h, e) \) is a probability measure over \( h' \) we have

\( \sum_{h'} \frac{\partial Q(h'; h, e)}{\partial e} = 0 \) and thus \( e_t(h, \gamma) = 0 \) for all \( h \), on account of our assumptions on \( q'(\cdot) \).
But then

$$v_t(h) = u(w_t - P_t) + \left\{-0 + \beta v_{t+1} \sum_{h'} Q(h'; h, 0)\right\} = u(w_t - P_t) + \beta v_{t+1} = v_t$$

since $\sum_{h'} Q(h'; h, 0) = 1$ for all $h$. Thus $v_t$ is independent of $h$. The evolution of the health distributions follows from (2.23), and given these health distributions wages and health insurance premia are given by (2.39) and (2.41).

**B.2 Further Analysis of the No-Wage Discrimination Case**

**B.2.1 Health Insurance Distortions with No-Wage Discrimination**

The firm’s break-even condition is

$$\sum_h \left\{ g(h)F(h, 0) + (1 - g(h)) \int_0^\varepsilon f(\varepsilon)[F(h, \varepsilon - x^{NP}(\varepsilon, h))]d\varepsilon - w(h) \right\} \Phi(h) = 0,$$

and hence on average the production level of a worker will equal his gross wage. Taking $\varepsilon_w > 0$ and $\delta > 0$ as given, workers for whom the wage limits, $\max_{h,h'} |w(h) - w(h')| \leq \varepsilon_w$, bind will be paid either more or less than their production level depending on whether the wage discrimination bound binds from above or below. The firm will optimally choose to hire less than the population share of any health type $h$ whose wage is above their production level, and hence some of these workers will be unemployed. Since we have assume that there is no cost to working and workers pay for their own insurance, competition over health insurance will lead these workers to increase their health insurance, $x(e, h)$, so that their productivity is within $\varepsilon_w$ of their wage $w(h)$. In the limit as $\varepsilon_w \to 0$, this implies that

$$w(h) = g(h)F(h, 0) + (1 - g(h)) \int_0^\varepsilon f(\varepsilon)[F(h, \varepsilon - x^{NP}(\varepsilon, h))]d\varepsilon, \quad (B.10)$$

holds and they are fully employed, or $w(h) - P(h) = 0$. On the flip side, there will be excess demand for workers whose expected production is more than $w(h)$, they will therefore find
it optimal to either lower their insurance, and in the limit as \( \varepsilon \to 0 \) either (B.10) holds they or set \( x(e, h) = 0 \) if they end up at corner with respect to health insurance. Assuming that neither corner binds, this implies that the no-wage discrimination policy will be undone by adjustments in the health insurance market. This motivated our assumption that the government will choose to regulate the health insurance market to prevent this outcome as part of the no-wage discrimination policy.

For health types for which the bounds do not bind, market clearing implies that

\[
    w(h) = g(h)F(h, 0) + (1 - g(h)) \int_0^\varepsilon f(\varepsilon)[F(h, \varepsilon - x^{NP}(\varepsilon, h))]d\varepsilon
\]

while actuarial fairness implies that

\[
    P(h) = (1 - g(h)) \int_0^\varepsilon f(\varepsilon)x^{NP}(\varepsilon, h))d\varepsilon.
\]

Hence, an efficient health insurance contract for this type will maximize

\[
    w(h) - P(h) = w^{CE}(h) - P^{CE}(h).
\]

Since \( w^{CE}(h) - P^{CE}(h) \) is increasing in \( h \), it follows that the wage bound binds for the lowest and highest health types.

### B.2.2 No-Wage Discrimination with Realized Penalties in Equilibrium

Here we assume that the firm must pay a cost for having wage dispersion conditional on health type or for having the health composition of its work force differ from the population average. The wage variation penalty is assumed to take the form

\[
    C \sum_h [w(h) - w(0)]^2 n(h),
\]

since health type 0 will have the lowest wage in equilibrium, and where \( C \) is the penalty parameter and \( n(h) \) is measure of type \( h \) workers the firm hires. Note that with this penalty
function the penalty will apply to all workers with health \( h > 0 \). The penalty from having one’s composition deviate from the population average is given by

$$\sum_h D \left[ \frac{n(h)}{\sum n(h)} - \frac{\Phi(h)}{\sum \Phi(h)} \right]^2.$$ 

Since these penalties are small for small deviations, it will turn out that penalty costs will be realized in equilibrium. Since both of these penalties are real we need to subtract them from production. We will assume that there too the government will regulate the insurance market to prevent workers low health status workers raising their productivity by over-insuring themselves against health risks and high health status workers lowering their productivity by under-insuring themselves.

We begin analyzing this case by assuming that the penalties for wage discrimination \( C \) and hiring discrimination \( D \) are both finite and then we examine the equilibrium in the limit as they become large. The firm takes as given the health policy of the worker and the equilibrium wage \( w(h) \) and chooses the measure of each health type to hire \( n(h) \) so as to maximize

$$\max_{n(h)} \left[ \sum_h \left[ g(h) \left[ F(h, -x(0, h)) - x(0, h) \right] 
+ (1 - g(h)) \int_0^\varepsilon f(\varepsilon) \left[ F(h, \varepsilon - x(\varepsilon, h)) - x(\varepsilon, h) \right] d\varepsilon - w(h) \right] n(h) 
- C \sum_h [w(h) - w^*]^2 n(h) - \sum_h \left[ \frac{n(h)}{\sum n(h)} - \frac{\Phi(h)}{\sum \Phi(h)} \right]^2 \right] n(h),$$

where \( w^* \) is taken here to mean the lowest wage. Trivially, the firm will want to hire more

\(^{63}\)If we have assumed that the form of the penalty was

$$C \int_h [w(h) - w^*]^2 \psi(h) dh,$$

where \( w^* \) is the average wage, this would mean that low productivity workers are more costly and less productive, which will discourage hiring them. Hence, with this form the low productivity workers will only be employed because of the compositional penalty, which means that the hiring penalty must bind at the margin. Hence the less than average productivity workers will be in positive net supply in equilibrium, which will complicate the analysis because some of these workers will be employed and some will not be.
than the population share of any type $h$ for whom

$$N(h) \equiv \left[ g(h) [F(h, -x(0, h)) - x(0, h)] \\
+ (1 - g(h)) \int_{0}^{\varepsilon} f(\varepsilon) [F(h, \varepsilon - x(\varepsilon, h)) - x(\varepsilon, h)] d\varepsilon - w(h) \\
- C [w(h) - w^*]^2 \right]$$

is positive and less than the population share if $N(h)$ is negative. Since all firms share this condition, they will all choose the same relative shares of each type of worker. Since workers are willing to work so long as $w(h) - P(h) > 0$, it follows that $w(h)$ cannot be more than $w^*$ if $N(h)$ is not positive. To see this note that there would be excess supply of type $h$ workers and hence the labor market would not clear. Moreover, a firm would rather hire a worker of type $h$ at $w^* - \varepsilon$ than for $w^*$ for $\varepsilon$ small. Hence, if $w(h) = w^*$, then $N(h) = 0$ so long as $w^* - P(h) > 0$. Hence, for the labor market to clear for each health type, either $N(h) = 0$ for type $h$ or $N(h) > 0$ but $w(h) - P(h) = 0$. This implies the following proposition.

**Proposition 10.** If $C$ and $D$ are positive but finite, and $w(h) - P(h) > 0$ for all $h$, then in equilibrium all households are hired, all firms are representative, and the wage $w(h)$ is equal to a worker’s productivity less the cost of paying him.

Since the government can set $x(\varepsilon, h) = 0$ which implies that $P(h) = 0$, we assume that $w(h) - P(h) > 0$ for all health types.

**B.2.3 Realized Penalties with Both Policies**

Since all that workers care about is their net wage $\tilde{w}(h)$, which is also equal to their consumption, it follows that workers are indifferent over contracts that offer combinations of a gross wage $w(h)$ and medical costs $P(h)$ for which $\tilde{w}(h) = w(h) - P(h)$ is constant. Hence, it is natural to assume that the firm takes the equilibrium net wage function $\tilde{w}(h)$ as given and chooses the measure of each health type to hire, $n(h)$, and its health plan, $x(\varepsilon, h)$, to solve the following problem
\[
\max_{n(h), x(\varepsilon, h)} \sum_h \left[ g(h) \left[ F(h, -x(0, h)) - x(0, h) \right]
\right.
\]
\[
+ (1 - g(h)) \int_0^\varepsilon f(\varepsilon) \left[ F(h, \varepsilon - x(\varepsilon, h)) - x(\varepsilon, h) \right] d\varepsilon - \tilde{w}(h) \n\]
\[
- C \sum_h |\tilde{w}(h) - \tilde{w}(0)|^2 n(h) - \sum_h D \left[ \frac{n(h)}{\sum n(h)} - \frac{\Phi(h)}{\sum \Phi(h)} \right]^2.
\]

**Proposition 11.** If \( C \) and \( D \) are positive but finite, then in equilibrium all households are hired, all firms are representative, the net wage \( \tilde{w}(h) \) is equal to a worker’s productivity less the cost of paying him more than \( \tilde{w}(0) \), and \( \tilde{w}(0) = w^{CE}(0) - P(0) \). The firm optimally sets \( x(\varepsilon, h) = x^{CE}(\varepsilon, h) \). As \( C \to \infty \), \( \tilde{w}(h) \to \tilde{w}(0) \).

**Proof:** The optimality condition for \( x(h, \varepsilon) \) if \( \varepsilon = 0 \) is
\[
F(h, -x(0, h)) - 1 \leq 0
\]
and if \( \varepsilon > 0 \) is
\[
F(h, \varepsilon - x(\varepsilon, h)) - 1 \leq 0 \text{ w. equality if } x(\varepsilon, h) > 0.
\]

These are the same conditions as in the competitive equilibrium.

Next, we show that \( \tilde{w}(h) \) has to be increasing in \( h \) and hence \( \tilde{w}(0) \) is the lowest paid type. The wage penalty is w.r.t. to the lowest paid worker type, which we denote by \( w^* \).

Given that optimum insurance is the same as in the competitive equilibrium, it follows that the net earnings per worker is \( w^{CE}(h) - P^{CE}(h) - \tilde{w}(h) \), and from before \( w^{CE}(h) - P^{CE}(h) \) is increasing in \( h \). Hence, for the firm to break even
\[
\sum_h \left[ w^{CE}(h) - P^{CE}(h) - \tilde{w}(h) \right] n(h)
\]
\[
- C \sum_h (\tilde{w}(h) - w^*)^2 n(h) - \sum_h D \left[ \frac{n(h)}{\sum n(h)} - \frac{\Phi(h)}{\sum \Phi(h)} \right]^2 = 0,
\]
and the optimality condition for $n(h)$ is

$$
\frac{\left[w^{CE}(h) - P^{CE}(h) - \tilde{w}(h)\right] - C [\tilde{w}(h) - w^*]^2}{D \left[\frac{n(h)}{\sum n(h)} \frac{\Phi(h)}{\sum \Phi(h)} \right] \left[1 - \frac{n(h)}{\sum n(h)}\right] \frac{1}{\sum n(h)}} = 0.
$$

This condition implies that a firm will hire more that the population share of any type $h$ for whom

$$
\tilde{N}(h) \equiv w^{CE}(h) - P^{CE}(h) - \tilde{w}(h) - C [\tilde{w}(h) - w^*]^2 > 0,
$$

and less than the population share if the reverse is true. However any health type $h$ that are not fully employed in equilibrium would have excess members who would be happy to be hired any positive wage. Hence, either type $h$ is paid the lowest equilibrium wage or they are fully employed. Hence, any type $h$ for whom $w(h) > w^*$ are fully employed. Any type receiving the lowest wage must be fully employed since the firm would be willing to hire more of these workers if we lowered the bottom wage by $\varepsilon$. Since all workers are fully employed, it follows that all firms will choose to be representative to avoid the hiring penalty, and that $\tilde{w}(0) = w^{CE}(0) = w^*$ and $\tilde{w}(h)$ is increasing $h$. Finally, since the marginal penalty for a deviation in a type’s net wage from the economy-wide lowest type’s wage is given by

$$
-C [\tilde{w}(h) - \tilde{w}(0)]^2,
$$

and since this cost goes to infinity as $C \to \infty$ for any positive wage gap, it follows that as $C$ becomes large $\tilde{w}(h) \rightarrow \tilde{w}(0)$, and all of the workers are paid as if they were the lowest health status type and all of their productivity gap is absorbed by the cost of discriminating on wages. Q.E.D.

The fact that the productivity advantage of higher health status individuals is completely absorbed by the discrimination costs means that the society as a whole gets no gain from their productivity advantage. So the health expenditures that raise their productivity above the lowest type are inefficient. In addition, expenditure on the lowest health type relaxes the wage discrimination penalty on other types. So this equilibrium outcome is not socially
efficient.

B.3 Wages in the Competitive Equilibrium

To understand the implications of proposition 2 for the behavior of equilibrium wages, note that our results imply that the equilibrium competitive wage is given by

$$w_{CE}(h) = g(h)F(h, 0) + (1 - g(h)) \int_{0}^{\bar{\varepsilon}_{CE}(h)} f(\varepsilon)F(h, \varepsilon - x(\varepsilon, h))d\varepsilon$$

$$+(1 - g(h)) \int_{\bar{\varepsilon}_{CE}(h)}^{\bar{\varepsilon}} f(\varepsilon)F(h, \bar{\varepsilon}_{CE}(h))d\varepsilon.$$ 

Hence

$$\frac{dw_{CE}(h)}{dh} = g'(h) \left[ F(h, 0) - \int_{0}^{\bar{\varepsilon}_{CE}(h)} f(\varepsilon)F(h, \varepsilon - x(\varepsilon, h))d\varepsilon \right.$$

$$+g(h)F_1(h, 0) + (1 - g(h)) \int_{0}^{\bar{\varepsilon}_{CE}(h)} f(\varepsilon)F_1(h, \varepsilon - x(\varepsilon, h))d\varepsilon$$

$$+(1 - g(h)) \int_{\bar{\varepsilon}_{CE}(h)}^{\bar{\varepsilon}} f(\varepsilon)F_1(h, \bar{\varepsilon}_{CE}(h))d\varepsilon$$

$$\left. + (1 - g(h)) \int_{\bar{\varepsilon}_{CE}(h)}^{\bar{\varepsilon}} f(\varepsilon)F_2(h, \bar{\varepsilon}_{CE}(h))\frac{d\bar{\varepsilon}_{CE}(h)}{dh}d\varepsilon, \right\]$$

since net effect of the change in the integrand bounds generated by $\frac{d\bar{\varepsilon}_{CE}(h)}{dh}$ is zero. Next note that our optimality condition for $\bar{\varepsilon}_{CE}(h)$, (2.17), implies that

$$F_{12}(h, \bar{\varepsilon}_{CE}(h))dh + F_{22}(h, \bar{\varepsilon}_{CE}(h))d\bar{\varepsilon}_{CE}(h) = 0,$$

and hence

$$\frac{d\bar{\varepsilon}_{CE}(h)}{dh} = -\frac{F_{12}(h, \bar{\varepsilon}_{CE}(h))}{F_{22}(h, \bar{\varepsilon}_{CE}(h))}.$$
This result, along with (2.17), implies that

\[
\frac{d w^{CE}(h)}{dh} = g'(h) \left[ F(h,0) - \int_0^{\epsilon^{CE}(h)} f(\varepsilon) F(h,\varepsilon - x(\varepsilon,h)) d\varepsilon - \int_{\epsilon^{CE}(h)}^{\epsilon} f(\varepsilon) F(h,\epsilon^{CE}(h)) d\varepsilon \right]
\]

\[+ g(h) F_1(h,0) + (1 - g(h)) \int_0^{\epsilon^{CE}(h)} f(\varepsilon) F_1(h,\varepsilon - x(\varepsilon,h)) d\varepsilon \]

\[+ (1 - g(h)) \int_{\epsilon^{CE}(h)}^{\epsilon} f(\varepsilon) F_1(h,\epsilon^{CE}(h)) d\varepsilon \]

\[-(1 - g(h)) \int_{\epsilon^{CE}(h)}^{\epsilon} f(\varepsilon) F_2(h,\epsilon^{CE}(h)) \frac{F_{12}(h,\epsilon^{CE}(h))}{F_{22}(h,\epsilon^{CE}(h))} d\varepsilon.\]

All of the terms in (B.11) are trivially positive except the last, which is negative since \(F_{22} < 0\). However, so long as the spillover ratio \(F_{12}/F_{22}\) evaluated at \((h, \epsilon^{CE}(h))\) is not too negative then, then wages will vary positive with health status. Note that this is trivially implied if the direct effect of the change in health status offsets the spillover, or

\[F_1(h, \epsilon^{CE}(h)) - F_2(h, \epsilon^{CE}(h)) \frac{F_{12}(h, \epsilon^{CE}(h))}{F_{22}(h, \epsilon^{CE}(h))} > 0.\]

Note that this is a condition purely on the fundamentals of the economy since \(\epsilon^{CE}(h)\) is given by an (implicit) equation that depends only on exogenous model elements. We summarize our results in the following proposition:

**Proposition 12.** The competitive wage is increasing in \(h\) if (B.11) is positive.

### B.4 Computation of the Social Planner Problem

The idea to solve the problems in (2.25) is to iterate on sequences \(\{c_t, e_t(h), \Phi_t(h)\}\), using the first order condition (2.26) for the optimal effort choice and the envelope condition.
To initialize the iterations, note that

\[
V_T(\Phi_T) = u(c_T)
\]

\[
\frac{\partial V_T(\Phi_T)}{\partial \Phi_T(h)} = u'(c_T) \cdot \left[ g(h) F(h,0) + (1 - g(h)) \int_0^1 f(\varepsilon) \left[ F(h,\varepsilon - x^{SP}(\varepsilon,h)) - x^{SP}(\varepsilon,h) \right] d\varepsilon \right]
\]

\[
\equiv u'(c_T) \cdot \Psi(h) \quad (B.13)
\]

For these expressions we only need to know \(c_T\), the term \(\Psi(h)\) is just a number that depends on \(h\) and is known once we have solved the static insurance problem. This suggests the following algorithm to solve the dynamic social planner problem:

**Algorithm 13.**

1. Guess a sequence \(\{c_t\}_{t=0}^T\)

2. Determine \(\frac{\partial V_T(\Phi_T)}{\partial \Phi_T(h)}\) from (B.13)

3. Iterate on \(t\) to determine \(\{e_t(h)\}_{t=0}^{T-1}\)
   
   (a) For given \(\frac{\partial V_{t+1}(\Phi_{t+1})}{\partial \Phi_{t+1}(h)}\) use (2.26) to determine \(e_t(h)\).
   
   (b) Use \(c_t, e_t(h), \frac{\partial V_{t+1}(\Phi_{t+1})}{\partial \Phi_{t+1}(h)}\) and (2.27) to determine \(\frac{\partial V_t(\Phi_t)}{\partial \Phi_t(h)}\)

4. Use the initial distribution \(\Phi_0\) and \(\{e_t(h)\}_{t=0}^{T-1}\) to determine \(\{\Phi_t\}_{t=0}^T\) and thus \(\{c_t^{\text{new}}\}_{t=0}^T\).

5. If \(\{c_t^{\text{new}}\}_{t=0}^T = \{c_t\}_{t=0}^T\) we are done. If not, set \(\{c_t\}_{t=0}^T = \{c_t^{\text{new}}\}_{t=0}^T\) and go to 1.

This algorithm is straightforward to implement numerically, since we only have to iterate on the aggregate consumption sequence, not on the sequence of distributions. In particular, the only moderately costly operation comes in step 2a) but even there we only have to solve one nonlinear equation in one unknown (although we have to do it \(T \ast \text{card}(H)\) times per iteration).
B.5 Computation of the Equilibrium with a No-Prior-Conditions Law and/or a No-Wage Discrimination Law

The algorithm to solve this version of the model shares its basic features with that for the social planner problem, but differs in terms of the sequence of variables on which we iterate:

**Algorithm 14.**

1. Guess a sequence \(\{E_{u_t}, P_t\}_{t=0}^T\).

2. Given the guess use equations (2.30)-(2.33) to determine health cutoffs and wages \(\{\bar{\varepsilon}_t^{NP}(h), w_t(h)\}\).

3. Given \(\{w_t(h), P_t\}\), solve the household dynamic programming problem (2.34) for a sequence of optimal effort policies \(\{e_t(h)\}_{t=0}^T\).

4. From the initial health distribution \(\Phi_0\) use the effort functions \(\{e_t(h)\}_{t=0}^T\) to derive the sequence of health distributions \(\{\Phi_t\}_{t=0}^T\) from equation (2.23).

5. Obtain a new sequence \(\{E_{u_t}^{new}, P_t^{new}\}_{t=0}^T\) from (2.32) and (2.33).

6. If \(\{E_{u_t}^{new}, P_t^{new}\}_{t=0}^T = \{E_{u_t}, P_t\}_{t=0}^T\) we are done. If not, go to step 1. with new guess \(\{E_{u_t}^{new}, P_t^{new}\}_{t=0}^T\).

The algorithm for no-wage discrimination is a slight modification of that for no-prior conditions. The algorithm iterates over \(\{E_{u_t}, w_t\}_{t=0}^T\). In Step 1 given the guess use equations (2.36)-(2.40) to determine health cutoffs and premia \(\{\bar{\varepsilon}_t^{NP}(h), P_t(h)\}\). In Step 4 obtain a new sequence \(\{E_{u_t}^{new}, w_t^{new}\}_{t=0}^T\) from (2.39) and (2.38). With both policies, equation (2.41) replaces (2.40) in all expressions.

---

64Instead of \(\{E_{u_t}\}\) one could iterate on \(\{w_t(h)\}\) which is more transparent, but significantly increases the dimensionality of the problem.
B.6 Details for Data and Calibration

B.6.1 Details of the Augmented Model Analysis: Inclusion of the $z$-shock

We assume that households \textit{must} incur the cost $z$, when the $z$-shock hits. This assumption and the fact that households are risk averse imply that the $z$-shock will be fully insured in the competitive equilibrium under any policy (and of course by the social planner).

Moreover, we assume that households receiving a $z$-shock can still work, but that their productivity is only $\rho$ times that of a healthy worker. Therefore, in a competitive equilibrium, the wage of a worker with health status $h$ is given by

$$w(h) = g(h)F(h,0) + \rho \kappa(h) F(h,0) + (1 - g(h) - \kappa(h)) \int F(h, \varepsilon - x(\varepsilon, h)) f(\varepsilon) d\varepsilon$$

and the health insurance premium is determined as

$$P(h) = (1 - g(h) - \kappa(h)) \int x(\varepsilon, h) f(\varepsilon) d\varepsilon + \mu_z(h)$$

Given our assumptions there is no interaction between the $z$-shocks and the health insurance contract problem associated with the $\varepsilon$-shock since it is prohibitively costly by assumption not to bear the $z$-expenditures. The role of the $z$-expenditures is to soak up the most extreme health expenditures observed in the data associated with catastrophic illnesses, but to otherwise leave our theory from the previous sections unaffected.

The static analysis goes through completely unchanged in the presence of the $z$-shocks. In the dynamic analysis the benefits of higher effort $e$ and thus a better health distribution $\Phi_t(h)$ now also include a lower probability $\kappa(h)$ of receiving a positive $z$-shock and a lower mean expenditure $\mu_z(h)$ from that shock with better health $h$. This extension of the model leads to straightforward extensions of the expressions derived in the analysis of the dynamic model in section 2.4, and does not change any of the theoretical properties derived in sections 2.3 and 2.4.
### B.6.2 Descriptive Statistics of the PSID Data

Before we proceed to descriptive statistics of the PSID data, we summarize, in Table B.1, the mapping between variables in our model and data.

<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
<th>Data PSID Variable</th>
<th>Actual Data Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x, \mu_z$</td>
<td>Medical Expenditure</td>
<td>Average of total expenditure reported in 1999, 2001, 2003</td>
<td>1997-2002</td>
</tr>
<tr>
<td>$h$</td>
<td>Health Status</td>
<td>Self-reported Health in 1997</td>
<td>1997</td>
</tr>
</tbody>
</table>

Since our model period is six years, we take average of reported medical expenditure and wages over six year periods that we observe. Moreover, we use health status data from 1997 (rather than 1999) to capture the effect of health on wages and medical expenditure.

Table B.2 documents descriptive statistics of key variables from the 1999 PSID data that we use in our analysis.

<table>
<thead>
<tr>
<th>Model</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>41</td>
<td>10</td>
<td>23</td>
<td>65</td>
</tr>
<tr>
<td>Labor Income</td>
<td>30,170</td>
<td>40,573</td>
<td>0</td>
<td>1,153,588</td>
</tr>
<tr>
<td>if Labor Income &gt; 0</td>
<td>32,076</td>
<td>41,097</td>
<td>0.55</td>
<td>1,153,588</td>
</tr>
<tr>
<td>Excellent</td>
<td>38,755</td>
<td>55,406</td>
<td>0</td>
<td>940,804</td>
</tr>
<tr>
<td>Very Good</td>
<td>32,768</td>
<td>40,351</td>
<td>0</td>
<td>1,153,588</td>
</tr>
<tr>
<td>Good</td>
<td>25,516</td>
<td>25,908</td>
<td>0</td>
<td>384,783</td>
</tr>
<tr>
<td>Fair</td>
<td>12,605</td>
<td>13,926</td>
<td>0</td>
<td>81,300</td>
</tr>
<tr>
<td>Medical Expenditure</td>
<td>1,513</td>
<td>4,624</td>
<td>0</td>
<td>127,815</td>
</tr>
<tr>
<td>Excellent</td>
<td>1,234</td>
<td>2,374</td>
<td>0</td>
<td>28,983</td>
</tr>
<tr>
<td>Very Good</td>
<td>1,647</td>
<td>5,812</td>
<td>0</td>
<td>127,815</td>
</tr>
<tr>
<td>Good</td>
<td>1,486</td>
<td>4,283</td>
<td>0</td>
<td>93,298</td>
</tr>
<tr>
<td>Fair</td>
<td>1,792</td>
<td>4,950</td>
<td>0</td>
<td>65,665</td>
</tr>
<tr>
<td>Health Status</td>
<td>2.77</td>
<td>0.95</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>

Physical Activity: fraction(number) of days in a year

- **Light**
  - Mean: 0.63 (230.99)
  - Std. Dev.: 0.39 (142.28)
  - Min: 0
  - Max: 1 (365)

- **Heavy**
  - Mean: 0.29 (105.69)
  - Std. Dev.: 0.35 (126.85)
  - Min: 0
  - Max: 1 (365)

In the PSID, each individual (head of household) self-reports his health status in a 1 to 5 scale, where 1 is Excellent, 2, Very Good, 3, Good, 4, Fair, and 5 is Poor. Even with large number of observations, only about 1% of total individuals report their health status
to be poor. Thus, for our analysis, we will use four levels of health status (merge poor and fair together). Since PSID reports household medical expenditure, we control for family size using modified OECD equivalence scale.

As we model working-age population, each household starts his life as a 24 year old and makes economic decisions until he is 65 years old. Our model time period is 6 years and thus they live for 7 time periods. We choose six year time period to capture the effect of exercises on health transition. Since exercises tend to have positive longer-term effects than do medical expenditure, by allowing for a medium-term time period, we are able to quantify the impact of exercises in a more reliable way.

**Data on Health Transitions** Table B.3 presents the transition matrix of health status over six years. We see that health status is quite persistent.

<table>
<thead>
<tr>
<th></th>
<th>Excellent</th>
<th>Very Good</th>
<th>Good</th>
<th>Fair</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excellent</td>
<td>1,286</td>
<td>904</td>
<td>335</td>
<td>92</td>
<td>2,617</td>
</tr>
<tr>
<td></td>
<td>49.14 %</td>
<td>34.54 %</td>
<td>12.80 %</td>
<td>3.52 %</td>
<td>100 %</td>
</tr>
<tr>
<td>Very Good</td>
<td>482</td>
<td>1,844</td>
<td>1,217</td>
<td>274</td>
<td>3,817</td>
</tr>
<tr>
<td></td>
<td>12.63 %</td>
<td>48.31 %</td>
<td>31.88 %</td>
<td>7.18 %</td>
<td>100 %</td>
</tr>
<tr>
<td>Good</td>
<td>187</td>
<td>712</td>
<td>1,592</td>
<td>637</td>
<td>3,128</td>
</tr>
<tr>
<td></td>
<td>5.98 %</td>
<td>22.76 %</td>
<td>50.90 %</td>
<td>20.36 %</td>
<td>100 %</td>
</tr>
<tr>
<td>Fair</td>
<td>36</td>
<td>109</td>
<td>358</td>
<td>957</td>
<td>1,460</td>
</tr>
<tr>
<td></td>
<td>2.47 %</td>
<td>7.47 %</td>
<td>24.52 %</td>
<td>65.55 %</td>
<td>100 %</td>
</tr>
<tr>
<td>Total</td>
<td>1,991</td>
<td>3,569</td>
<td>3,502</td>
<td>1,960</td>
<td>11,022</td>
</tr>
<tr>
<td></td>
<td>18.06 %</td>
<td>32.38 %</td>
<td>31.77 %</td>
<td>17.78 %</td>
<td>100 %</td>
</tr>
</tbody>
</table>

**Physical Activity Data** Here, we report some statistics on physical activity.

- *Variation of Physical Activity and Its Impact on Health Transition*

Density of light and heavy physical activity levels by health are summarized in Figures B.1 and B.2. From variations in health evolution by physical activity and initial health status, we find that about 30% of variance in health status in the future is explained by

---

65 Labor income and medical expenditure data for fair health in Table B.2 include poor (5) in data.
66 Each additional adult gets the weight of 0.5, and each child, 0.3.
health status today, whereas, light and physical activity explains about 8% and 14%, respectively. Moreover, both initial health status and light (heavy) exercise explains 46% (41%) of variance in future health outcome.\textsuperscript{67}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{density.png}
\caption{Density of Light Activity}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{density.png}
\caption{Density of Heavy Activity}
\end{figure}

- \textit{Physical Activity Over Time}

Light physical activity has steadily decreased over time, whereas heavy physical activity decreased for a while, but started increasing in 2005 (Figures B.3 and B.4).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{activity.png}
\caption{Light Activity}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{activity.png}
\caption{Heavy Activity}
\end{figure}

\textsuperscript{67}From the law of total variance, we know

\[ \text{var}(Y) = E(\text{var}(Y|X)) + \text{var}(E(Y|X)), \]

where the former is the \textit{unexplained} and the latter, \textit{explained} component of the variance.
B.6.3 Health Shocks, Distribution of Medical Expenditures, and Discussion of Categorization of Health Shocks

Before going into discussing the medical expenditure distribution in data, we briefly discuss the appropriate counterparts of data moments for our model. In our model, households do not consume medical care when they do not get a health shock (although, they can choose not to spend any in case of health shock, since \( x^*(h, \varepsilon) = \max\{0, \varepsilon(h)\} \)). Therefore, in data, we are interested in the distribution of medical expenditure conditional on having gotten any health shocks (which we have some information in PSID).

Table B.4 summarizes medical expenditure by shock. Note that all numbers reported are yearly average taken over six years (1997-2002).

<table>
<thead>
<tr>
<th>Category</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>4,226</td>
<td>1,513</td>
<td>4,624</td>
<td>0</td>
<td>127,815</td>
</tr>
<tr>
<td>No Shock</td>
<td>1,419</td>
<td>1,350</td>
<td>4,447</td>
<td>0</td>
<td>101,952</td>
</tr>
<tr>
<td>Any Shock</td>
<td>2,807</td>
<td>1,595</td>
<td>4,710</td>
<td>0</td>
<td>127,815</td>
</tr>
<tr>
<td>Catastrophic Disease Shock</td>
<td>168</td>
<td>3,745</td>
<td>9,363</td>
<td>0</td>
<td>93,298</td>
</tr>
<tr>
<td>Cancer</td>
<td>51</td>
<td>5,210</td>
<td>15,134</td>
<td>0</td>
<td>93,298</td>
</tr>
<tr>
<td>Heart Attack</td>
<td>46</td>
<td>3,334</td>
<td>4,705</td>
<td>0</td>
<td>27,161</td>
</tr>
<tr>
<td>Heart Disease</td>
<td>94</td>
<td>3,382</td>
<td>5,535</td>
<td>0</td>
<td>38,500</td>
</tr>
<tr>
<td>Light Shock</td>
<td>2,767</td>
<td>1,585</td>
<td>4,732</td>
<td>0</td>
<td>127,815</td>
</tr>
<tr>
<td>Diabetes</td>
<td>183</td>
<td>2,088</td>
<td>7,196</td>
<td>0</td>
<td>93,298</td>
</tr>
<tr>
<td>Stroke</td>
<td>33</td>
<td>2,200</td>
<td>4,905</td>
<td>0</td>
<td>27,161</td>
</tr>
<tr>
<td>Arthritis</td>
<td>322</td>
<td>1,684</td>
<td>3,166</td>
<td>0</td>
<td>38,500</td>
</tr>
<tr>
<td>Hypertension</td>
<td>566</td>
<td>1,825</td>
<td>6,143</td>
<td>0</td>
<td>93,298</td>
</tr>
<tr>
<td>Lung Disease</td>
<td>63</td>
<td>1,705</td>
<td>2,476</td>
<td>0</td>
<td>12,595</td>
</tr>
<tr>
<td>Asthma</td>
<td>61</td>
<td>1,135</td>
<td>1,444</td>
<td>0</td>
<td>7,170</td>
</tr>
<tr>
<td>Ill</td>
<td>2,351</td>
<td>1,637</td>
<td>5,040</td>
<td>0</td>
<td>127,815</td>
</tr>
<tr>
<td>( z )-shock</td>
<td>297</td>
<td>4,704</td>
<td>12,834</td>
<td>0</td>
<td>127,815</td>
</tr>
<tr>
<td>( \varepsilon )-shock</td>
<td>2,510</td>
<td>1,227</td>
<td>2,023</td>
<td>0</td>
<td>32,909</td>
</tr>
</tbody>
</table>

We see that cancer, heart attack, and heart disease incur the most medical expenditure, and thus we categorize them to be catastrophic shocks (\( z \)-shocks). Although the diseases PSID specifically reports information on are those that are common, they are not, by all means, exhaustive of the kind of health diseases that one can be diagnosed with. And this is hinted when we look at the medical expenditure statistics for those who report to have missed work due to illness. The maximum amount of medical expenditure they spend exceeds those of the others, and this might be due to some severe diseases for which they
had to be treated.

Therefore, in addition to cancer, heart attack, and heart disease, we categorize those who have spent more than their labor income on medical expenditure as having had a catastrophic \((z)\) health shock.\(^{68}\) Those who had a health shock that were not cancer, heart attack, or heart disease, and who spent less than their income on medical expenditure is considered to have had an \(\varepsilon\)-shock.\(^{69}\)

Figures B.5 - B.8, plot logs of medical expenditure distribution for all population, for those with ANY health shock, those with \(z\)-shock, and those with \(\varepsilon\)-shock. By definition, mean medical expenditure of \(z\)-shock households are higher than those of \(\varepsilon\)-shock, and so are standard deviations.

\(^{68}\)Categorizing catastrophic health shocks using expenditures as percentage of income is not new. There has been discussion on insuring catastrophic health shocks, and they mostly refer to high amount of expenditure as percentage of income.

\(^{69}\)In PSID sample, median of percentage of labor income spent on medical expenditure is 2\%, and the mean, 132\%. Only about 5\% of households with health shocks spend medical expenditure in excess of their labor income.
B.6.4 Estimation Results

Health Transition  Using the functional form described in the main body of the paper, we estimate the health transition function in the following way.

Let set of parameters to be estimated be \( \theta = \{ G(h, h'), \delta, \phi(h), \lambda(h), \alpha_1(h), \alpha_2(h) \} \). We use Generalized Method of Moments to estimate these parameters.

We first determine the exercise intervals and assign each individual initial health status and exercise level bins, \( k \). Using the transition from the data \( \mathbb{E}(q^k(h')) \), we minimize the distance between our estimated transition function and data, i.e.

\[
\theta = \arg \min_{\theta} \left( \frac{1}{K} \sum_{k=1}^{K} \left[ Q(h'_k; \theta) - \mathbb{E}(q^k(h')) \right] \right) \hat{W} \left( \frac{1}{K} \sum_{k=1}^{K} \left[ Q(h'_k; \theta) - \mathbb{E}(q^k(h')) \right] \right),
\]

where \( K \) denotes the total number of groups and \( \hat{W} \), weighting matrix. Here, we use the efficient weighting matrix.

With exercise step size of nine,\(^70\) we get the following estimated parameter values (\( h = 1, 2, 3, 4 \) corresponds to health being fair, good, very good, and excellent, respectively, i.e. the higher the \( h \) the better one’s health status.).

\[
\hat{G}(h, h') = \begin{bmatrix} 0.8742 & 0.0927 & 0.0230 & 0.0101 \\ 0.6597 & 0.2547 & 0.0609 & 0.0249 \\ 0.1404 & 0.3949 & 0.3442 & 0.1204 \\ 0.0850 & 0.3170 & 0.5406 & 0.0573 \end{bmatrix}
\]

\[
\delta = 1
\]

\[
\phi = [2.2796, 1.1063, 0.5179, 8.4123]
\]

\[
\lambda = [0.3308, 0.0193, 0.5939, 0.1878]
\]

\[
\alpha_1 = [1.3274, 12.8747, 7.0260]
\]

\[
\alpha_2 = [0.8035, 5.8693]
\]

\(^70\)The PSID has exercise data from 1999 to 2009. The total number of observations for 6 year transition is 11,022.
The estimated transition functions are plotted in Figures B.9 - B.12. In the figures, the smoothed functions are estimated transition, whereas the straight lines represent the data. We see that our functional form fits the data quite well.

Table B.5: Health Shock Probabilities by Health Status

<table>
<thead>
<tr>
<th>Health Status</th>
<th>Observations</th>
<th>Any Health Shock $1 - g(h)$</th>
<th>$z$ Shock $\kappa(h)$</th>
<th>$\varepsilon$ Shock $1 - g(h) - \kappa(h)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>4,226</td>
<td>0.66</td>
<td>0.07</td>
<td>0.59</td>
</tr>
<tr>
<td>Fair</td>
<td>458</td>
<td>0.66</td>
<td>0.21</td>
<td>0.45</td>
</tr>
<tr>
<td>Good</td>
<td>1,139</td>
<td>0.71</td>
<td>0.07</td>
<td>0.63</td>
</tr>
<tr>
<td>Very Good</td>
<td>1,618</td>
<td>0.68</td>
<td>0.05</td>
<td>0.62</td>
</tr>
<tr>
<td>Excellent</td>
<td>1,143</td>
<td>0.60</td>
<td>0.03</td>
<td>0.57</td>
</tr>
</tbody>
</table>

Health Shock Probabilities  As seen in Table B.5, there are about 7% of households who receive $z$-shocks over six years, and the probabilities are decreasing in health status. However, probability of getting any health shock is not the highest for the Fair health individuals (from Good to Excellent, it is monotone). This might be due to the fact that given that the health status is already bad, probabilities that one would get other minor adverse health shocks ($\varepsilon$ shocks in the model) are not very high.
**Effect of Health Shock on Productivity**  In Table B.6, we summarize working hours and labor income reported by those with different health shock categories.

The six year average hours worked of those with $z$-shocks are about half that of the ones who did not get any shock (and worked) and they earn about half on average. Therefore, we take $\rho = 0.4235$, which is the percentage of labor income earned by those with $z$-shock, compared to those who have worked and did not experience any health shock (since we denote earnings of those with $z$-shock as $\rho F(h, 0)$).

Table B.6: Hours Worked and Labor Income by Health Shock

<table>
<thead>
<tr>
<th></th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hours Worked</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>4.226</td>
<td>1.823</td>
<td>856</td>
<td>0</td>
<td>5,300</td>
</tr>
<tr>
<td>Positive Hours</td>
<td>3.903</td>
<td>1.974</td>
<td>704</td>
<td>7</td>
<td>5,300</td>
</tr>
<tr>
<td>No Shock, Positive Hours</td>
<td>1.259</td>
<td>1.987</td>
<td>781</td>
<td>14</td>
<td>4,732</td>
</tr>
<tr>
<td>$z$-shock</td>
<td>297</td>
<td>998</td>
<td>1,033</td>
<td>0</td>
<td>3,640</td>
</tr>
<tr>
<td>$\epsilon$-shock</td>
<td>2.639</td>
<td>1,892</td>
<td>763</td>
<td>0</td>
<td>5,300</td>
</tr>
<tr>
<td><strong>Labor Income</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>4.226</td>
<td>30.171</td>
<td>40,573</td>
<td>0</td>
<td>1,153,588</td>
</tr>
<tr>
<td>Positive Hours</td>
<td>3.903</td>
<td>32.362</td>
<td>41,364</td>
<td>0</td>
<td>1,153,588</td>
</tr>
<tr>
<td>No Shock, Positive Hours</td>
<td>1.259</td>
<td>32.606</td>
<td>49,358</td>
<td>0</td>
<td>940,804</td>
</tr>
<tr>
<td>$z$-shock</td>
<td>297</td>
<td>13,809</td>
<td>25,470</td>
<td>0</td>
<td>253,560</td>
</tr>
<tr>
<td>$\epsilon$-shock</td>
<td>2.639</td>
<td>31,163</td>
<td>36,883</td>
<td>0</td>
<td>1,153,588</td>
</tr>
</tbody>
</table>
### B.6.5 Calibration Results

#### Table B.7: Data Targets

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Data Targets</th>
</tr>
</thead>
<tbody>
<tr>
<td>{h_i}_{i=1,2,3,4}</td>
<td>Income of (h_i) relative to (h_1)</td>
</tr>
<tr>
<td></td>
<td>( \log \frac{w(h_2)}{w(h_1)} = 0.2739 )</td>
</tr>
<tr>
<td></td>
<td>( \log \frac{w(h_3)}{w(h_1)} = 0.4691 )</td>
</tr>
<tr>
<td></td>
<td>( \log \frac{w(h_4)}{w(h_1)} = 0.5948 )</td>
</tr>
<tr>
<td></td>
<td>Income of Old relative to Young</td>
</tr>
<tr>
<td></td>
<td>( \log \frac{w(O)}{w(Y)} = 0.1114 )</td>
</tr>
</tbody>
</table>

| Production Function \( A(t, educ) \) | Income in \( t \) of less than HS relative to Income (Young, Fair) |
| | \( t = 1, < HS : -0.0042 \) |
| | \( t = 2, < HS : 0.1449 \) |
| | \( t = 3, < HS : 0.1715 \) |
| | \( t = 4, < HS : 0.1980 \) |
| | \( t = 5, < HS : 0.0907 \) |
| | \( t = 6, < HS : -0.0969 \) |
| | \( t = 7, < HS : -0.1112 \) |
| Income in \( t \) of HS Grad relative to Income (Young, Fair) |
| | \( t = 1, HS : 0.2980 \) |
| | \( t = 2, HS : 0.4738 \) |
| | \( t = 3, HS : 0.5082 \) |
| | \( t = 4, HS : 0.5988 \) |
| | \( t = 5, HS : 0.6060 \) |
| | \( t = 6, HS : 0.5395 \) |
| | \( t = 7, HS : 0.2406 \) |

| \( \phi(a, educ) \) | \% Income spent on Med Exp. by Health |
| | \( \mathbb{E}(x|h_1) = 0.0525 \) |
| | \( \mathbb{E}(x|h_2) = 0.0429 \) |
| | \( \mathbb{E}(x|h_3) = 0.0353 \) |
| | \( \mathbb{E}(x|h_4) = 0.0308 \) |

<p>| ( \xi(a, educ) ) | % Income on Med Exp. by Educ. and Age(a ( \in {Y, O})) |
| | ( \mathbb{E}(x|Y, &lt; HS) = 0.0386 ) |
| | ( \mathbb{E}(x|Y, HS) = 0.0348 ) |
| | ( \mathbb{E}(x|O, &lt; HS) = 0.0428 ) |
| | ( \mathbb{E}(x|O, HS) = 0.0356 ) |</p>
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Data Targets</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\varepsilon$-shock Distribution</td>
<td>$\mu, \sigma$ Mean, St. Dev of Med. Exp. on Light</td>
</tr>
<tr>
<td></td>
<td>$E(x) = 0.0362$</td>
</tr>
<tr>
<td></td>
<td>$E(w) = 0.362$</td>
</tr>
<tr>
<td></td>
<td>$\sigma(x) = 1.6462$</td>
</tr>
<tr>
<td>$z$-shock Distribution</td>
<td>$\mu_z(h)$ % Income on Cata. Shock by Health</td>
</tr>
<tr>
<td></td>
<td>$E(z</td>
</tr>
<tr>
<td></td>
<td>$E(w</td>
</tr>
<tr>
<td></td>
<td>$E(z</td>
</tr>
<tr>
<td></td>
<td>$E(w</td>
</tr>
<tr>
<td>Exercise Disutility</td>
<td>${\gamma_1(educ), \gamma_2(educ)}$ Mean and St. Dev of Exercise in $t = 1$</td>
</tr>
<tr>
<td></td>
<td>$E(e_t = 1) = 0.5735$</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2(e_t = 1) = 0.2828$</td>
</tr>
<tr>
<td></td>
<td>Measure of Fair and Excellent in $t = 7$</td>
</tr>
<tr>
<td></td>
<td>$\Phi_{t-T}(h_1) = 0.1944$</td>
</tr>
<tr>
<td></td>
<td>$\Phi_{t-T}(h_4) = 0.1618$</td>
</tr>
<tr>
<td>Preference Distribution</td>
<td>$p(\gamma</td>
</tr>
<tr>
<td></td>
<td>$E(e_t = 1</td>
</tr>
<tr>
<td></td>
<td>$E(e_t = 1</td>
</tr>
<tr>
<td></td>
<td>$E(e_t = 1</td>
</tr>
<tr>
<td></td>
<td>$E(e_t = 1</td>
</tr>
<tr>
<td></td>
<td>Mean Exercise by Education in $t = 1, 7$</td>
</tr>
<tr>
<td></td>
<td>$E(e_t = 1 &lt; HS) = 0.5303$</td>
</tr>
<tr>
<td></td>
<td>$E(e_t = 1</td>
</tr>
<tr>
<td></td>
<td>$E(e_t = 1</td>
</tr>
<tr>
<td></td>
<td>$E(e_t = 1</td>
</tr>
<tr>
<td>Terminal (Marginal) Value</td>
<td>${\Delta_2, \Delta_3, \Delta_4}$ Exercise in the Last Period by Health</td>
</tr>
<tr>
<td></td>
<td>$E(e_t = T</td>
</tr>
<tr>
<td></td>
<td>$E(e_t = T</td>
</tr>
<tr>
<td></td>
<td>$E(e_t = T</td>
</tr>
<tr>
<td>Parameters</td>
<td>Description</td>
</tr>
<tr>
<td>------------</td>
<td>-------------</td>
</tr>
<tr>
<td>$h_1$</td>
<td>Health Status</td>
</tr>
<tr>
<td>$h_2$</td>
<td></td>
</tr>
<tr>
<td>$h_3$</td>
<td></td>
</tr>
<tr>
<td>$h_4$</td>
<td></td>
</tr>
<tr>
<td>$A(t = 1, &lt; HS)$</td>
<td></td>
</tr>
<tr>
<td>$A(t = 2, &lt; HS)$</td>
<td></td>
</tr>
<tr>
<td>$A(t = 3, &lt; HS)$</td>
<td></td>
</tr>
<tr>
<td>$A(t = 4, &lt; HS)$</td>
<td></td>
</tr>
<tr>
<td>$A(t = 5, &lt; HS)$</td>
<td></td>
</tr>
<tr>
<td>$A(t = 6, &lt; HS)$</td>
<td></td>
</tr>
<tr>
<td>$A(t = 7, &lt; HS)$</td>
<td>Age, Education</td>
</tr>
<tr>
<td>$A(t = 1, HS)$</td>
<td>on Productivity</td>
</tr>
<tr>
<td>$A(t = 2, HS)$</td>
<td></td>
</tr>
<tr>
<td>$A(t = 3, HS)$</td>
<td></td>
</tr>
<tr>
<td>$A(t = 4, HS)$</td>
<td></td>
</tr>
<tr>
<td>$A(t = 5, HS)$</td>
<td></td>
</tr>
<tr>
<td>$A(t = 6, HS)$</td>
<td></td>
</tr>
<tr>
<td>$A(t = 7, HS)$</td>
<td></td>
</tr>
<tr>
<td>$\phi(Y, &lt; HS)$</td>
<td></td>
</tr>
<tr>
<td>$\phi(O, &lt; HS)$</td>
<td></td>
</tr>
<tr>
<td>$\phi(Y, HS)$</td>
<td></td>
</tr>
<tr>
<td>$\phi(O, HS)$</td>
<td></td>
</tr>
<tr>
<td>$\xi(Y, &lt; HS)$</td>
<td></td>
</tr>
<tr>
<td>$\xi(O, &lt; HS)$</td>
<td></td>
</tr>
<tr>
<td>$\xi(Y, HS)$</td>
<td></td>
</tr>
<tr>
<td>$\xi(O, HS)$</td>
<td></td>
</tr>
<tr>
<td>$\mu_\varepsilon$</td>
<td>Mean: health shock</td>
</tr>
<tr>
<td>$\sigma_\varepsilon$</td>
<td>St. Dev.: health shock</td>
</tr>
<tr>
<td>$\mu_{\varepsilon}(h_1)$</td>
<td>Mean: $z$-shock</td>
</tr>
<tr>
<td>$\mu_{\varepsilon}(h_2)$</td>
<td></td>
</tr>
<tr>
<td>$\mu_{\varepsilon}(h_3)$</td>
<td></td>
</tr>
<tr>
<td>$\mu_{\varepsilon}(h_4)$</td>
<td></td>
</tr>
<tr>
<td>$\gamma_1(&lt; HS)$</td>
<td>Disutility by Education</td>
</tr>
<tr>
<td>$\gamma_2(&lt; HS)$</td>
<td>by Education</td>
</tr>
<tr>
<td>$\gamma_1(HS)$</td>
<td>Measure of Fair in $t = T$</td>
</tr>
<tr>
<td>$\gamma_2(HS)$</td>
<td>Measure of Ex. in $t = T$</td>
</tr>
<tr>
<td>$p(\gamma_1</td>
<td>&lt; HS, h_1)$</td>
</tr>
<tr>
<td>$p(\gamma_1</td>
<td>&lt; HS, h_2)$</td>
</tr>
<tr>
<td>$p(\gamma_1</td>
<td>&lt; HS, h_3)$</td>
</tr>
<tr>
<td>$p(\gamma_1</td>
<td>&lt; HS, h_4)$</td>
</tr>
<tr>
<td>$p(\gamma_1</td>
<td>HS, h_1)$</td>
</tr>
<tr>
<td>$p(\gamma_1</td>
<td>HS, h_2)$</td>
</tr>
<tr>
<td>$p(\gamma_1</td>
<td>HS, h_3)$</td>
</tr>
<tr>
<td>$p(\gamma_1</td>
<td>HS, h_4)$</td>
</tr>
<tr>
<td>$\Delta_2$</td>
<td>Value of health in $t = T$</td>
</tr>
<tr>
<td>$\Delta_3$</td>
<td></td>
</tr>
<tr>
<td>$\Delta_4$</td>
<td></td>
</tr>
</tbody>
</table>

Table B.9: Calibrated Parameters
B.7 Additional Quantitative Results

B.7.1 Model Fit

Figures B.13–B.16 represent the model fit for average effort of each health level.

Figure B.13: Fair

Figure B.14: Good

Figure B.15: Very Good

Figure B.16: Excellent
Insurance Benefits  Tables 12 and 13 present the weighted-averages (across education and exercise preference) of the cross-subsidies by health level under different policy regimes. We measure cross subsidies in premium by the differences between the actuarially fair health premium and premium paid under policies; and cross subsidies in wage by the differences between the aggregate wage and productivity of the worker (of a given health level). As discussed in the main text, the negative cross-subsidy implies that the worker is paying higher premium than the actuarially fair price and/or getting paid less in wages than he produces.

Since under no-prior conditions law, only premium is subsidized, and under no-wage discrimination law, only wage is subsidized, we report cross-subsidies of premium and wages under each law. The second row under each health level reports separately the subsidies of premium and wage, under both policies.

Table B.10: Cross-Subsidy by Health Level under Different Policy Regimes: Young

<table>
<thead>
<tr>
<th>Health</th>
<th>Policy</th>
<th>24–29</th>
<th>30–35</th>
<th>36–41</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Prem.</td>
<td>Wage</td>
<td>Prem.</td>
</tr>
<tr>
<td>Fair</td>
<td>One Policy</td>
<td>0.276</td>
<td>0.285</td>
<td>0.306</td>
</tr>
<tr>
<td></td>
<td>Both Policies</td>
<td>0.141</td>
<td>0.247</td>
<td>0.123</td>
</tr>
<tr>
<td>Good</td>
<td>One Policy</td>
<td>0.041</td>
<td>0.107</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>Both Policies</td>
<td>0.011</td>
<td>0.102</td>
<td>-0.007</td>
</tr>
<tr>
<td>Very Good</td>
<td>One Policy</td>
<td>-0.030</td>
<td>-0.029</td>
<td>-0.085</td>
</tr>
<tr>
<td></td>
<td>Both Policies</td>
<td>-0.011</td>
<td>-0.026</td>
<td>-0.034</td>
</tr>
<tr>
<td>Excellent</td>
<td>One Policy</td>
<td>-0.071</td>
<td>-0.139</td>
<td>-0.114</td>
</tr>
<tr>
<td></td>
<td>Both Policies</td>
<td>-0.033</td>
<td>-0.129</td>
<td>-0.054</td>
</tr>
</tbody>
</table>

Table B.11: Cross-Subsidy by Health Level under Different Policy Regimes: Old

<table>
<thead>
<tr>
<th>Health</th>
<th>Policy</th>
<th>42–47</th>
<th>48–53</th>
<th>54–59</th>
<th>60–65</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fair</td>
<td>One Policy</td>
<td>0.352</td>
<td>0.481</td>
<td>0.3402</td>
<td>0.470</td>
</tr>
<tr>
<td></td>
<td>Both Policies</td>
<td>0.105</td>
<td>0.411</td>
<td>0.103</td>
<td>0.404</td>
</tr>
<tr>
<td>Good</td>
<td>One Policy</td>
<td>-0.034</td>
<td>0.094</td>
<td>-0.044</td>
<td>0.080</td>
</tr>
<tr>
<td></td>
<td>Both Policies</td>
<td>-0.024</td>
<td>0.073</td>
<td>-0.025</td>
<td>0.066</td>
</tr>
<tr>
<td>Very Good</td>
<td>One Policy</td>
<td>-0.151</td>
<td>-0.318</td>
<td>-0.155</td>
<td>-0.333</td>
</tr>
<tr>
<td></td>
<td>Both Policies</td>
<td>-0.050</td>
<td>-0.327</td>
<td>-0.052</td>
<td>-0.337</td>
</tr>
<tr>
<td>Excellent</td>
<td>One Policy</td>
<td>-0.173</td>
<td>-0.505</td>
<td>-0.177</td>
<td>-0.524</td>
</tr>
<tr>
<td></td>
<td>Both Policies</td>
<td>-0.073</td>
<td>-0.307</td>
<td>-0.075</td>
<td>-0.519</td>
</tr>
</tbody>
</table>
Welfare Implications

Tables B.12 and B.13 present the static and dynamic consumption equivalent variations for each \((educ, \gamma)\)-groups as well as the aggregates.

Table B.12: Welfare Comparisons in Static Economy

\[
\begin{array}{|l|c|c|c|c|}
\hline
& (< HS, \gamma_L) & (< HS, \gamma_H) & (HS \text{ Grad}, \gamma_L) & (HS \text{ Grad}, \gamma_H) & \text{Aggregate} \\
\hline
\text{Social Planner} & 2.6341 & 8.4298 & 1.6131 & 7.5427 & 5.6527 \\
\text{Competitive Equilibrium} & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
\text{No Prior Conditions Law} & 1.9557 & 7.1836 & 1.1334 & 5.2777 & 4.1593 \\
\text{No Wage Discrimination Law} & 2.4443 & 7.4681 & 1.5778 & 7.2692 & 5.3486 \\
\text{Both Policies} & 2.6341 & 8.4298 & 1.6131 & 7.5427 & 5.6527 \\
\hline
\end{array}
\]

Table B.13: Welfare Comparisons in Dynamic Economy

\[
\begin{array}{|l|c|c|c|c|}
\hline
& (< HS, \gamma_L) & (< HS, \gamma_H) & (HS \text{ Grad}, \gamma_L) & (HS \text{ Grad}, \gamma_H) & \text{Aggregate} \\
\hline
\text{Social Planner} & 7.8120 & 14.4481 & 17.1213 & 17.8447 & 16.4799 \\
\text{Competitive Equilibrium} & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
\text{No Prior Conditions Law} & 5.4108 & 7.7063 & 5.8094 & 7.6374 & 6.9782 \\
\text{Both Policies} & 4.9908 & 6.7668 & 8.3978 & 8.8680 & 8.1656 \\
\hline
\end{array}
\]

Moreover, in Table B.14 are the lifetime welfare comparisons in the dynamic economy, conditional on health and \((educ, \gamma)\)-group.

Table B.14: Lifetime Welfare Comparisons in the Dynamic Economy Conditional on Type and Health

\[
\begin{array}{|l|l|l|l|l|}
\hline
\text{Type} & \text{Policy} & \text{Fair} & \text{Good} & \text{Very Good} & \text{Excellent} \\
\hline
\text{Low Educ, Low } \gamma & \text{Social Planner} & 45.6618 & 8.3078 & 7.3379 & 2.5327 \\
& \text{Comp. Eq.} & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
& \text{No Prior} & 34.3609 & 6.9402 & 3.0803 & -0.4095 \\
& \text{No Wage} & 34.9972 & 9.7912 & 3.1518 & -3.5944 \\
& \text{Both} & 46.4916 & 9.0499 & 1.0363 & -7.0393 \\
\hline
\text{Low Educ, High } \gamma & \text{Social Planner} & 46.4190 & 9.7199 & 8.1054 & 3.1847 \\
& \text{Comp. Eq.} & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
& \text{No Prior} & 31.4642 & 5.7174 & 1.7938 & -1.5060 \\
& \text{No Wage} & 33.8613 & 9.1959 & 2.4047 & -4.2708 \\
& \text{Both} & 42.5672 & 6.9865 & -1.2339 & -9.1102 \\
\hline
\text{High Educ, Low } \gamma & \text{Social Planner} & 69.4954 & 18.5571 & 18.0361 & 14.2185 \\
& \text{Comp. Eq.} & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
& \text{No Prior} & 49.3466 & 13.5551 & 6.6835 & 2.3228 \\
& \text{No Wage} & 66.9188 & 24.0447 & 11.2267 & 1.8703 \\
& \text{Both} & 78.9321 & 25.3647 & 11.2843 & 0.7706 \\
\hline
\text{High Educ, High } \gamma & \text{Social Planner} & 62.4530 & 15.3438 & 13.7657 & 9.6096 \\
& \text{Comp. Eq.} & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
& \text{No Prior} & 38.6852 & 8.1877 & 2.9285 & -0.4819 \\
& \text{No Wage} & 52.4471 & 15.7693 & 4.0127 & -4.5987 \\
& \text{Both} & 60.3707 & 14.3834 & 1.4297 & -8.1014 \\
\hline
\end{array}
\]
B.8 Sensitivity Analysis

B.8.1 Robustness of Results with Respect to Age and Gender

The PSID asks questions on ethnicity, and among them, we take those who answered to be of a national origin (47% of the total sample in 1997) to test robustness. We also restrict our sample to males (about 77%) for the second robustness check.

The health transition function and production function related parameters are the key driving forces of our quantitative results. Therefore, we provide evidence for the similarity in health transition and the labor earnings over the life cycle between the total population and the subsamples.

For the health transition function $Q(h'|h,e)$, we obtain a measure of differences in the estimated probabilities and the data moments, i.e., $\chi^2 = \sum_{i=1}^{N} \frac{q^{data}(h'^{i}) - Q^{est}(h'^{i})}{Q^{est}(h'^{i})}$, where the $q^{data}(h'^{i})$ and $Q^{est}(h'^{i})$ are the actual data and the estimated probability of a worker with initial health status $h$ with exercise level $e^{i72}$ ending up being health status of $h'$ in the next period. The $\chi^2$ value for the health transition is 1.16 and 1.02 for whites and males, where the $\chi^2_{49,0.05}^{2}$ is 79.

With regards to the production function, we provide in Table B.15, the data moments associated with the subsamples, in comparison with the full sample. The qualitative features of the moments are similar across different samples: although the absolute numbers for the changes in income over the life-cycle vary in their levels, the gradients over the life cycle are similar. Thus our quantitative results are robust to restricting our samples to white and males.

---

71The exact choices are American (5%); Hyphenated American (e.g., African-American, Mexican-American) (14%); National origin (e.g., French, German, Dutch, Iranian, Scots-Irish) (47%); Nonspecific Hispanic identity (e.g., Chicano, Latino) (2%); Racial (e.g., white or Caucasian, black) (29%) and; 6 Religious (e.g., Jewish, Roman, Catholic, Baptist).

72We divide the population into five exercise bins, and use them to evaluate the differences, as we do in our estimation procedure. The only difference is that due to the shortage of observations (since we only use half the total sample), instead of nine bins (in the full model), we use five bins.

73The degrees of freedom is 49, as the number of observations are $4 \times 4 \times 5 \times 2 \times 2 \times 5$ (Health Today $\times$ Health Tomorrow $\times$ Exercise Bins), and the number of parameters, 30 (80-1-30). Using the full sample, the $\chi^2$ value is 0.9986.
Table B.15: Moments for the Subsample of Population

<table>
<thead>
<tr>
<th>Moments Description</th>
<th>All</th>
<th>Whites</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$t = 1$</td>
<td>-0.0042</td>
<td>-0.0892</td>
</tr>
<tr>
<td></td>
<td>$t = 2$</td>
<td>0.1449</td>
<td>0.2026</td>
</tr>
<tr>
<td></td>
<td>$t = 3$</td>
<td>0.1715</td>
<td>0.2464</td>
</tr>
<tr>
<td>Income by Age of Less than HS</td>
<td>$t = 4$</td>
<td>0.1980</td>
<td>0.2901</td>
</tr>
<tr>
<td></td>
<td>$t = 5$</td>
<td>0.0907</td>
<td>0.0014</td>
</tr>
<tr>
<td></td>
<td>$t = 6$</td>
<td>-0.0969</td>
<td>-0.3306</td>
</tr>
<tr>
<td></td>
<td>$t = 7$</td>
<td>-0.1112</td>
<td>-0.0970</td>
</tr>
<tr>
<td>Income by Age of HS Grad.</td>
<td>$t = 1$</td>
<td>0.2980</td>
<td>0.3019</td>
</tr>
<tr>
<td></td>
<td>$t = 2$</td>
<td>0.4738</td>
<td>0.5867</td>
</tr>
<tr>
<td></td>
<td>$t = 3$</td>
<td>0.5082</td>
<td>0.6073</td>
</tr>
<tr>
<td></td>
<td>$t = 4$</td>
<td>0.5988</td>
<td>0.6274</td>
</tr>
<tr>
<td></td>
<td>$t = 5$</td>
<td>0.6060</td>
<td>0.6500</td>
</tr>
<tr>
<td></td>
<td>$t = 6$</td>
<td>0.5395</td>
<td>0.5276</td>
</tr>
<tr>
<td></td>
<td>$t = 7$</td>
<td>0.2406</td>
<td>0.1792</td>
</tr>
<tr>
<td>% Income Spent on Med. Exp.</td>
<td>Fair</td>
<td>0.0525</td>
<td>0.0573</td>
</tr>
<tr>
<td></td>
<td>Good</td>
<td>0.0429</td>
<td>0.0428</td>
</tr>
<tr>
<td></td>
<td>Very Good</td>
<td>0.0353</td>
<td>0.0376</td>
</tr>
<tr>
<td></td>
<td>Excellent</td>
<td>0.0308</td>
<td>0.0320</td>
</tr>
<tr>
<td>% Income Spent on Med. Exp. by Less than HS</td>
<td>Young</td>
<td>0.0386</td>
<td>0.0350</td>
</tr>
<tr>
<td></td>
<td>Old</td>
<td>0.0348</td>
<td>0.0357</td>
</tr>
<tr>
<td>% Income Spent on Med. Exp. by HS Grad</td>
<td>Young</td>
<td>0.0428</td>
<td>0.0465</td>
</tr>
<tr>
<td></td>
<td>Old</td>
<td>0.0356</td>
<td>0.0379</td>
</tr>
</tbody>
</table>

B.8.2 Benefits of Effort Not Related to Labor Productivity

So far the only benefit of effort $e$ consisted in probabilistically raising health in the future which in turn impacts positively future wages and health insurance premia. As a result, a combination of both policies reduces optimal effort to zero, unless a health-dependent terminal continuation utility (as in the quantitative version of our model) is introduced. We now briefly argue that our main results do not necessarily hinge on this assumption. Suppose that the net cost of providing effort is given by

$$\gamma [q(e) - \theta e].$$

Our previous specification is a special case with $\theta = 0$, and $\gamma \theta$ measures the direct utility benefit from one unit of exercise. In the absence of any other benefits from exercise (say, from higher wages or lower health insurance premia), as in the economy with both laws in
place, the optimal effort level \( e^{BP} \) now solves

\[
q'(e^{BP}) = \theta
\]

and thus \( e^{BP} > 0 \) if and only if \( \theta > 0 \). Thus for a given function \( q \) the parameter \( \theta \) governs the minimal effort level that each household will provide, and thus a lower bound below which no policy can distort effort levels.

The equations determining optimal effort levels (equation (2.26) for the social planner problem and equation (2.29) for the competitive equilibrium under the various policies) with preference shocks \( \gamma \) and direct utility benefits from exercising \( \gamma \theta e \) now become

\[
q'(e_t(h)) = \theta + \frac{\beta}{\gamma} \sum_{h'} \frac{\partial V_{t+1}(\Phi_{t+1})}{\partial \Phi_{t+1}(h')} \cdot \frac{\partial Q(h'; h, e_t(h))}{\partial e_t(h)}
\]

\[
q'(e_t(h)) = \theta + \frac{\beta}{\gamma} \sum_{h'} \frac{\partial Q(h'; h, e_t(h))}{\partial e_t(h)} v_{t+1}(h')
\]

and for any given initial health level \( h \), for any preference shock \( \gamma \) and any policy the optimal effort level is simply shifted upwards.
Bibliography


and Intergenerational Mobility,” Journal of Political Economy, 87, 1153-1189.

from the GHS,” The Economic Journal, 107(441), 333-344.

What Levels of Redistribution Maximize Growth and Efficiency?,” Econometrica,
70(2), 481-517.

forthcoming.


in Britain,” Review of Economic Dynamics, 13(1) 76-102.

nomic Review, 77 531-553.


Earnings,” The Mirrlees Review: Reforming the Tax System for the 21st Century,
Oxford University Press.

librium Models,” in J.B. Taylor and M. Woodford, eds, Handbook of Macroeconomics,
17,18.


Research Memorandum*, No. 172, The Hague

[44] Grogger, J. and Hanson, G.H., 2011, “Income Maximization and the Selection and 


*Journal of Political Economy*, 80(2), 223-255.


Working Paper, University of Hawaii.


