Essays on Public Economics

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Hitoshi Tsujiyama

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
Doctor of Philosophy

José-Víctor Ríos-Rull, Advisor

May, 2013

Acknowledgements

I am deeply grateful to Jonathan Heathcote, José-Víctor Ríos-Rull, Motohiro Yogo and Roger Feldman for their advice and continuous encouragement. I also thank Cristina Arellano, Jahiz Barlas, Elisa Belfiori, Manuel Macera, Ellen McGrattan, Sam Schulhofer-Wohl, Kjetil Storesletten, Naoki Takayama, Satoshi Tanaka, Yuichiro Waki and participants at seminars at the University of Minnesota and the Federal Reserve Bank of Minneapolis for comments and discussions.

I would also like to thank the staff members of the University of Minnesota and the Federal Reserve Bank of Minneapolis.

I dedicate this dissertation to my wife, Natsuki, who always supports me, and my parents in Japan.

Abstract

This dissertation consists of two essays. The first essay examines how upcoming health care reform aiming to increase insurance coverage affects savings and welfare. I develop a heterogeneous-agent life-cycle model of insurance choice in which households face medical expenditure risk in addition to uninsurable income risk. The existence of an uninsured population reflects limited liability in the medical services market and actuarially unfair premiums in the health insurance market. I estimate key structural parameters so that the model broadly replicates the joint distribution of insurance coverage by age, health, earnings and wealth across active participants in the individual insurance market. I then use the model to explore the implications of the Affordable Care Act which prohibits insurers from price-discriminating based on health risk and mandates the purchase of insurance. I find that the health care reform induces wealth accumulation among the uninsured poor who no longer take advantage of limited liability, leading to lower wealth inequality. The reform also generates welfare gains for the rich who are less exposed to risk, but welfare losses for the poor who lose the benefit of access to free emergency care.

In the second essay co-authored with Jonathan Heathcote, we quantitatively study the optimal income tax schedule. We build on the standard static Mirrlees economy in which heterogeneous households choose their labor supply given their productivity, and the planner chooses income tax schedule to maximize the social welfare. In the model, the planner can condition the tax schedule on observable personal attributes, and we calibrate the relative variances of the observable and unobservable components from wage regressions on the relative variances of between-group versus within-group wage dispersion. We then solve for the constrained-efficient allocation following Mirrlees approach and look for simple approximate implementations in the class of polynomial functions, following Ramsey approach. We find that the efficient allocation is highly distortionary; the output decreases by around 12% relative to a baseline flat tax, whereas welfare gains from redistribution are as large as 16%. More importantly, we show that the efficient allocation can be approximately implemented by very simple and realistic tax schemes. Specifically, a linear income tax with lump-sum transfers can achieve almost 96% of the gains, and a third-order polynomial tax function approximately implements the constrained-efficient allocation. It is crucial to condition both lump-sum transfers and marginal tax rates on observables. If those are type-independent, 25% of the potential gains are lost.

Contents

List of Tables

List of Figures

Chapter 1

Implications of Health Care Reform for Inequality and Welfare

1.1 Introduction

This paper examines how health care reform affects health insurance coverage, savings and welfare. Medical expenditure is a major source of risk. However, in the United States among those who are neither eligible for public health insurance nor offered employer-provided insurance, more than three-fourths are uninsured. In order to expand insurance coverage in the private health insurance market, the Affordable Care Act (ACA) introduces guaranteed issue and community rating regulations that prevent insurers from rejecting or price-discriminating against applicants on the basis of health status. The Act also mandates holding insurance and provides means-tested price subsidies. This study explores the implications of these reforms for participation in private health insurance markets, individual saving behavior, and distribution of welfare gains and losses.

To understand the implications of health insurance reform, I first develop a heterogeneous-agent life-cycle model that incorporates idiosyncratic income and health risk, as well as key features of the current U.S. insurance system. I then estimate structural parameters so that the model replicates the current insurance coverage pattern in the data. Finally, using the calibrated model as a laboratory, I conduct an experiment of implementing the ACA to investigate the effects of the health insurance reform.

Households in the model face a combination of idiosyncratic income and health risk. They are also subject to medical expenditure shocks. Poor health translates into higher expected medical expenses, and lower expected future income. The model also captures the following key features of the current U.S. insurance system. A small fraction of individuals of working age are covered by Medicaid. The rest can buy private health insurance either from their employer, if employer-sponsored insurance is offered, or directly from an insurance company. Employer-sponsored group insurance is not risk-rated and a large fraction of the premium is paid by the employer. In contrast, in the individual health insurance market the insurance company is allowed to price-discriminate based on health status and age.

I assume fixed costs of issuing health insurance which capture the combined effects of administrative and screening costs, and translate into premiums that exceed actuarially fair values. This drives some rich and/or healthy agents to choose to be uninsured. I also model the provision of U.S. federal law which requires hospitals to treat individuals in urgent need of medical attention, regardless of ability to pay (Emergency Medical Treatment and Active Labor Act). As a result some poor individuals choose to be uninsured because they anticipate receiving free care in the event of a large negative health shock. This free rider problem due to limited liability leads to uncompensated care costs to the hospital, and to make up these costs, the hospital charges those who can pay more for the services provided, resulting in higher insurance premiums (Gruber (2008))[1].

I calibrate a joint stochastic process for earnings, health status, medical expenses, and availability of public and employer-provided insurance using panel data from the Survey of Income and Program Participation (SIPP) and the Medical Expenditure Panel Survey (MEPS). The process for health status and medical expenses varies with insurance status, to capture the idea that the insured enjoy easier access to primary care. I then estimate the key structural parameters – the degree of risk aversion and the fixed costs of issuing health insurance – so that the model broadly replicates the current insurance coverage pattern in the U.S. Specifically, the model replicates the joint distribution of insurance coverage by age, health, earnings and wealth of active participants in the individual insurance market – the population that is neither eligible for public health insurance nor offered employer-provided insurance, the target population of the policy reform. The resulting

[1]Chatterjee, Corbae, Nakajima, and Ros-Rull (2007) document that household bankruptcy generates a mark-up in the medical service market. The mechanism whereby the uninsured population and the free rider problem generate negative externalities in the form of higher premiums in this paper is similar to Smith and Wright (1992) who study automobile insurance markets.

estimates show that risk aversion is in the range of the standard models but relatively low (1.23), and the fixed costs of issuing insurance are high and constitute 26% of the average premium. I then conduct a decomposition exercise in which I assess the relative importance in accounting for low current coverage rates of (i) high fixed costs in the insurance market, (ii) relatively low risk aversion within the pool of potential buyers of insurance, and (iii) limited liability in the health services market.

With a diagnosis of what ails the current insurance market in hand, I move to assess whether the ACA is an appropriate remedy. In particular, I use the model to examine the implications of the two key provisions in the ACA aiming to expand the coverage rate in the individual health insurance market. First, community rating together with guaranteed issue prohibits insurers from rejecting or price-discriminating against applicants based on health status, and creates a pool of risks in the individual insurance market. Second, to avoid adverse selection problems, the ACA penalizes those without health insurance, and establishes health insurance exchanges that provide insurance premium subsidies for individuals with income up to 400% of the poverty line[2].

The counter-factual experiment of implementing the ACA delivers three sets of quantitative results. First, the reform substantially increases the participation rate in the individual insurance market, from 22.9% to 88.1%. It reduces the mark-up in the medical services market due to uncompensated care, because the number of people who take advantage of the free care substantially decreases. Those who remain uninsured tend to be wealth rich, have moderate income, and be in good health. This group is well positioned to self-insure expenditure shocks, and faces an especially actuarially unfair premium due to community rating on top of administrative costs. The higher aggregate coverage results in an improvement in the average health status among the working age population because more people have access to primary care. Nevertheless the fraction of aggregate health spending in GDP increases by 0.3 percentage points because of a significant increase in the usage of primary care.

Second, wealth inequality decreases: the Gini coefficient of the active participants declines from 0.653 to 0.634. The uninsured poor – the major users of free care before the reform – are mostly insured after the reform. Once they no longer fear that any savings could be lost to pay for

[2]Some states have already introduced similar types of regulations. For example, Maine, Massachusetts, Vermont, and Washington introduced modified community rating that prohibits insurers from risk-rating applicants based on health. New Jersey and New York enacted pure community rating that bans any price-discrimination. Massachusetts also has mandated health insurance holdings since 2006.

health care expenses, they have stronger incentive to accumulate savings. Thus they are no longer trapped at the bottom of the wealth distribution. On the other hand, as the reform provides easier access to private insurance when having an adverse health or income shock, it effectively reduces the risk exposure of the rich, discouraging precautionary saving. Therefore, the wealth accumulation increases among the poor and decreases among the rich.

Third, the reform generates a significant welfare gain on average for the active participants of the individual insurance market (equivalent to 0.2% of consumption). More than 50% of them are in favor of the reform. The unhealthy and those with low income tend to gain thanks to the direct transfers associated with community rating and income-tested premium subsidies. However, because the reform forces those with little or no wealth to obtain insurance rather than take advantage of limited liability, they experience overall welfare losses. In contrast, the wealth rich tend to experience welfare gains, because the reform makes it easier to access to insurance when one's health deteriorates and/or one is hit by a negative income shock.

The contribution of this paper is threefold. First, this paper is one of the first quantitative macroeconomic studies to tackle the issue of the large uninsured population in the United States, using a dynamic general equilibrium incomplete markets model with heterogeneous agents in the tradition of Bewley (1986), Imrohoroglu (1989), Huggett (1993), and Aiyagari (1994). Specifically, I explicitly incorporate the two main explanations for the large uninsured population proposed in Gruber (2008), i.e., actuarial unfairness in the U.S. insurance market, and implicit insurance through uncompensated care. I then estimate the individual risk attitudes and the fixed costs of issuing health insurance from the observed pattern of the insurance coverage in the U.S. Understanding why the U.S. has such a large uninsured population is an important prerequisite for studying the implications of upcoming health care reform.

The second contribution of this paper is that, departing from the literature, I carefully estimate a process for eligibility for public health insurance jointly with the other shocks using panel data on income, health status, and insurance status. This is important because existing papers that analyze the effect of the health care reform using macroeconomic models assume – counterfactually – that low income guarantees public insurance eligibility. Thus those models generate too few low income households in the pool of the uninsured, the target population of the reform (Jeske and Kitao (2009), Janicki (2011), Jung and Tran (2011), Hansen, Hsu, and Lee (2012), Pashchenko and Porapakkarm (forthcoming))[3]. In contrast, my model successfully replicates the income and wealth distribution within the uninsured population. This further allows me to investigate the welfare effects of the reform on the population with different characteristics – age, earnings, health status and wealth.

The third contribution is that I show that the free care option due to limited liability distorts the savings decision of the poor. A similar mechanism is proposed in Hubbard, Skinner, and Zeldes (1995) who study the role of social insurance on savings and document that the social safety net discourages low income households from savings. I show that limited liability in the medical services market plays the same role as the social safety net in their model, and poor households are trapped at the bottom of the wealth distribution by a similar mechanism. I then analyze the effect of the reform on this margin, showing that higher insurance coverage potentially weakens this distortion and reduces wealth inequality. This linkage between higher insurance coverage and lower wealth inequality is not typically present in the existing papers that assume income-tested public health insurance before the policy reform and hence exhibit no change in the insurance status of low income households before and after the reform.

This paper contributes to the growing literature that incorporates idiosyncratic health risks over the life-cycle into the incomplete market model with heterogeneous agents and analyzes households' insurance decision. The paper by Jeske and Kitao (2009) is the first work that has endogenous health insurance decision built into the standard incomplete market model. They analyze the policy implications of tax benefits for the employer-provided health insurance. Attanasio, Kitao, and Violante (2010) study the effect of Medicare reform in a heterogeneous-agent economy. Hansen, Hsu, and Lee (2012) investigate the effect of introducing Medicare Buy-in that allows people of age 55 to 64 to purchase Medicare. Other quantitative papers that incorporate health production in the incomplete market model include Zhao (2011), Jung and Tran (2011), He and Huangy (2012), and Cole, Kim, and Krueger (2012).

The closest paper to this study is the recent work by Pashchenko and Porapakkarm (forthcoming) who build a rich model of insurance and labor participation choice, and analyze the welfare implications of the policy reform similar to this paper[4]. The main innovation of my paper relative

[3] In the United States, poverty alone does not guarantee eligibility for public insurance. Individuals must also fall into a specific eligibility category such as veterans, pregnant women, the disabled, parents of the eligible child. Specifically, in 43 states, adults without any dependent child can never qualify for Medicaid even if they are penniless. Among the active participants with income in the bottom quartile, i.e., those who are automatically eligible for Medicaid in the existing papers, 82.3% are uninsured and 75.3% are not having any dependent child in the SIPP data.

[4] Unlike Pashchenko and Porapakkarm (forthcoming), this paper does not consider the effect of the Medicaid expansion proposed in the ACA. The reason is because the ACA originally planned to expand Medicaid to people with income up to 133% of the federal poverty line. However, in 2012, the Supreme Court ruled that participation

to theirs is the focus here on replicating the current pattern of insurance coverage, which allows for a sharper characterization of the effects of the reform on savings, insurance purchase and welfare for actual U.S. participants in the individual insurance market. Pashchenko and Porapakkarm (forthcoming) model the discrete labor supply decision from which this paper abstracts and find a reduction in the aggregate labor participation after the reform, although the main driving force of this change is not the tax distortion associated with the redistributive policy change but the Medicaid expansion in their model.

This paper also contributes to the literature that studies precautionary savings in response to health risks. Kotlikoff (1989) argues that idiosyncratic health shocks can have a large effect on saving behavior. Kopecky and Koreshkova (2011) investigate how medical and nursing home expenses affect the savings of the wealthy. De Nardi, French, and Jones (2010) study the effect of uncertainty about longevity and health expenditures on the saving behavior of the elderly.

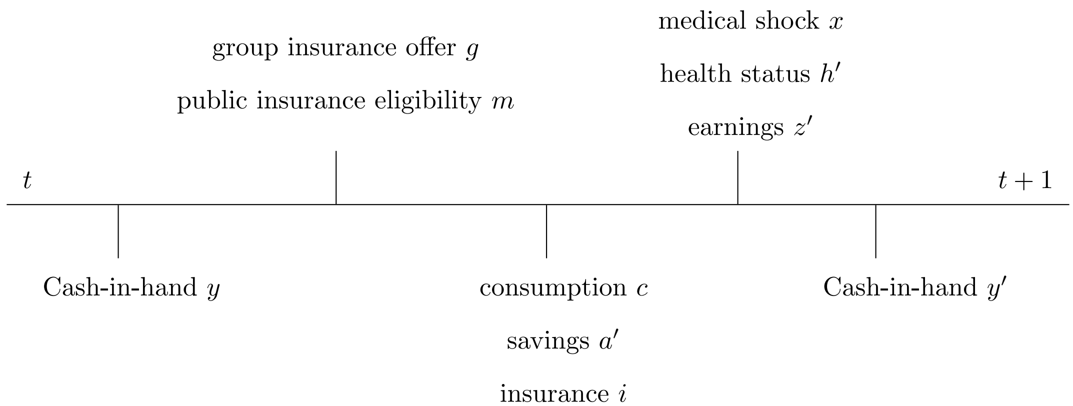
1.2 Dynamic Model of Insurance Choice

I describe the heterogeneous-agent life-cycle model of insurance choice in a stationary economic environment to account for the *pre-Act* situation in the United States. In the model, households face a combination of idiosyncratic income and health risk. They are also subject to medical expenditure shocks. They have two ways to bear these risks: obtaining health insurance and accumulating assets.

in the Medicaid expansion is voluntary for states. Hence, the eligibility would vary across states. In fact, five states have already announced that they would decline to participate. Also, because of this, there is still the potential for the eligibility condition to be altered before becoming effective. Therefore, I do not analyze this policy change in this paper.

1.2.1 Economic environment

The timing of the economy is illustrated in the following figure and described in detail below.



Demographics

Time is discrete. There are J overlapping generations of households of equal measure in the economy. Households enter the labor market at age $j = 0$, retire at $j = J_R$ and die at $j = J$ with certainty.

Health Status

Households face uncertainty about their health status h and it takes one of the two values, $h \in \mathbb{H} \equiv \{good, bad\}$. The health status evolves stochastically according to the Markov chain $\Gamma_h(h'|h, j, i)$ whose evolution depends on the current health status h, age j and insurance status i. The dependence of Γ_h on the insurance status captures the fact that households with health insurance have access to *primary care* in the U.S. that improves their health status. This is a reduced-form notion of health production that is originally considered in Grossman (1972), and having health insurance and hence receiving primary care are interpreted as health investment.

Medical Services

Medical treatment is non-discretionary. The medical shock $x \in \mathbb{X} \subseteq \mathbb{R}_+$ conceptually comprises two types of care, *urgent care* and *primary care*. The urgent care corresponds to urgent medical attention, including emergency care, whereas the primary care corresponds to all treatments that improve one's health outcome such as preventive care, treatments of diabetes and kidney dialysis treatment. Only insured people have access to the primary care, for in the United States hospitals can and often do decline treatment of the uninsured unless it is urgent. Therefore, the random medical shocks x are from the distribution that depends on insurance status i as well as health status h, and age j, i.e., $\Pi(x|h, j, i)$. The persistence in medical shocks is captured by the persistence in health status. For an amount of medical care x, households are charged qx by the medical service provider, where $q \geq 1$ is the mark-up of the medical services, which is described in Section 1.2.3.

Earnings

Households obtain labor earnings of $w\varepsilon z$ every period, where $w \in \mathbb{R}_+$ is the market wage, $\varepsilon \in \mathbb{R}_+$ is the age-dependent labor efficiency, and $z \in \mathbb{R}_+$ is the idiosyncratic labor productivity shock. The age-dependent labor efficiency ε captures the hump-shaped labor productivity over the life-cycle and evolves deterministically. On the other hand, z evolves stochastically according to the Markov chain $\Gamma_z(z'|z, h')$ that depends on the previous productivity z and the health status in the same period h'. Healthy people are more likely to be productive.

Insurance

There are three types of health insurance for the working age population in this economy: public insurance (i.e., Medicaid), employer-provided group insurance and individual insurance. Households are randomly assigned Medicaid eligibility, but those with low income and/or in poor health are more likely to be eligible for Medicaid. This captures the fact that Medicaid is designed for designated eligibility categories such as the disabled, pregnant women and low income parents, but poverty alone does not guarantee eligibility. Specifically, the eligibility for Medicaid $m \in \{0, 1\}$ is stochastic and follows the Markov chain $\Gamma_m(m'|m, z', h')$. Also, Medicaid is free.

When households are not eligible for Medicaid, they can purchase private health insurance before the medical shock realizes. Employer-provided group insurance is available only if it is offered by the employer, while individual insurance is available for everybody. They both have the same reimbursement schedule, but differ in the premium $p \in \mathbb{R}_+$. In the United States, group plans benefit from tax-deductibility of the premium and the employer's contribution, which is implicitly captured through the premium p in the model. The employer-provided group insurance offer $g \in \{0, 1\}$ is stochastic and follows the Markov chain $\Gamma_g(g'|g, z')$. Note that the process depends on the realization of the labor productivity z'. More precisely, the higher is the labor income, the higher is the probability of the group plan being offered.

If employer-provided group insurance is not offered by the employer, households choose whether to purchase private insurance from the individual insurance market. Individual insurance have the same reimbursement schedule as employer-provided group insurance, but agents need to pay premium p on their own. The premium is described in Section 1.2.4.

After the retirement age J_R, all households are covered by the Medicare which is assumed to be free for simplicity.

The reimbursement schedule is given by a function λ. That is, given an amount of the medical shock qx and an insurance plan, the function λ determines the fraction of qx that is reimbursed to the agent by the insurance plan. The insurance plan is either private insurance or public insurance for the working age population, and Medicare for the retired population.

Social Security and Government

After the retirement, households get Social Security benefit $ss \in \mathbb{R}_+$ in each period. The government imposes a proportional tax τ on the labor income of the working age population in order to finance the Social Security benefit in addition to Medicaid and Medicare.

1.2.2 Household Problem

Households choose consumption, savings and insurance status to maximize their lifetime utility:

$$\mathbb{E}_0 \sum_{j=0}^{J} \beta^j u(c_j).$$

I separate the problem into the working age problem and the after-retirement problem. I construct these household problems in a recursive fashion.

Working Age Problem

The timing is as follows. Households start a period with cash-in-hand, $y \in \mathbb{R}_+$. They observe the group insurance offer $g \in \{0, 1\}$ and Medicaid eligibility $m \in \{0, 1\}$. Next, households make decisions about consumption $c \in \mathbb{R}_+$, savings $a' \in \mathbb{R}_+$, and insurance status $i \in \{0, 1\}$. Then, the medical shock $x \in \mathbb{R}_+$ is realized. Finally, the new health status $h' \in \mathbb{H}$ and idiosyncratic labor productivity

$z' \in \mathbb{R}_+$ are given, and tomorrow's cash-in-hand y' is determined with which households enter next period. See also the figure in Section 1.2.1.

The determination of y' reflects limited liability in the medical services market. Suppose a household faces the medical bill of qx. A fraction determined by a function λ is covered if insured, and the residual represents out-of-pocket expenses. If the household's asset holdings after earning interest $(1 + r)a'$ are not large enough to settle the out-of-pocket medical expenditure, then it pays as much as it can, and is exempted from the remaining charge. Therefore, the asset holdings net of medical expenditure is determined by $\max\{(1 + r)a' - (1 - \lambda(qx, m)i)qx, 0\}$. Note that the reimbursement schedule $\lambda : \mathbb{R}_+ \times \{0, 1\} \to [0, 1]$ is a function of the size of the medical expenditure and the types of the insurance, i.e., $m = 0$ or 1. Then, y' is given by the sum of this term and the current income, that is,

$$y' = \max\{(1 + r)a' - (1 - \lambda(qx, m)i)qx, 0\} + (1 - \tau)wz'\varepsilon'.$$

Note that even when households go into medical bankruptcy, they can keep the income in the same period, which is consistent with U.S. law (Chatterjee, Corbae, Nakajima, and Ros-Rull (2007)).

I define the joint state vector of the working age problem as $s = (j, z, g, m, h) \in \mathbb{S}$ where $j \in \{0, \cdots, J_R - 1\}$ stands for the household's age. Then the recursive formulation of the working age problem is given by

$$\begin{cases} V(y, s) = & \max\limits_{c \geq 0, a' \geq 0, i \in \{0,1\}} & u(c) + \beta \mathbb{E} V(y', s') \\ & \text{s.t.} & c + a' = y - ip(s), \\ & & y' = \max\{(1 + r)a' - [1 - \lambda(qx, m)i]qx, 0\} + (1 - \tau)wz'\varepsilon_{j+1}. \end{cases} \tag{1.1}$$

The expectation is described as the product of $\Gamma^i_{ss'}$ and $\Pi^i_{s,x}$ where

$$\begin{aligned} \Gamma^i_{ss'} &= \Gamma_h(h'|h, j, i)\Gamma_z(z'|z, h')\Gamma_m(m'|m, z', h')\Gamma_g(g'|g, z'), \\ \Pi^i_{s,x} &= \Pi(x|h, j, i). \end{aligned}$$

Note that the insurance premium is a function of the state variable, $p : \mathbb{S} \to \mathbb{R}_+$.

After-retirement Problem

After the retirement age J_R, households get Social Security benefit ss and are covered by Medicare. They die at age J and there is no bequest. Then the recursive formulation of the after-retirement problem is given by

$$\begin{cases} V(y,h,j) = & \max_{c\geq 0, a'\geq 0} & u(c) + \beta \mathbb{E} V(y', h', j+1) \\ & \text{s.t.} & c + a' = y, \\ & & y' = \max\{(1+r)a' - [1-\lambda(qx)]qx, 0\} + ss. \end{cases} \tag{1.2}$$

with $V(\cdot,\cdot,J+1) = 0$. The expectation is described as the product of $\Gamma_h(h'|h,j)$ and $\Pi(x|h,j)$[5]. Also the dependence of the reimburse schedule λ on m is suppressed because all the retired people are covered by Medicare.

1.2.3 Medical Service Sector

I assume a competitive medical service sector and hence zero profits. The medical service technology transforms one unit of the composite good into one unit of medical services.

U.S. federal law requires hospitals to treat individuals in urgent need of medical attention, regardless of ability to pay (Emergency Medical Treatment and Active Labor Act). This free rider problem of limited liability leads to uncompensated care costs to the hospital, and to make up these costs, hospital charge those who can pay more for the services provided (Gruber (2008)). However, this cost shift is borne almost exclusively by private insurance providers through the medical services market (Kathleen Stoll and Kim Bailey (2009)). This is because the Medicare and Medicaid programs use restrictive and inflexible reimbursement schedules. I therefore assume that the medical services market for Medicaid and Medicare holders is separated from that for private insurance holders and the uninsured, and the uncompensated amounts in those markets are paid for by the government through the Medicaid and Medicare Disproportionate Share Hospital payments. See 1.2.6 for details.

Hospitals charge a mark-up $q \geq 1$ for the supply of medical services x.

[5]These are degenerated (conditional) distributions of $\Gamma_h(h'|h,j,i)$ and $\Pi(x|h,j,i)$ when $i = 1$.

Hospital Revenue

Suppose a patient with resources $(1+r)a'$ is charged the amount qx. Suppose further he is uninsured. He pays the entire bill if he can, and if not, he just pays as much as possible. Thus the hospital revenue from the treatment of this uninsured agent is

$$\min\{(1+r)a', qx\}.$$

Now suppose he has insurance. Then again he pays the out-of-pocket expenditure if he can, and otherwise, he just pays as much as possible. The insurer pays the amount specified in the plan, $\lambda(qx)qx$, where the dependence of the reimbursement schedule λ on m is suppressed because here I am considering the private insurance provider only, i.e., $m = 0$. Thus the hospital revenue from the treatment of this insured agent is

$$\begin{aligned} & \underbrace{\min\{(1+r)a', [1-\lambda(qx)]qx\}}_{\text{Payment by the patient}} + \underbrace{\lambda(qx)qx}_{\text{Payment by the insurer}} \\ = \quad & \min\{(1+r)a' + \lambda(qx)qx, qx\}. \end{aligned}$$

Therefore, in sum, the hospital revenue from the treatment of a patient is

$$\begin{aligned} & \underbrace{(1-i)\min\{(1+r)a', qx\}}_{\text{Uninsured patient}} + \underbrace{i\min\{(1+r)a' + \lambda(qx)qx, qx\}}_{\text{Insured patient}} \\ = \quad & \min\{(1+r)a' + i\lambda(qx)qx, qx\}. \end{aligned}$$

Hospital's Break Even

The hospital's break even condition balances the cost and benefit of the medical services. The break even within the medical services market of the working age population without public health insurance coverage becomes

$$\underbrace{\int \mathbb{E}_x \min\{(1+r)a' + i\lambda(qx)qx, qx\} d\mu(m=0, j < J_R)}_{\text{Revenue}} = \underbrace{\int \mathbb{E}_x x d\mu(m=0, j < J_R)}_{\text{Cost}}, \tag{1.3}$$

where μ denotes the probability measure over the measurable space defined by the state space $\mathbb{R}_+ \times \mathbb{S}$ (see 1.2.7 for the formal definition). Note that the mark-up q is an equilibrium object pinned down by this condition.

1.2.4 Insurance Company

I assume a competitive private insurance market and hence zero profits. The markets for group insurance and nongroup insurance are segmented, so there is no cross subsidy. Households' states are public information[6].

Issuing an insurance plan incurs fixed costs ϕ that capture the combined effects of administrative and screening costs. The premium is a function of age j and health status h, and is given by

$$p(j, h) = (1 + r)^{-1}\mathbb{E}_x[\lambda(qx)qx|j, h] + \phi. \tag{1.4}$$

Note that the insurance premium is actuarially unfair due to the fixed costs ϕ. The dependence of the reimbursement schedule λ on m is suppressed (i.e., $m = 0$).

When households have a employer-provided group insurance offer, $g = 1$, the employer fully contributes the premium, and hence agents take up the employer-provided group insurance as long as it is available. This assumption is made because the employer's decision about its contribution is out of the scope of this paper, but is also innocuous since in the data almost all employees take up the offer whenever they are offered[7].

The premium collected by the insurance companies is saved and used as a capital by the production sector.

[6] In the data, as opposed to the implication of the standard adverse selection literature, healthy people are more likely to be insured than unhealthy people. There are several possibilities to account for this fact. One interpretation is that the insurance companies spend costs ϕ (or at least a part of ϕ) and successfully verify agents' health status, which is the assumption in the present paper On the other hand, according to the advantageous selection, people who have strong tastes for being healthy might be at the same time more risk averse (Cutler, Finkelstein, and McGarry (2008)). This paper does not dig into this issue, and further assumes that the fixed costs ϕ are policy-invariant. Separating the screening costs from the other administrative costs in ϕ using micro-data is my future work.

[7] Brgemann and Manovskii (2010) study the employers' decision of providing group insurance to their employees.

1.2.5 Firm

The aggregate production technology is given by a Cobb-Douglas function, $F(K, L) = AK^\theta L^{1-\theta}$. I assume that the firm pays the group insurance premium for the employees who have the employer-provided insurance, but the assignment of the employer-provided insurance is unknown when the firm is choosing its labor input L. Denote by η the expected marginal payment of employer's group insurance premium per efficiency unit. Then, the firm's problem is given by

$$\max_{K,L} \; AK^\theta L^{1-\theta} - wL - (r+\delta)K - \eta L$$

where the last term is the employer's premium contribution. Then by zero profit condition, the factor prices are given by

$$\begin{aligned} r &= F_K(K, L) - \delta, \\ w &= F_L(K, L) - \eta \\ &= F_L(K, L) - \frac{\int p d\mu(g = 1, j < J_R)}{L}, \end{aligned} \tag{1.5}$$

where the last equality comes from the fact that the marginal employer's premium payment per efficiency unit is simply given by the total premium payment divided by the total labor input.

1.2.6 Government

The government imposes a proportional tax τ on the labor income of the working age population in order to finance the Social Security benefit. The tax revenue is also used to finance Medicaid and Medicare Disproportionate Share Hospital payments that compensate the deficit of the Medicaid and Medicare providers. Then, assuming the balanced budget for each period, the government budget constraint is given by

$$\begin{aligned} \tau \int wz\varepsilon d\mu(j < J_R) = & \int ss d\mu(j \geq J_R) \\ + & \int \mathbb{E}_x \left[x - \min\{(1+r)a' + \lambda(qx, 1)qx, qx\} \right] d\mu(m = 1, j < J_R) \\ + & \int \mathbb{E}_x \left[x - \min\{(1+r)a' + \lambda(qx, m)qx, qx\} \right] d\mu(j \geq J_R), \end{aligned} \tag{1.6}$$

where the second and third lines represent the net cost of the Medicaid and the Medicare, respectively.

1.2.7 Stationary Equilibrium

Define the state space as $\Omega \equiv \mathbb{R}_+ \times \mathbb{S}$ and the sigma algebra on Ω as Σ_Ω. Denote the probability measure over the measurable space (Ω, Σ_Ω) by μ. I then formally define the stationary equilibrium of this economy.

Definition 1 *A stationary equilibrium of this economy is a set of policy functions* $\{c, a', i | c : \Omega \to \mathbb{R}_+,\ a' : \Omega \to \mathbb{R}_+,\ i : \Omega \to \{0,1\}\}$, *a value function* $V : \Omega \to \mathbb{R}$, *market prices* $\{w, r, p | w \in \mathbb{R}_+,\ r \in \mathbb{R}_+,\ p : \mathbb{S} \to \mathbb{R}_+\}$, *a mark-up in the medical services market* $q \in \mathbb{R}_+$, *government policies* $\{\tau, ss | \tau \in \mathbb{R}_+,\ ss \in \mathbb{R}_+\}$ *and a stationary distribution* μ *such that*

(i) Given prices, $\{c, a', i\}$ *solves the households' problem (1.1) and (1.2), and achieves* V.

(ii) q *satisfies the hospitals' zero profit condition (1.3).*

(iii) p *satisfies the insurance company's zero profit condition (1.4).*

(iv) $\{w, r\}$ *satisfies the firms' marginal profit conditions (1.5).*

(v) $\{\tau, ss\}$ *satisfies the government budget constraint (1.6).*

(vi) All markets clear:

$$
\begin{aligned}
K &= \int a' d\mu + \int ip(s) d\mu(m = 0, j < J_R), \\
L &= \int z\varepsilon d\mu(j < J_R), \\
F(K, L) &= \int c d\mu + \int x d\mu + \delta K.
\end{aligned}
$$

(vii) The distribution μ *is stationary.*

1.2.8 Policy Reform

The Affordable Care Act comprises two key provisions to increase the coverage rate in the individual insurance market.

Community Rating and Guaranteed Issue

Community rating and guaranteed issue prohibit insurers from price-discriminating or rejecting against applicants based on health status. Therefore, after the reform, the insurance premium is no longer a function of h :

$$\tilde{p}(j) = (1+\tilde{r})^{-1} \frac{\int i \mathbb{E}_x[\lambda(\tilde{q}x)\tilde{q}x|j,h] d\tilde{\mu}(j)}{\int i d\tilde{\mu}(j)} + \phi$$

Here I put the tildes on the objects that are different from those of the pre-reform economy. Note that the premium is an equilibrium object that depends on the individual decision i. This is interpreted as a fixed point problem for each age j, hence J_R fixed point problems. Note also that the premium depends on the new mark-up $\tilde{q}$ derived from the hospitals' break even constraint, which is different from q because the individual insurance decision changes.

Insurance Mandate and Insurance Exchanges

The other key provision is the individual mandate which penalizes those without health insurance. The size of the penalty is based on the income and given by $695 or 2.5% of the income, whichever is greater.

The ACA also establishes health insurance exchanges that provide insurance premium subsidies for individuals with low income. Specifically, the insurance exchanges are going to provide the premium subsidy for those with income from 100% to 400% of the federal poverty line (FPL). The ACA does not specify the premium subsidies for those with income under 100% of FPL, as it originally planned to expand Medicaid to all the people with income up to 133% of FPL. However, because in 2012 the Supreme Court of the United States ruled that participation in the Medicaid expansion is voluntary for states, the treatment of poor individuals who are not eligible for Medicaid after the reform is unclear. Thus, I assume that the premium subsidy is extended to those with income below 100% of FPL, and apply the same premium cap of those with 100% of FPL (i.e., 2% of income) to those population. Figure 1.1 displays this extended subsidy schedule. I assume this subsidy is financed by a proportional income tax, and hence the tax rate changes to $\tilde{\tau}$.

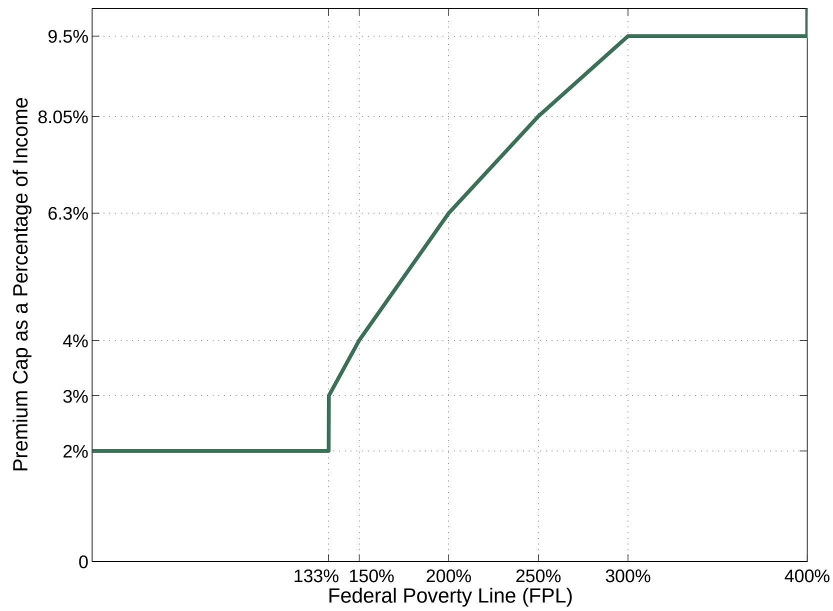


Figure 1.1: Premium Cap as a Percentage of Income, by Federal Poverty Line (FPL). Note the subsidy is extended to 0-100% of FPL (see the text). Source: Congressional Research Service (2010), "Private Health Insurance Provisions in PPACA"

Working Age Household's Budget Constraint after the Reform

The budget constraint of the working age household can be written by

$$c + a' = y - i \min\{\tilde{p}(j), pcap(\tilde{w}z\varepsilon)\} - (1 - i) \max\{\$695, 2.5\% \times \tilde{w}z\varepsilon\},$$

where *pcap* is the premium cap illustrated in Figure 1.1.

1.3 Data and Estimation

This section describes the data I use and parametrization of the model.

1.3.1 The data

I use data from the Survey of Income and Program Participation (SIPP), and the Medical Expenditure Panel Survey (MEPS). These continuous series of national panels provide information about health insurance and medical expenditures of the U.S. households. I use the 2001 panel and the

2004 panel of the SIPP, and the five panels of the MEPS from 1999 to 2004. Taking into account the fact that some states such as Massachusetts have introduced drastic policy reforms after 2006, I assume the U.S. economy is in a steady state prior to 2006.

I use the SIPP as the main data set since it provides key information about health insurance, self-assessed health status and wealth. I do not use panels before 1996, because the key question asking the source of the health insurance was significantly different from the panels after 2001.

I also use the MEPS as it provides information about medical expenditures and medical charges. To the best of my knowledge, no other national data provide this information. I match the information in the MEPS with that in the SIPP using self-reported health status whose summary statistics are comparable as is shown in Table 1.1.

The SIPP consists of core wave files and topical module files. Respondents are surveyed every 4 months. In the core wave files, for many variables such as income and health insurance coverage they are questioned about each of the previous 4 months since the last interview wave. Some of the topical module files report key variables for this analysis such as net worth and self-reported health status. I merge all the core waves of each panel with the topical module files that report this key information. Specifically, for the 2001 panel I use all the core waves and the wave 3, 6 and 9 of the topical module files, whereas for the 2004 panel I use all the core waves and the wave 3 and 6 of the topical module files. I also use the medical expenditure and charge data from the MEPS.

Samples

The decision making unit in the model is the Health Insurance Eligibility Unit (HIEU), which is different from the Census definition of a family or a household. The basic concept of this subfamily relationship unit is that an HIEU comprises individuals in the household who are eligible for one health insurance policy. The policyholder, policyholder's spouse, and their unmarried minor children are typically considered an HIEU. Thus, one household can potentially consist of multiple HIEUs. For the general definition of the HIEU, see State Health Access Data Assistance Center (2012)[8].

I use the HIEU identifiers provided by the MEPS. However, unlike the MEPS, the SIPP does not provide identifiers for HIEUs. Therefore I construct HIEUs for the SIPP samples from other variables regarding family relationship, following the definition provided by State Health Access Data Assistance Center (2012). To be consistent with the HIEU definition in the MEPS data, I also

[8] They use the term "Health Insurance Unit (HIU)" instead of the Health Insurance Eligibility Unit (HIEU).

Table 1.1: Summary Statistics of the Working Age Population in the SIPP and the MEPS

	Obs.*	Mean	Median	Std. Dev.
SIPP				
Age	64417	44	44	11
Health Score**	63507	2.27	2	1.08
Income***	64310	$33,629	$25,865	$35,246
Wealth***	62243	$180,838	$95,725	$295,423
MEPS				
Age	30673	43	43	11
Health Score**	30573	2.29	2	1.10
Medical Expenditure***	30673	$2,689	$667	$9,066

* Pooled sample of heads of HIEU of age 25-64
** Health score is calculated from the answers to the subjective health status question that range from 1 (excellent) to 5 (poor).
*** Amounts are in 2001 dollars.

include children under age 24 who are full-time students as eligible children.

A head of a HIEU is defined as the single adult for the HIEUs of single adults and the male adult for the HIEUs of married couples. The sample consists of the heads of the HIEUs whose age is from 25 to 80[9].

The sample weights provided by the SIPP and the MEPS are then rescaled proportionally so that the sum of the weights in each panel is equal to the number of the observations of the panel. All the dollar values are converted into 2001 dollars.

Summary Statistics

Table 1.1 provides the summary statistics of the working age population of the SIPP and the MEPS. Note that the statistics of the self-reported health status, which I use to link the two data sets, are similar.

1.3.2 Calibration

This section constructs stochastic processes and parameters I use in the numerical analysis.

[9] I do not use samples that have a sample weight of zero.

Demographics

One model period corresponds to one year. The decision making unit in the model which I call as a household is the HIEU described above. I assume households enter the economy at age 25, get retired at age 65, and die at the end of age 80.

Health Status

I use self-reported health status in the SIPP as the measure of health[10]. The SIPP asks respondents' general health in the topical module files, and the respondents report their health as being "excellent", "very good", "good", "fair", or "poor" every year. Given the fact that the median health status of the sample is "very good", I define healthy adults as those who report their health status as "excellent" or "very good", and unhealthy adults as the complement[11]. In the same way, using self-reported health status in the MEPS, I also construct a health status variable for each individual for the MEPS samples.

I compute the transition probability of health status for each age and insurance status. Specifically, I first construct 10-year age bins from 25-34 to 55-64 for each panel. Then for each bin, I compute the conditional probability of being healthy or unhealthy in the next period given today's health and insurance status. When doing this, I limit my sample to those who are not disabled in order to capture the effect of the primary care on the health. I then take the weighted average of all the transition probabilities over the panels using the relative weights. Finally, I interpolate the conditional probabilities along the age in order to obtain smooth transition matrix for each age.

Figure 1.2 displays the conditional probability of being healthy next period given health and insurance status today. As is seen in the Figure, the insured agents who have access to primary care have higher probability of being healthy in the next period than the uninsured for both the healthy and the unhealthy.

For the elderly, I construct 5-year age bins from 65-69 to 75-79 for each panel. I then use the same procedure as the estimation for the working age population described above, but for old households, the transition does not depend on the insurance status, because all of them are covered

[10] Self-reported health status is widely used as the measure of health in the literature, e.g. Clemens (2012), Imrohoroglu and Kitao (2012), and Yogo (2012). For example, although he uses different survey data (i.e., the Health Retirement Survey), Yogo (2012) sets self-reported health status as the primary measure of health in his study about the elderly. He argues that "self-reported general health status is highly correlated with doctor-diagnosed health problems, difficulty with activities of daily living, health care utilization, and future mortality."

[11] Clemens (2012) uses the similar definition of healthy/unhealthy agents in his analysis using the Current Population Survey.

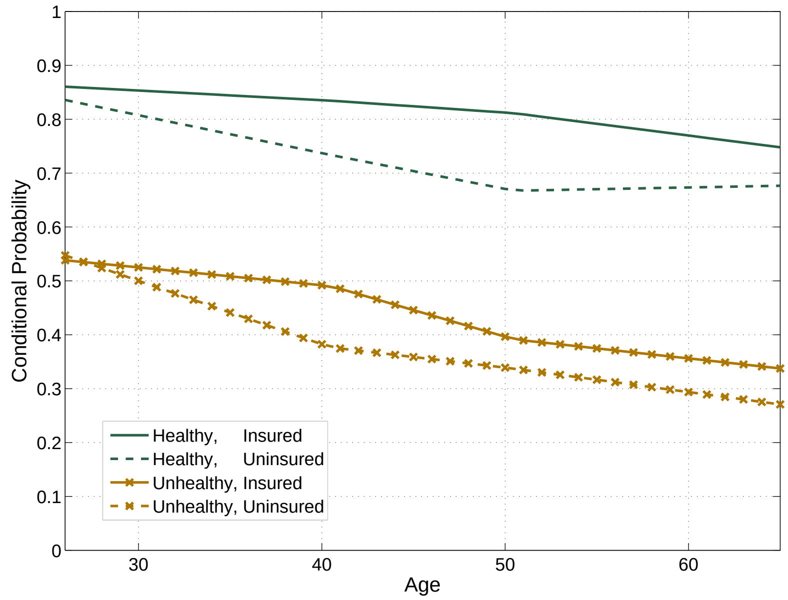


Figure 1.2: Conditional Probability of Being Healthy Next Period for Each Health and Insurance Status from the SIPP

by Medicare.

Earnings

Households' endowment is a product of age-dependent productivity ε, idiosyncratic earnings shock z, and the market wage rate w that is equal to the marginal product of labor. I calibrate ε and z.

For the idiosyncratic earnings shock, I use non-asset income of the heads of the HIEUs in the data. The definition of non-asset income follows Glover, Heathcote, Krueger, and Ros-Rull (2012), and includes wage and salary income, two thirds of business or farm income, social security income, and a variety of public and private transfers such as unemployment compensation and money from relatives. To construct annual income, I sum up all these variables for 12 months in the core wave files of the SIPP.

The basic calibration strategy of the earnings process is similar to Jeske and Kitao (2009), but in this study, the evolution of the earnings depends also on health status, as described in Section 1.2.1. Given the data of households' annual income, I create five income classes with equal size of population for each year. Using these five states, I then compute the conditional transition probabilities separately for each health status for each panel. Specifically, given h', the transition probabilities are directly calculated from the panel data and I obtain $\Gamma_z(z'|z,h')$ for each panel. Finally, I take the weighted average of these transition probabilities and the income grids over the panels using the relative weights. The associated income grids are normalized. The resulting earnings grid space is $\mathbb{Z} \equiv \{0.14, 0.46, 0.77, 1.18, 2.45\}$ relative to the average annual income of $33,575 in 2001 dollars.

The age-dependent productivity is taken from Hansen (1993) as in Imrohoroglu and Kitao (2012).

Public Health Insurance

I define that one is covered by public health insurance if she is covered by Medicaid, Medicare, or other public insurance such as TRICARE for more than 6 months of the year. For more detailed definition of insurance, see section 1.3.3.

Given the data of public insurance status, I compute the conditional transition probabilities separately for each level of tomorrow's earnings and health status for each panel. Specifically, given z' and h', the transition probabilities are directly calculated from the panel data and I obtain $\Gamma_m(m'|m,z',h')$ for each panel. I then take the weighted average of these transition probabilities over the panels using the relative weights.

Employer-Provided Group Health Insurance

As described in Section 1.2.1, because the actual take-up rate of the employer-provided group health insurance in the data is above 95% when it is available, I assume that households obtain group insurance as long as it is offered from the employer. Hence, I use the group insurance status in the data as a proxy of the group insurance offer.

I define that one is covered by employer-provided group insurance if she is covered by a health plan through her current or prior employer, or union for more than 6 months of the year. For more detailed definition of insurance, see Section 1.3.3.

Given the data of employer-provided group insurance status, I compute the conditional transition probabilities separately for each level of tomorrow's earnings for each panel. Specifically, given z', the transition probabilities are directly calculated from the panel data and I obtain $\Gamma_g(g'|g,z')$ for each panel. Finally, I take the weighted average of these transition probabilities over the panels using the relative weights.

Medical Expenditure

It is important to note that what is observed in the data is not the medical resource cost x but the medical expenditures qx.

I estimate the medical expenditure shock separately for the insured and the uninsured. This is because the medical spending of the insured is substantially higher than that of the uninsured. The insured have access to primary care, and receive medical care to maintain their health status. On the other hand, the uninsured mainly receive urgent care, and they often delay non-urgent medical care because it is very expensive.

I use the variables of total medical expenditures and charges in the MEPS. The variable of total medical expenditures is sum of the amounts actually paid for each medical event, and does not include the amount unpaid due to medical bankruptcy. On the other hand, the variable of total medical charges is sum of the amounts charged by the medical service providers, and includes the amounts unpaid due to medical bankruptcy. The former is suitable for the medical shocks for the insured, and the latter is suitable for those for the uninsured[12].

For each panel, I construct 10-year age bins from 25-34 to 55-64 for each health and insurance status. For each age bin, I create five states for the expenditures/charges with the bins of size of 25%, 25%, 25%, 20%, 5% population separately for each year. For example, the first state means the expenditures of the lowest 25%, and the second state means the next 25% and so on. The top bin is intended to capture the catastrophic events. Then I compute the weighted average of each bin and normalize the values in 2001 dollars. Next, I take the weighted average of the values over the panels for each health and insurance status using the relative weights. I approximate these average medical expenditures of each bin by second order polynomial functions of ages, separately for each health and insurance status. Figure 1.3 displays these average medical expenditures and the fitted values. The amounts are higher for the insured, while the catastrophic events are equally severe for both the insured and the uninsured.

For the elderly, I construct 5-year age bins from 65-69 to 75-79 for each panel. I then use the same procedure as the estimation for the working age population described above, but for old households, the expenditure distribution does not depend on the insurance status, because all of

[12] Jeske and Kitao (2009) and Pashchenko and Porapakkarm (forthcoming) use the medical expenditure data. This might underestimates the size of the shocks for the uninsured, since the variable largely reflects the insurance discount that is unavailable for the uninsured. On the other hand, they do not distinguish the medical usage of the insured from that of the uninsured. However, the medical spending of the insured is substantially higher than that of the uninsured. Thus, the size of the shocks for the uninsured might be overestimated.

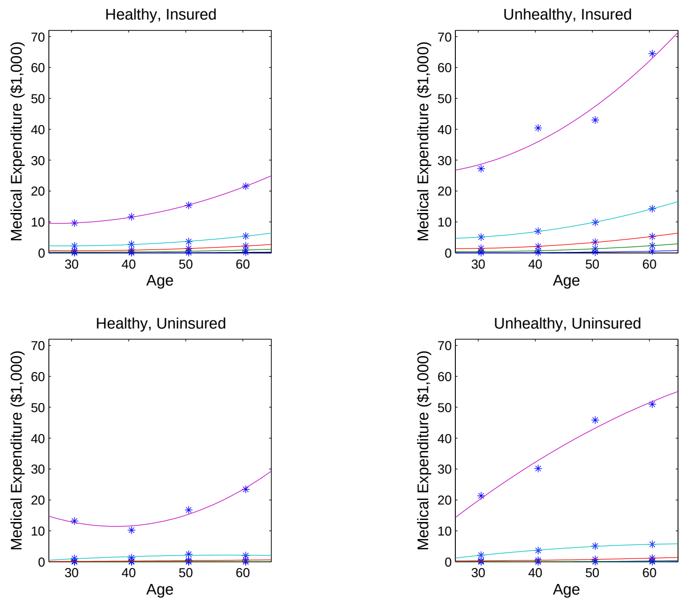


Figure 1.3: Distribution of Medical Expenditures of the Working Age Population by Insurance and Health Status from the MEPS. Each figure displays the average medical expenditures for each 10-year age bin (stars) and its approximation (lines) of each expenditure group ranging from the bottom 25% to the top 5% (see the text for the detailed information).

them are covered by Medicare.

Insurance Reimbursement Schedule

The model abstracts from the detail of insurance plans such as deductible and copayment. Instead, following Jeske and Kitao (2009), I estimate the reimbursement schedule of health insurance for private insurance, Medicaid and other public insurance, and Medicare separately. Specifically, first I compute the reimbursed amounts out of medical expenditures for each agent using MEPS. Then I run regressions for each insurance type of the form

$$\log(oop) = \beta_0 + \beta_1 \log(ex) + \beta_2 (\log(ex))^2,$$

where ex is the total medical expenditure, and oop is the out-of-pocket expenditure that is not reimbursed by the insurance, i.e., ex less the reimbursed amounts[13]. Here, I limit the sample to

[13] I do not use the reimbursed amounts directly as a dependent variable because people sometimes report the reimbursed amounts of zero and hence I cannot take the logarithm.

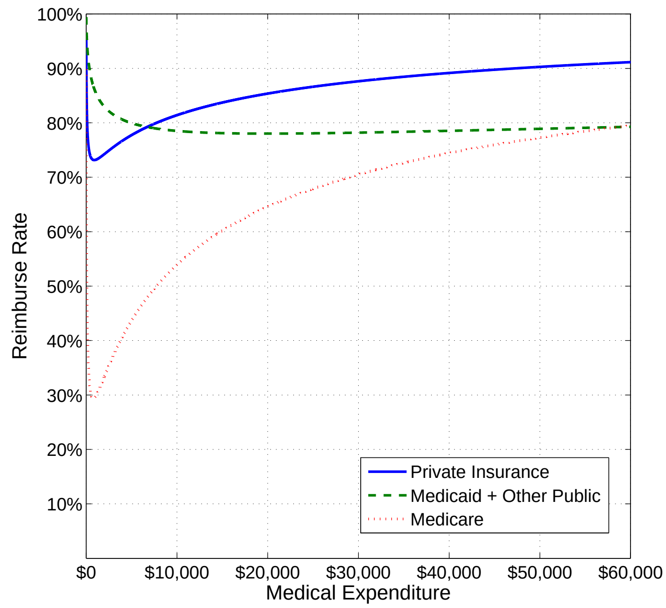


Figure 1.4: Estimated Insurance Reimbursement Schedule from the MEPS

those with positive expenditures. The weighted least squares estimation finds coefficient estimates that are all significant at the 1% level[14]. Using the resulting coefficients, I compute the reimburse schedule, which is displayed in Figure 1.4. It shows that public insurance is more generous for small amounts of medical expenditures than private plans, whereas the opposite is true for large amounts of medical expenditures. Medicare is not as generous as Medicaid.

Preference

The discount factor β is chosen so that the capital-output ratio of the economy is equal to three. I use the standard CRRA utility function of the form $\frac{c^{1-\gamma}}{1-\gamma}$. The estimation for γ is described in Section 1.3.3.

Technology

I use a constant-return-to-scale Cobb-Douglas production technology $F(K,L) = AK^{\theta}L^{1-\theta}$ with the capital share of $\theta = 0.33$. The TFP A is calibrated so that the average income is equal to one. The

[14]The R-squares are 0.558 for private insurance, 0.501 for Medicaid and other public insurance, and 0.788 for Medicare, respectively.

Table 1.2: Parameters of the Model

Remark	Parameter	Value	Target
Max Age	J	55	Live from age 25 to 80
Capital Share	θ	0.33	-
SS Replacement	ss	0.45	45% of Ave. Income
Discount Factor	β	0.958	Capital-output Ratio: 3
TFP	A	0.965	Ave. Income = 1
Depreciation	δ	0.082	Interest Rate: 3%

capital depreciation is set so that the real interest rate is equal to 3%. Note that after the policy reform, the average income and interest rate will change because of the general equilibrium effects.

Social Security

The Social Security replacement rate is set to 45% of the average income as in Jeske and Kitao (2009). Although this rate might be slightly higher than the values usually used in the literature, this is reasonable given that the population on which this study focuses, the active participants of the individual health insurance market, has substantially lower income on average through their working age[15]. The proportional income tax rate is calibrated so that the government's budget is balanced.

Calibrated Parameters

Table 1.2 displays the summary of the calibrated parameters. The top panel shows the parameters exogenously given and the bottom panel shows the parameters calibrated inside the model.

1.3.3 Estimation

This section describes empirical moments with which I match the model, and the estimation strategy. Before explaining the estimation procedure, I first illustrate the data construction of households' net worth and insurance status. I then describe the estimation.

[15] Pashchenko and Porapakkarm (forthcoming) use 30% and 40% for each people with low and high education, respectively.

Wealth

I construct households' wealth using detailed data about assets and liabilities in the SIPP, following the U.S. Census Bureau (2008), "Net Worth and the Assets of Households: 2002". To briefly explain, the net worth is defined as the sum of the assets such as interest-earning assets, stocks, bonds and properties owned, subtracted by the sum of the liabilities such as mortgages, debt, and loans. To construct the net worth of heads of HIEUs, I use variables in the SIPP describing the entitlements for each property. For some asset and liability variables for which the amounts are already aggregated within the household, I split the amounts by the fraction calculated from the heads' income.

There are two reasons why it is desirable to construct households' wealth from each component of assets and liabilities in the data rather than simply using the households' total net worth provided by the SIPP. One is that the decision making units in the model are not households but HIEUs, and there is no systematic way to allocate total net worth of a household to multiple HIEUs that form the household together. The other is that the variable for households' total net worth in the SIPP is top-coded by itself. Hence direct utilization of the variables about assets and liabilities in the SIPP, even though some are also top-coded, provides more precise information about households' wealth distribution.

Insurance Status

For each month, the SIPP provides the information about insurance sources in addition to the insurance status for each individual. I define agents to be insured if they have insurance for more than 6 months of the year from any source. The uninsured are the complement.

One is insured by public insurance if she is covered by Medicaid, Medicare, or other public plans for specific populations such as TRICARE for more than 6 months. I define that one is covered by employer-provided group insurance if she is not covered by the public plan and is covered by a health plan through her current or prior employer, or union for more than 6 months of the year[16]. Finally, one is covered by individual insurance, if she is not covered by the public plan and directly purchases a health plan for more than 6 months of the year. When one is insured but no single source covers her for more than 6 months, then I choose the one with the longest length of the coverage as the source.

[16] By the Consolidated Omnibus Budget Reconciliation Act (COBRA), employees can continue health insurance coverage after leaving employment (for up to 18 months in most cases).

Table 1.3: Estimated Parameters

Remark	Parameter	Value	Target
Risk Aversion	γ	1.234	Joint dist. of insurance coverage
Fixed Costs of Insurance	ϕ	\$803	Joint dist. of insurance coverage

Estimation

There are two key parameters for the insurance choice in the model, the risk aversion γ and the fixed costs of issuing insurance ϕ, i.e., the extent to which health insurance is priced above its actuarially fair values. Intuitively, the higher is the risk aversion, the more insured are people. Similarly, the lower are the fixed costs, the more insured are people.

To capture the insurance coverage distribution in the individual health insurance market, I target the joint distribution of insurance coverage of active participants of the individual health insurance market. Specifically, I use the insurance coverage rate of those with neither public insurance nor employer-provided insurance for each of three income classes ($z \times 3$), three wealth classes ($a \times 3$), two age classes ($j \times 2$), and two health status ($h \times 2$). With 36 moments in total, I solve

$$\min_{\gamma,\phi} \sum \alpha_{z,a,j,h}[i_{Data}(z,a,j,h) - i_{Model}(z,a,j,h)]^2$$

where i_{Data} and i_{Model} are the insurance coverage rate of each subpopulation in the data and the model, respectively, and α is the population weights of each subpopulation taken from the SIPP data.

Table 1.3 provides the estimated parameters. The estimated risk aversion is in the range of the standard models, but relatively low. Also the fixed costs of insurance are high in that it constitutes of 26% of the average premium. Given that the very low rate of coverage among the active participants (i.e., 23.5%), these seem to be reasonable. In the next section, I further conduct a decomposition exercise in order to investigate the driving force of the low insured rate.

1.4 Quantitative Analysis

This section provides the results of the quantitative analysis. First, I describe how the model fits the data. I compare the model and the data regarding the insurance coverage and the wealth and income distribution of the uninsured. Second, I conduct a decomposition exercise in which I assess the relative importance for explaining the high uninsured rate. Finally, I present the implications of the health care reform.

1.4.1 Model Performance

This section describes how the model fits the data. First, I look at the insurance status of the working age population. Then I focus on the active participants of the individual health insurance market, those who are neither eligible for public health insurance nor offered employer-provided insurance, because those are the target population of the reform. Finally I compare the wealth and income distribution of the uninsured in the model with those in the data.

Insurance Coverage

The aggregate insurance rate of the working age population is presented in Table 1.4. The top panel shows the insurance status of all the working age population. The model fits the data well, although it overestimates the number of the active participants. This results in the higher uninsured rate in the model than that in the data. The bottom panel shows the insurance status of the active participants, the target population of the policy reform. This population is the main focus both of this study and the upcoming reform, and the model captures the fraction of the insured among this population.

Next, I construct subgroups from these active participants. Specifically, I decompose the population by wealth, income, health status and age. Figure 1.5 plots the uninsured rate for each wealth and income quintile of the model and the data, and shows that the model replicates the insurance pattern reasonably well. In the bottom panel, more than 20% people with the income of the 1st quintile, i.e., very low income households, are insured in the data. The model cannot capture this pattern because those are typically rich individuals retired earlier than age 65, yet the model assumes that everyone gets retired all at once at age 65.

The top panel of Table 1.5 is the decomposition by age and health status, and displays the

Table 1.4: Insurance Status (Data vs Model)

Insurance Status	Data	Model
All Working Age		
Individual	5.0%	5.9%
Uninsured	16.2%	19.7%
Employer-based	66.3%	60.3%
Public	12.5%	14.1%
Active Participants*		
Insured	23.5%	22.9%
Uninsured	76.5%	77.1%

* "Active participants" are those who are neither eligible for public health insurance nor offered employer-provided insurance.

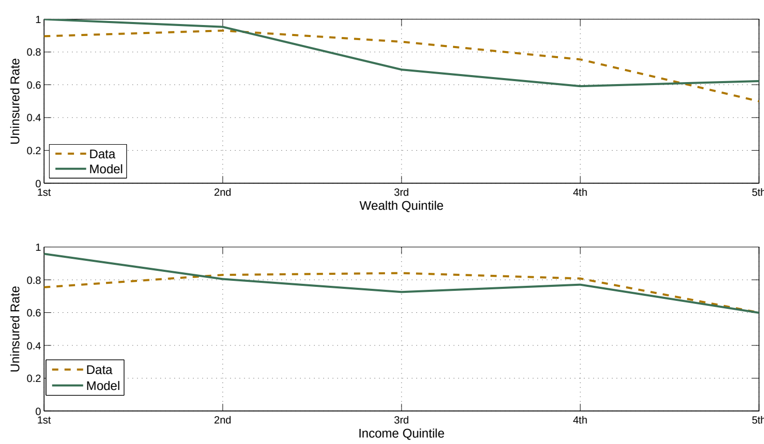


Figure 1.5: Uninsured Rate of the Active Participants for Each Quintile of Wealth and Income (Data vs Model)

Table 1.5: Insurance Coverage Pattern of the Active Participants (Data vs Model)

Statistics	Data	Model
Uninsured Rate		
Age 25-44	82.3%	81.4%
Age 45-64	67.9%	72.4%
Unhealthy	83.4%	88.4%
Healthy	71.1%	70.7%
Correlation of Insurance Status		
$Corr(insurance, wealth)$	0.396	0.261
$Corr(insurance, income)$	0.208	0.259
$Corr(insurance, age)$	0.196	0.089
$Corr(insurance, health)$	0.143	0.202

uninsured rate of the young (age 25-44), the old (age 45-64), and the healthy and the unhealthy. The model overestimates the uninsured rate of the young population, but the overall insurance pattern is similar. The bottom panel shows the correlation of the insurance status with wealth, income, age and health. Those correlations are all positive both in the data and the model.

Characteristics of the Uninsured

Covering the uninsured population is one of the main goals of the health care reform. Therefore, it is important to look at the characteristics of the uninsured. While the model does not target those moments, it replicates the income and wealth distribution of the uninsured fairly well, as displayed in Table 1.6. There is a large skewness in the wealth distribution among the uninsured in the data, which is only partially captured in the present model. Note that the uninsured population includes many low income households, because they are not necessarily eligible for Medicaid. The model replicates this important margin, and the reform has a significant implication for these uninsured poor households.

1.4.2 Source of Low Insured Rate

This section conducts a decomposition exercise. I assess the relative importance in accounting for low current insured rates, 23.5%, of (i) high fixed costs in the insurance market, (ii) low risk aversion

Table 1.6: Income and Wealth Distribution of the Uninsured (Data vs Model, 2001 dollars)

Statistics	Data	Model
Income Percentile		
25%	5,720	3,853
50%	12,792	12,068
75%	19,832	20,127
Wealth Percentile		
25%	0	0
50%	6,027	13,137
75%	71,273	79,286

within the pool of potential buyers of insurance, and (iii) limited liability in the health services market. Through the exercise, I fix all the other parameters but the parameter under consideration and the prices. Hence this is a general equilibrium analysis.

First, I change the fixed costs of issuing insurance in the baseline economy. I gradually increase the parameter ϕ from zero to twice as much as the baseline.

The top-left panel in Figure 1.6 displays the result. Specifically, it plots the change in the insured rate of the active participants when the fixed costs of insurance change. The vertical line corresponds to the estimated value in the baseline model. The figure shows that the insured rate is decreasing in the size of the fixed costs, and people are very sensitive to how actuarially unfair the premium is. If there is no cost ($\phi = 0$) and hence the premiums are actuarially fair values, then more than 50% of the population purchase the individual insurance. On the other hand, if the fixed costs are increased so that it constitutes more than 70% of the average premium, then almost nobody would buy the individual insurance.

Second, I change the risk aversion in the baseline economy. I gradually increase γ from almost risk neutral $\gamma = 0.1$ to $\gamma = 5$, a relatively higher value in the standard models.

The top-right panel in Figure 1.6 displays the result. It plots the change in the insured rate of the active participants when the risk aversion changes, and the vertical line corresponds to the estimated value in the baseline model. The figure shows that the insured rate is increasing in the risk aversion. With high risk aversion, households have incentive to accumulate a lot of assets, yet they also have incentive to hold insurance. On the other hand, some households buy insurance even when they are almost risk neutral, i.e., the risk aversion is close to zero. The reason is because having insurance gives people to access to primary care, and in some cases people are able to reduce the expected medical expenditure by having insurance since the primary care improves the health status.

Third, I change the condition of the limited liability in the baseline economy. Previously, as stated in the U.S. law, households' income was secured even when filing medical bankruptcy. Now, I consider a situation in which a fraction $\alpha \in [0, 1)$ of their income could be confiscated. The cash-in-hand after being hit by the medical shock y' is then given by

$$\begin{aligned} y' &= \max\{(1+r)a' - (1-\lambda(qx)i)qx + \alpha(1-\tau)wz'\varepsilon', 0\} + (1-\alpha)(1-\tau)wz'\varepsilon' \\ &= \max\{(1+r)a' - (1-\lambda(qx)i)qx + (1-\tau)wz'\varepsilon', (1-\alpha)(1-\tau)wz'\varepsilon'\}. \end{aligned}$$

The parameter α describes the intensity of the limited liability condition, and $\alpha = 0$ corresponds to the baseline model.

The bottom-left panel in Figure 1.6 displays the result. It plots the change in the insured rate of the active participants when I change this intensity parameter. The insured rate gradually increases as the condition of the limited liability gets intensified. When the intensity parameter is really high in the sense that more than 90% of the current income is confiscated after medical bankruptcy, then the insured rate substantially increases. When people lose 99.9% of the income, then the insured rate is doubled from the baseline, i.e., 43.3% people would purchase health insurance. However, more than half of the households still remain uninsured. This is because the limited liability is not a relevant consideration for the uninsured rich. Moreover, there are some households who simply cannot afford the insurance.

Lastly, I change the condition of limited liability in the economy without the fixed costs, i.e., $\phi = 0$. This exercise is informative because it illustrates how many of the uninsured arise from limited liability and from poverty under the actuarially fair premiums.

The bottom-right panel in Figure 1.6 displays the result. The insured rate increases as the condition of the limited liability gets stronger as is the previous case, and if people lose 99.9% of the income when filing medical bankruptcy, then more than three-fourths become insured (75.9%). This exercise uncover the fact that those who chose to be uninsured even with the actuarially fair premiums become insured if the condition of limited liability is very strict. The remaining

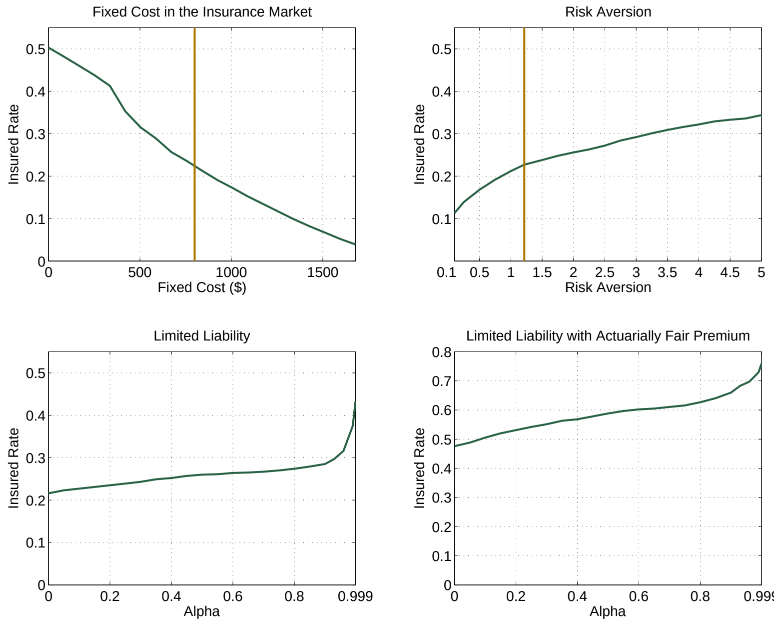


Figure 1.6: Relative Importance in Accounting for High Uninsured Rates among the Active Participants: (i) Fixed Cost in the Insurance Market, (ii) Risk Aversion, (iii) Limited Liability, (iv) Limited Liability when Premiums are Actuarially Fair. The vertical lines in the top panels represent the estimated values in the baseline economy.

households, one-fourths of the active participants, are too poor to afford insurance. The premium subsidy of the policy reform is effective to help them purchase insurance.

1.4.3 Implications of the Reform

This section describes the implications of the reform.

First, the reform has effects on the aggregate variables and prices, as displayed in Table 1.7. The insured rate of the active participants significantly increases after the reform, resulting in the low uninsured rate among the working age population, 3.1%. This means that the reform achieves its primary goal, i.e., the expansion of the insurance coverage. Those who remain uninsured tend to be rich, have moderate income, and be in good health. For this group the penalty for not buying insurance is relatively small and the premium is actuarially unfair due to community rating.

The low uninsured rate induces the reduction in the mark-up in the medical services market because the number of people who take advantage of the free care substantially decreases. The

Table 1.7: Implications of the Reform for Aggregate Variables and Prices

	Before	After
Uninsured Rate: working age population	19.8%	3.1%
Uninsured Rate: active participants	77.1%	11.9%
Aggregate Output	1.126	1.133
Aggregate Capital	3.31	3.32
Interest Rate	3.00%	3.06%
Income Tax Rate	25.0%	25.9%
Mark-up in the Medical Services	6.70%	1.62%
Fraction of Healthy	63.7%	70.3%
Health Care Spending in GDP (age 25+)	9.61%	9.85%

mark-up decreases from 6.7% to 1.6%, and hence the medical cost for the same care decreases by 5%[17]. However, the fraction of the aggregate health spending in GDP increases by 0.2 percentage points. This is because more people are insured after the reform and they use primary care. This results in the improvement in the average healthy status among the working age population. The number of people who enjoy healthy life increases by about 7% by the reform.

There are also slight increases in the aggregate output and the aggregate capital.

Wealth Inequality

The reform has an implication for wealth inequality. As is shown in Table 1.8, the Gini coefficient drops from 0.555 to 0.545 for all working age people and from 0.653 to 0.634 for the active participants. This decline comes from the significant increase in the wealth accumulation among the poor and decrease among the rich. In fact, the asset holdings of the household in the top 20th percentiles drops from 55.0% of the aggregate wealth to 54.6%, whereas that in the bottom 40th percentiles rises from 5.0% to 6.0%.

Before the reform, the uninsured poor are the major users of free care. Because their accumulated asset is confiscated when a sizable medical shock hits, they have strong incentive to dissave. Hence, they are trapped at the bottom of the wealth distribution. This mechanism is similar to the one proposed in Hubbard, Skinner, and Zeldes (1995), but in this case, limited liability in the

[17] This 6.7% of mark-up before the reform is comparable to Kathleen Stoll and Kim Bailey (2009) which reports 7.7% of the mark-up using the MEPS data.

Table 1.8: Wealth Distribution before and after the Reform

	Before	After
Gini wealth: working age population	0.555	0.545
Gini wealth: active participants	0.653	0.634
Wealth Percentile (active participants)		
25%	$2,820	$4,979
50%	$26,857	$30,692
75%	$106,032	$104,182

Table 1.9: Welfare Effect of the Reform (Consumption Equivalent Variation)

	All Working Age	Active Participants
Welfare Gain	0.08%	0.19%
Fraction who gains	51.8%	52.8%

medical services market is acting the same role as the social safety net in their paper. The reform gets this population to obtain insurance, and hence they do not take advantage of the free care opportunity. Once they no longer fear that any savings could be lost to pay for health care expenses after receiving emergency care, they have stronger incentive to accumulate savings.

On the other hand, the biggest fear before the reform is to experience a large negative medical shock as well as a negative income shock. As the reform induces richer households to purchase insurance, it effectively reduces the risk exposure of that group. Specifically, the reform makes it easier to access to insurance through the community rating and the premium subsidy. This discourages precautionary saving.

Welfare Analysis

This section describes the implications of the reform on the welfare. The welfare effects are measured by the consumption equivalent variation for the economies before and after the reform.

As Table 1.9 displays, the reform generates a moderate welfare gain for the working age population (0.08%) and more than 50% are in favor of the reform. On the other hand, there is a more gain on average for the active participants (equivalent to 0.19% of consumption), and 52.8% of them are in favor of the reform.

Table 1.10: Welfare Effect of the Reform on the Active Participants (CEV, %)

			Wealth Quartile	
Age	Income*	Health	Bottom	Top
25-34	Low	Good	−0.15	1.00
		Bad	−0.21	0.97
	High	Good	−0.17	0.03
		Bad	−0.19	0.05
55-64	Low	Good	−0.21	0.98
		Bad	−0.44	1.02
	High	Good	−0.87	−0.44
		Bad	−0.88	−0.40

* "Low income" and "High income" mean the least and the most productive households, respectively.

Next, I decompose the active participants into subgroups. In Table 1.10, I compute the average welfare gains for the group of (i) age 25-34 and age 55-64, (ii) those who are the least and the most productive, (iii) the healthy and the unhealthy, and (iv) those with wealth in the bottom and top quartile.

It is clear from Table 1.10 that there is a contrast in the welfare effects between the rich and the poor. Perhaps surprisingly, the poor who have wealth in the bottom quartile are worse off by the reform. This is because the reform gets them to buy insurance, and hence they no longer take advantage of limited liability. Therefore the reform brings welfare losses for this population, even with the premium subsidy. The unhealthy poor are suffered more than the healthy poor, because they rely more on the free care before the reform.

On the other hand, for the rich who have wealth in the top quartile, the limited liability is not a relevant consideration. The reform eliminates the risk exposure of this population. As described in the previous section, the reform makes it easier to access to insurance when one's health is deteriorated and one is hit by a negative income shock. Thus the rich decrease their precautionary saving and increase consumption, resulting in the welfare gains among this population. Especially, the rich with low income have substantial welfare gains because of the direct transfers associated with the premium subsidies.

1.5 Conclusion

This paper quantifies the implications of the upcoming health care reform in the United States participation in private health insurance markets, individual saving behavior, and distribution of welfare gains and losses. The reform substantially decreases the uninsured population. As the poor obtain insurance and no longer take advantage of the free care opportunity in the medical services market, they accumulate more assets. On the other hand, the rich decrease the precautionary saving, because they are less exposed to medical expenditure risks after the reform. Therefore, the reform decreases the wealth inequality. The reform also brings welfare gains for the rich, but losses for the poor because they are much less likely to use the free care option.

Chapter 2

Optimal Income Taxation: Mirrlees Meets Ramsey

2.1 Introduction

This chapter quantitatively studies the optimal income tax schedule, following both Mirrlees approach and Ramsey approach. The Ramsey approach to optimal taxation assumes functional forms and solves for the optimal income tax schedule within that class of functions, while the Mirrlees approach solves for the constrained-efficient allocation. The former is simple but has been criticized because it assumes ad hoc functional forms, whereas the latter is theoretically appealing but often suggests highly non-linear tax schedules (Saez (2001)), which are very different from the observed tax schedule in the U.S. Recent papers compare Mirrlees solution to optimal linear policy, but there is not much consensus on whether welfare differences are large or small. This paper systematically measures the difference, in terms of welfare, between Mirrlees taxation and simple but ad hoc Ramsey policies for a range of static economies.

We build on the standard static Mirrlees economy in which heterogeneous households choose their labor supply given their productivity, and the planner chooses income tax schedule to maximize the social welfare. Departing from the traditional Mirrlees approach, we assume in the model that each individual's log wage is the sum of two orthogonal components: a component observable and unobservable to the tax authority. This is motivated by the fact that the wage in the U.S. varies systematically with observable characteristics of workers such as education, age, gender and race. The relative variances of the two components are calibrated to evidence from wage regressions on the relative variances of between-group versus within-group wage dispersion. We then solve for the constrained-efficient allocation, i.e., the second best allocation, following Mirrlees approach. Also, following the Ramsey approach, we look for simple approximate implementations in the class of polynomial functions.

We find that the constrained-efficient allocation is highly distortionary; the output decreases by around 12% relative to a baseline flat tax. However, the welfare gains from redistribution are as large as 16% of consumption. The key force is that under the optimal tax schedule, agents are distorted by income tax more than the flat tax economy, but there is higher redistribution mainly through transfers. More importantly, we show that the constrained-efficient allocation and hence the welfare gains can be approximately implemented by very simple and realistic tax schemes. Specifically, we find that a linear income tax with lump-sum transfers can achieve almost 96% of the gains, and a third-order polynomial tax function approximately implements the constrained-efficient allocation. We show, however, that to approximate the second best allocation, it is crucial to condition both lump-sum transfers and marginal tax rates on observables. If the planner does not take advantage of the observability, 25% of the potential gains are lost.

The finding that the simple tax system approximately implements the constrained-efficient allocation is robust to various settings, including different wage distributions and preferences. This is useful because it helps reconcile various results in the literature for papers that consider a variety of alternative special cases. Therefore, our finding tells us that the discussion in the literature about how wildly marginal tax rates change with income (for example, Saez (2001)) might be of little quantitative relevance. It also suggests that both approaches are effectively very similar in terms of resulting allocations, as long as we are judicious in the choice of tax function. The important implication, however, is that even if only a quarter of the wage variance reflects the observable component, it is critical to condition taxes on observables.

Another finding is that the welfare gains are sensitive to the coarseness of the productivity distribution; if we use very coarse grid points, which is common in the dynamic public finance literature, the Mirrlees planner has too much power to control the agents' incentive and can achieve much larger gains than the Ramsey planner. This could be one reason why several papers find that Ramsey style tax schemes are very costly in welfare terms (see Conesa, Kitao, and Krueger (2009)) and Fukushima (2010)).

Based on the seminal paper Mirrlees (1971), Diamond (1998) uses quasi-linear preferences and derives characteristics of the second best allocation. Saez (2001) generalizes the argument to the general preferences and argues that the optimal tax schedule is highly non-linear. Mankiw and Weinzierl (2010) consider the effect of the type-dependent tax system in the static environment and find large welfare gains, as in this chapter. Weinzierl (2011) analyzes the effect of age-dependent taxation in a dynamic economy and also finds large welfare gains.

2.2 Model

We describe the standard static Mirrlees economy in which agents are characterized by a pair of α and ε, two independently distributed components. $F(\alpha)$ and $F(\varepsilon)$ are the respective distributions. Log wage is given by

$$\log w = \alpha + \varepsilon.$$

Denote the agents' utility as $U(c,h)$ where c is consumption and h is hours worked, and the social welfare function as W. Let G be the government consumption which is exogenously determined.

We consider three cases. First, we consider a situation in which α and ε are both publicly observable, and call the equilibrium allocation in this case as the first best allocation. Second, we consider a situation in which α is private information but ε is publicly observable. We consider a Mirrlees planner who has access to unrestricted non-linear taxes, and call the equilibrium allocation in this case as the second best allocation. Third, under the same information structure as the previous case, we consider a Ramsey planner who has access to restricted tax functions and call the equilibrium allocation in this case as the Ramsey allocation.

First Best Problem

The planner maximizes the objective function subject to the resource constraint:

$$\begin{cases} \max\limits_{c(\alpha,\varepsilon),h(\alpha,\varepsilon)} & \int\int U(c(\alpha,\varepsilon),h(\alpha,\varepsilon))W(\alpha,\varepsilon)dF(\alpha)dF(\varepsilon) \\ \text{s.t.} & \int\int c(\alpha,\varepsilon)dF(\alpha)dF(\varepsilon) + G \leq \int\int \exp(\alpha+\varepsilon)h(\alpha,\varepsilon)dF(\alpha)dF(\varepsilon) \end{cases}$$

Mirrlees Problem

The Mirrlees planner maximizes the objective function subject to the resource constraint and incentive compatibility conditions:

$$\begin{cases} \max\limits_{c(\alpha,\varepsilon),h(\alpha,\varepsilon)} & \int\int U(c(\alpha,\varepsilon),h(\alpha,\varepsilon))W(\alpha,\varepsilon)dF(\alpha)dF(\varepsilon) \\ \text{s.t.} & \int\int c(\alpha,\varepsilon)dF(\alpha)dF(\varepsilon)+G \leq \int\int \exp(\alpha+\varepsilon)h(\alpha,\varepsilon)dF(\alpha)dF(\varepsilon) \\ & U(c(\alpha,\varepsilon),h(\alpha,\varepsilon)) \geq U\left(c(\tilde{\alpha},\varepsilon),h(\tilde{\alpha},\varepsilon)\frac{\exp(\tilde{\alpha}+\varepsilon)}{\exp(\alpha+\varepsilon)}\right) \quad \forall\tilde{\alpha}\neq\alpha \end{cases}$$

The incentive compatibility conditions say that agents have no incentive to pretend different types $\tilde{\alpha}$ from their true type α.

Ramsey Problem

Define $\mathbb{E}(\{T_\varepsilon\})$, a set of all competitive equilibrium allocations, given an arbitrary tax system $\{T_\varepsilon\}$. Note that the tax function might depend on the observable characteristics ε. The Ramsey problem then is

$$\begin{cases} \max\limits_{T_\varepsilon\in\mathcal{T}} & \int\int U(c(\alpha,\varepsilon),h(\alpha,\varepsilon))W(\alpha,\varepsilon)dF(\alpha)dF(\varepsilon) \\ \text{s.t.} & \{c(\alpha,\varepsilon),h(\alpha,\varepsilon)\}\in\mathcal{C}(T_\varepsilon) \end{cases}$$

where $\mathcal{T}$ is a set of a particular class of functions. The competitive equilibrium of this economy given $\{T_\varepsilon\}$ is a set of allocations $\{c,h\}$ such that (i) it solves agents' problem: for all α and ε, $\{c,h\}$ solves

$$\begin{cases} \max\limits_{c,h} & U(c,h) \\ \text{s.t.} & c \leq \exp(\alpha+\varepsilon)h - T_\varepsilon(\exp(\alpha+\varepsilon)h). \end{cases}$$

and (ii) it satisfies the feasibility condition: $\int c(\alpha,\varepsilon)F(d\alpha,d\varepsilon)+G \leq \int \exp(\alpha+\varepsilon)h(\alpha,\varepsilon)F(d\alpha,d\varepsilon)$.

Any Ramsey allocation is incentive compatible by construction, and coincides with the second best Mirrlees allocation when we let $\mathcal{T}$ be a set of any class of functions. In the following, we examine how these allocations and the social welfare quantitatively differ from each other.

2.3 Computation

In this section, we describe the allocation for each of the three cases above and provide an algorithm to compute the second best allocation which is different from Saez (2001) and Mankiw, Weinzierl, and Yagan (2009).

We discretize the types by $\alpha_1, \cdots, \alpha_{N_\alpha}$ and $\varepsilon_1, \cdots, \varepsilon_{N_\varepsilon}$ with fraction $\pi_{i,j}$ for each α_i and ε_j. Let $\theta_{i,j} = \exp(\alpha_i + \varepsilon_j)$ and $y_{i,j} = \theta_{i,j} h$. Assume the utility is separable: $U(c,h) = u(c) - v(h)$ and, for the sake of the explanation, we assume the Utilitarian social planner. In Appendix A we describe the allocation under the Rawlsian planner.

2.3.1 First Best Allocation

The planner's problem becomes:

$$\begin{cases} \max\limits_{c_{i,j}, y_{i,j}} & \sum \pi_{i,j} \left[u\left(c_{i,j}\right) - v\left(\frac{y_{i,j}}{\theta_{i,j}}\right) \right] \\ \text{s.t.} & \sum \pi_{i,j} c_{i,j} + G = \sum \pi_{i,j} y_{i,j}. \end{cases}$$

Solving the first order conditions, the first best allocation $\{c_{i,j}, y_{i,j}\}$ then is given by

$$\begin{cases} c_{i,j} = c = \sum \pi_{i,j} y_{i,j} - G \text{ for all } i, j, \\ u'\left(c\right) = v'\left(\frac{y_{i,j}}{\theta_{i,j}}\right) \frac{1}{\theta_{i,j}}. \end{cases}$$

The first best allocation is that everybody consumes the same amount and the hours worked are determined based on the productivity (and the social weight for non-Utilitalian cases).

2.3.2 Ramsey Allocation

The Ramsey problem becomes

$$\begin{cases} \max\limits_{\{T_\varepsilon\} \in \mathcal{T}} & \sum \pi_{i,j} \left[u\left(c_{i,j}\right) - v\left(\frac{y_{i,j}}{\theta_{i,j}}\right) \right] \\ \text{s.t.} & \{c_{i,j}, y_{i,j}\}_{i,j} \in \mathbb{E}(\{T_\varepsilon\}), \end{cases}$$

and the competitive equilibrium is a set of allocations $\{c_{i,j}, y_{i,j}\}$ such that it (i) solves agents' problem:

$$\begin{cases} \max\limits_{c_{i,j}, y_{i,j}} & u\left(c_{i,j}\right) - v\left(\frac{y_{i,j}}{\theta_{i,j}}\right) \\ \text{s.t.} & c_{i,j} \leq y_{i,j} - T_{\varepsilon_j}\left(y_{i,j}\right). \end{cases}$$

and (ii) satisfies the feasibility condition: $\sum \pi_{i,j} c_{i,j} + G = \sum \pi_{i,j} y_{i,j}$. It is straightforward to derive the equilibrium conditions:

$$\begin{cases} u'\left(c_{i,j}\right)\left(1 - T'_{\varepsilon_j}\left(y_{i,j}\right)\right) = v'\left(\frac{y_{i,j}}{\theta_{i,j}}\right)\frac{1}{\theta_{i,j}}, \\ c_{i,j} = y_{i,j} - T_{\varepsilon_j}\left(y_{i,j}\right), \\ \sum \pi_{i,j} c_{i,j} + G = \sum \pi_{i,j} y_{i,j}. \end{cases}$$

Particularly, if we have the flat tax with the constant tax rate τ, the equilibrium allocation is characterized by

$$\begin{cases} c_{i,j} = (1-\tau) y_{i,j} \text{ for all } i, j, \\ u'\left(c_{i,j}\right)(1-\tau) = v'\left(\frac{y_{i,j}}{\theta_{i,j}}\right)\frac{1}{\theta_{i,j}}, \\ G = \tau \sum \pi_{i,j} y_{i,j}. \end{cases}$$

In the numerical analysis below, we consider this flat tax economy as the benchmark. We also consider earnings tax systems in class of polynomial functions[1]

$$T_{\varepsilon}\left(y\right) = \tau_{\varepsilon}^{0} + \tau_{\varepsilon}^{1} y + \tau_{\varepsilon}^{2} y^{2} + ... + \tau_{\varepsilon}^{N} y^{N}$$

[1]With n-th order polynomial tax functions, it is possible to have multiple roots of the first order conditions, meaning that the first order conditions are necessary but no longer sufficient. However, the claim below says that with our specification of the utility and the tax function in the numerical analysis, it is sufficient to take a large $\bar{y} > 0$ and find the root(s) of the first order necessary conditions in the interval of $[0, \bar{y}]$.

Claim 1 *Assume an n-th order polynomial tax function. Suppose we solve*

$$\begin{array}{ll} \max_{c \geq 0, y \geq 0} & \frac{c^{1-\gamma}}{1-\gamma} - v(y) \\ s.t. & c = y - (\tau_0 + \tau_1 y + \tau_2 y^2 + \cdots + \tau_n y^n) \end{array}$$

with $\gamma \geq 1$, $v' > 0$ and $v'' > 0$. Then there exists $\bar{y} \in \mathbb{R}_+$ such that for all $y > \bar{y}$, $U'(y) < 0$ where U is the value function we get when substituting the optimal consumption.

Proof. We have

$$\begin{aligned} U'(y) &= (1 - \tau_1 - 2\tau_2 y - \cdots - n\tau_n y^{n-1})[y - (\tau_0 + \tau_1 y + \tau_2 y^2 \cdots + \tau_n y^n)]^{-\gamma} - v'(y) \\ &= (y^{n-1})^{1-\gamma}(\frac{1}{y^{n-1}} - \frac{\tau_1}{y^{n-1}} - \cdots - n\tau_n)[\frac{1}{y^{n-1}} - \frac{\tau_0}{y^{n-1}} - \frac{\tau_1}{y^{n-2}} - \cdots - \tau_{n-1} - \tau_n y]^{-\gamma} - v'(y). \end{aligned}$$

The first term converges to zero as y gets larger, whereas the second term diverges. ■

When numerically choosing tax coefficients within this class of functions, a good guess is obtained by a regression; given the second best allocation, we approximate it by a regression:

$$T_{i,j} = \tilde{\tau}_{0,j} + \tilde{\tau}_{1,j} y_{i,j} + \tilde{\tau}_{2,j} {y_{i,j}}^2 + \tilde{\tau}_{3,j} {y_{i,j}}^3 + (\text{approx. error}).$$

where the left hand side is the net tax payment for each type α_i, ε_j, $T_{i,j} = y_{i,j} - c_{i,j}$. Using these $\tilde{\tau}$'s makes the computation of the Ramsey solution much faster.

2.3.3 Second Best Allocation

Planner's Problem

The planner's problem becomes

$$\begin{cases} \max\limits_{c_{i,j},h_{i,j}} & \sum \pi_{i,j} \left[u\left(c_{i,j}\right) - v\left(h_{i,j}\right) \right] \\ \text{s.t.} & u\left(c_{i,j}\right) - v\left(h_{i,j}\right) \geq u\left(c_{k,j}\right) - v\left(h_{k,j} \frac{\theta_{k,j}}{\theta_{i,j}}\right) \quad \forall j, k \neq i \\ & \sum \pi_{i,j} c_{i,j} + G \leq \sum \pi_{i,j} h_{i,j} \theta_{i,j}. \end{cases}$$

Given the well-known fact that the local incentive constraints are sufficient for the standard utility functions, we can simplify the problem as

$$\begin{cases} \max\limits_{c_{i,j},h_{i,j}} & \sum \pi_{i,j} \left[u\left(c_{i,j}\right) - v\left(h_{i,j}\right) \right] \\ \text{s.t.} & u\left(c_{i,j}\right) - v\left(h_{i,j}\right) \geq u\left(c_{i-1,j}\right) - v\left(h_{i-1,j} \frac{\theta_{i-1,j}}{\theta_{i,j}}\right) \quad \forall i, j \\ & \sum \pi_{i,j} c_{i,j} + G \leq \sum \pi_{i,j} h_{i,j} \theta_{i,j}. \end{cases}$$

We put Lagrange multipliers $\mu_{i,j}$ to the incentive constraints and λ to the resource constraint.

The first order conditions are then for all $j = 1, \cdots, N_\varepsilon$

$$
\begin{cases}
\text{w.r.t } c_{N_\alpha,j} & : \quad \pi_{N_\alpha,j}\lambda = u'\left(c_{N_\alpha,j}\right)\left[\pi_{N_\alpha,j} + \mu_{N_\alpha,j}\right] \\
\text{w.r.t. } h_{N_\alpha,j} & : \quad \pi_{N_\alpha,j}\theta_{N_\alpha,j}\lambda = v'\left(h_{N_\alpha,j}\right)\left[\pi_{N_\alpha,j} + \mu_{N_\alpha,j}\right] \\
\text{w.r.t. } c_{i,j}, i = 2, \cdots, N_\alpha - 1 & : \quad \pi_{i,j}\lambda = u'\left(c_{i,j}\right)\left[\pi_{i,j} + \mu_{i,j} - \mu_{i+1,j}\right] \\
\text{w.r.t. } h_{i,j}, i = 2, \cdots, N_\alpha - 1 & : \quad \pi_{i,j}\theta_{i,j}\lambda = v'\left(h_{i,j}\right)\left[\pi_{i,j} + \mu_{i,j} - \mu_{i+1,j}\frac{v'\left(h_{i,j}\frac{\theta_{i,j}}{\theta_{i+1,j}}\right)}{v'\left(h_{i,j}\right)}\frac{\theta_{i,j}}{\theta_{i+1,j}}\right] \\
\text{w.r.t. } c_{1,j} & : \quad \pi_{1,j}\lambda = u'\left(c_{1,j}\right)\left[\pi_{1,j} - \mu_{2,j}\right] \\
\text{w.r.t. } h_{1,j} & : \quad \pi_{1,j}\theta_{1,j}\lambda = v'\left(h_{1,j}\right)\left[\pi_{1,j} - \mu_{2,j}\frac{v'\left(h_{1,j}\frac{\theta_{1,j}}{\theta_{2,j}}\right)}{v'\left(h_{1,j}\right)}\frac{\theta_{1,j}}{\theta_{2,j}}\right]
\end{cases}
$$

We get the intratemporal Euler equations which equate the marginal rate of substitution of consumption and hours worked.

$$
\begin{cases}
u'\left(c_{N_\alpha,j}\right)\theta_{N_\alpha,j} = v'\left(h_{N_\alpha,j}\right) \\
u'\left(c_{i,j}\right)\theta_{i,j} = v'\left(h_{i,j}\right)\frac{\pi_{i,j}+\mu_{i,j}-\mu_{i+1,j}\frac{v'\left(h_{i,j}\frac{\theta_{i,j}}{\theta_{i+1,j}}\right)}{v'\left(h_{i}\right)}\frac{\theta_{i,j}}{\theta_{i+1,j}}}{\pi_{i,j}+\mu_{i,j}-\mu_{i+1,j}} \\
u'\left(c_{1,j}\right)\theta_{1,j} = v'\left(h_{1,j}\right)\frac{\pi_{1,j}-\mu_{2,j}\frac{v'\left(h_{1,j}\frac{\theta_{1,j}}{\theta_{2,j}}\right)}{v'\left(h_{1,j}\right)}\frac{\theta_{1,j}}{\theta_{2,j}}}{\pi_{1,j}-\mu_{2,j}}
\end{cases}
$$

Note that there is no distortion at the top of the distribution as in the literature.

Forward Iteration Algorithm

This section lays out how to compute the allocation described above. Assume the standard utility function, $u(c,h) = \frac{c^{1-\gamma}}{1-\gamma} - \varphi\frac{h^{1+\sigma}}{1+\sigma}$. We utilize the forward iteration in order to avoid the situation of having complex numbers.[2] The gain of our algorithm is found in Appendix B.

We convert our conditions to more useful forms. The first order condition with respect to $c_{i,j}$ becomes

$$
\begin{aligned}
\pi_{i,j}\lambda &= c_{i,j}^{-\gamma}\left[\pi_{i,j} + \mu_{i,j} - \mu_{i+1,j}\right] \\
\Leftrightarrow \quad \mu_{i+1,j} &= \mu_{i,j} - \pi_{i,j}\left(\lambda c_{i,j}^{\gamma} - 1\right).
\end{aligned} \tag{2.1}
$$

[2]This happens when we have the population density π in a denominator of the equilibrium conditions because π potentially could be a very small number with fine grids and hence is sensitive to rounding errors.

Substituting this into the first order condition with respect to $h_{i,j}$, we get

$$\begin{aligned}
& \pi_{i,j}\theta_{i,j}\lambda = \varphi h_{i,j}^{\sigma}\left[\pi_{i,j} + \mu_{i,j} - \mu_{i+1,j}\left(\frac{\theta_{i,j}}{\theta_{i+1,j}}\right)^{1+\sigma}\right] \\
\Leftrightarrow\ & \pi_{i,j}\theta_{i,j}\lambda = \varphi h_{i,j}^{\sigma}\left[\pi_{i,j} + \mu_{i,j} - \left[\mu_{i,j} - \pi_{i,j}\left(\lambda c_{i,j}^{\gamma} - 1\right)\right]\left(\frac{\theta_{i,j}}{\theta_{i+1,j}}\right)^{1+\sigma}\right] \\
\Leftrightarrow\ & h_{i,j}^{\sigma} = \frac{\pi_{i,j}\theta_{i,j}\lambda}{\varphi}\left[\pi_{i,j} + \mu_{i,j} - \left[\mu_{i,j} - \pi_{i,j}\left(\lambda c_{i,j}^{\gamma} - 1\right)\right]\left(\frac{\theta_{i,j}}{\theta_{i+1,j}}\right)^{1+\sigma}\right]^{-1} \\
\Leftrightarrow\ & h_{i} = \left(\frac{\pi_{i,j}\theta_{i,j}\lambda}{\varphi}\right)^{\frac{1}{\sigma}}\left\{\pi_{i,j} + \mu_{i,j} - \left[\mu_{i,j} - \pi_{i,j}\left(\lambda c_{i,j}^{\gamma} - 1\right)\right]\left(\frac{\theta_{i,j}}{\theta_{i+1,j}}\right)^{1+\sigma}\right\}^{-\frac{1}{\sigma}}.
\end{aligned} \tag{2.2}$$

The incentive compatibility constraint for $i = 2, \cdots, N_\alpha - 1$ becomes

$$\begin{aligned}
& \frac{c_{i,j}^{1-\gamma}}{1-\gamma} - \varphi\frac{h_{i,j}^{1+\sigma}}{1+\sigma} = \frac{c_{i-1,j}^{1-\gamma}}{1-\gamma} - \varphi\frac{\left(h_{i-1,j}\frac{\theta_{i-1,j}}{\theta_{i,j}}\right)^{1+\sigma}}{1+\sigma} \\
\Leftrightarrow\ & \frac{c_{i,j}^{1-\gamma}}{1-\gamma} - \frac{\varphi}{1+\sigma}\left(\frac{\pi_{i,j}\theta_{i,j}\lambda}{\varphi}\right)^{\frac{1+\sigma}{\sigma}}\left\{\pi_{i,j} + \mu_{i,j} - \left[\mu_{i,j} - \pi_{i,j}\left(\lambda c_{i,j}^{\gamma} - 1\right)\right]\left(\frac{\theta_{i,j}}{\theta_{i+1,j}}\right)^{1+\sigma}\right\}^{-\frac{1+\sigma}{\sigma}} \\
& \qquad = \frac{c_{i-1,j}^{1-\gamma}}{1-\gamma} - \varphi\frac{\left(h_{i-1,j}\frac{\theta_{i-1,j}}{\theta_{i,j}}\right)^{1+\sigma}}{1+\sigma}.
\end{aligned} \tag{2.3}$$

Given $c_{i-1,j}, h_{i-1,j}, \mu_{i,j}$ and λ, this is a nonlinear equation for one unknown, $c_{i,j}$.

Algorithm

1. Guess the value of the Lagrange multiplier for the resource constraint: $\lambda > 0$.
2. For each $j = 1, \cdots, N_\varepsilon$, choose $c_{1,j}$ so that the implied sequence satisfies all the first order conditions and incentive compatibility constraints:

   i. Choose $c_{1,j} \in (c_{\min}, c_{\max})$.

   ii. Obtain $\mu_{2,j}$ and $h_{1,j}$ from the first order conditions:

$$\begin{aligned}
\mu_{2,j} &= \pi_{1,j}\left(1 - \lambda c_{1,j}^{\gamma}\right), \\
h_{1,j} &= \left(\frac{\pi_{1,j}\theta_{1,j}\lambda}{\varphi}\right)^{\frac{1}{\sigma}}\left[\pi_{1,j} - \mu_{2,j}\left(\frac{\theta_{1,j}}{\theta_{2,j}}\right)^{1+\sigma}\right]^{\frac{-1}{\sigma}}.
\end{aligned}$$

   iii. For $i = 2, \cdots, N_\alpha - 1$, given $c_{i-1,j}$, $h_{i-1,j}$ and $\mu_{i,j}$, solve for $c_{i,j}, h_{i,j}$ and $\mu_{i+1,j}$ from (2.3), (2.2) and (2.1), respectively.

iv. Given $\mu_{N_\alpha,j}$, compute $c_{N_\alpha,j}$ and $h_{N_\alpha,j}$ from the first order conditions:

$$c_{N_\alpha,j} = \left(\frac{\pi_{N_\alpha,j}+\mu_{N_\alpha,j}}{\pi_{N_\alpha,j}\lambda}\right)^{\frac{1}{\gamma}}$$
$$h_{N_\alpha,j} = \left(\frac{\pi_{N_\alpha,j}\theta_{N_\alpha,j}\lambda}{\varphi(\pi_{N_\alpha,j}+\mu_{N_\alpha,j})}\right)^{\frac{1}{\sigma}}$$

v. Check the incentive constraint for the type N_α agent;

$$\frac{c_{N_\alpha,j}^{1-\gamma}}{1-\gamma} - \varphi\frac{h_{N_\alpha,j}^{1+\sigma}}{1+\sigma} = \frac{c_{N_\alpha-1,j}^{1-\gamma}}{1-\gamma} - \varphi\frac{\left(h_{N_\alpha-1,j}\frac{\theta_{N_\alpha-1,j}}{\theta_{N_\alpha,j}}\right)^{1+\sigma}}{1+\sigma}.$$

If not satisfied, then go back to step 2-i and adjust $c_{1,j}$.

3. Check if the aggregate resource constraint is satisfied:

$$\sum \pi_{i,j}c_{i,j} + G = \sum \pi_{i,j}h_{i,j}\theta_{i,j}.$$

If not, go back to step 1 and adjust λ. Repeat until this is satisfied under a certain tolerance.

Several comments are following. First, how do we choose $c_{\min}$ and $c_{\max}$? It is natural to assume $c_{\min} = 0$. We know that the incentive constraint is satisfied with equality for $i = 2$ and hence $\mu_{2,j} = \pi_{1,j}\left(1 - \lambda c_{1,j}^{\gamma}\right) > 0$ at optimum. This implies that $\lambda c_{1,j}^{\gamma} < 1$, leading $c_{\max} = \left(\frac{1}{\lambda}\right)^{1/\gamma}$.

Second, how do we choose λ? The Lagrange multiplier for the resource constraint captures the marginal value of public funds and it has to be equal to the inverse of the marginal increase in average utility. That is, summing up all the first order conditions with respect to c, we get $\lambda = \left(\sum \frac{\pi_{i,j}}{u'(c_{i,j})}\right)^{-1}$. Since it is easy to compute the first best allocation, a good guess for the initial λ will be $\lambda = \left(\sum \frac{\pi_{i,j}}{u'(\delta c_{FB})}\right)^{-1}$ where c_{FB} is the first best consumption and $\delta < 1$ is some reduction to it.

Third, an important observation is that λ has a direct implication to the aggregate resource constraint. We then establish a simple relationship; define

$$\Phi(\lambda) \equiv \sum \pi h(\lambda)\theta - \sum \pi c(\lambda) - G,$$

where $h(\lambda)$ and $c(\lambda)$ are the "optimal" allocation given λ derived from the forward iteration which satisfies all the incentive compatibility constraints and first order conditions. Then Φ shows monotonically increasing pattern in λ.

2.4 Numerical Analysis

This section numerically solves the optimal tax schedule described in the previous section.

2.4.1 Calibration

The agents' utility is given by

$$u(c) - v(h) = \frac{c^{1-\gamma}}{1-\gamma} - \varphi \frac{h^{1+\sigma}}{1+\sigma}$$

As a baseline parametrization, we assume the relative risk aversion being unity $\gamma = 1$, i.e., log preference for consumption. Also assume equal weights on c and h, $\varphi = 1$. We choose $\sigma = 2$ so that the Frisch elasticity is 0.5.

We normalize average productivity for both α and ε to one, i,e, $\int \exp(\alpha) dF(\alpha) = 1$ and $\int \exp(\varepsilon) dF(\varepsilon) = 1$.

We consider the linear tax system as the benchmark, $T_\varepsilon(y) = \tau^1 y$

Under these assumptions, the hours worked and the output become independent of σ and τ^1, and we get the hours worked being one for each type $h(\alpha, \varepsilon) = 1$. Therefore we can simply characterize the equilibrium allocation with the linear tax rate τ^1 by

$$\begin{aligned} y(\alpha, \varepsilon) &= \exp(\alpha + \varepsilon), \\ c(\alpha, \varepsilon) &= (1 - \tau^1) \exp(\alpha + \varepsilon). \end{aligned}$$

Also in aggregate, we get the aggregate hours being one, i.e., $\int \int \exp(\alpha + \varepsilon) h(\alpha, \varepsilon) dF(\alpha) dF(\varepsilon) = 1$ and $G = \tau^1$. Because the government consumption and investment is 18.8% of GDP in 2005 in the U.S., we set $\tau^1 = 0.188$. In the rest of the analysis, we fix the *level* of the government expenditure as $G = 0.188$.

For the variance of wages, we follow Heathcote, Perri, and Violante (2010). The total variance of log male wages in 2005 is 0.499, $var(\log(w)) = 0.499$. Since the residual variance after controlling for age, education, and household composition is 0.389, we get $var(\alpha) = 0.389$ and $var(\varepsilon) = 0.110$.

Wage Distribution

We further assume two-point-equal-weight distribution for ε. Given the value of $var(\varepsilon)$ and normalization, we then get $\exp(\varepsilon_H)/\exp(\varepsilon_L) = 1.94$.

As for the uninsurable component of wage, we set the bounds of $\exp(\alpha) \in \left[\frac{\frac{1}{2}\times 5.15}{19.60}, \frac{200.56}{19.60}\right]$. Specifically, we use the Federal minimum wage in 2005, \$5.15, and assume the minimum productivity level is simply the half of it. We use \$200.56, the earnings per hour at 99.5th percentile of 2005 earnings distribution from Piketty and Saez (2003), assuming 2000 hours. Both bounds are normalized by 19.60, the average hourly earnings in 2005 taken from BLS data.

Saez (2001) argues that the wage distribution has a thick right tail and is well-approximated by a Pareto distribution rather than a log normal distribution. We address this issue by assuming the distribution being log-normal for $\exp(\alpha) \leq x$ and being Pareto for $\exp(\alpha) > x$ where $x = 2.32$ so that 95% of the population is in the log-normal range and the rest is in the Pareto range. This value for x corresponds to about \$45. For the Pareto parameter, we use 2.0 estimated from Piketty and Saez (2003). The shape of the wage distribution is found in Figure 2.1.

We use 1,000 evenly spaced grid points and as a sensitivity analysis, we also compute allocations for 2 to 1,000,000 grid points.

2.4.2 Numerical Analysis

This section describes the numerical results for each optimal tax scheme. We take the linear tax system as the benchmark, and evaluate other tax systems relative to this in welfare terms. Specifically, we compute the first best allocation, the Mirrlees second best allocation, and the Ramsey allocations in the class of polynomial tax functions. We consider both ε-type-contingent and no ε-type-contingent taxes.

No Type-Contingent Taxes

Table 2.1 shows the results with no type-contingent taxes. Relative to the benchmark proportional tax scheme, the second best allocation is highly distortionary; the output drops by 12% but

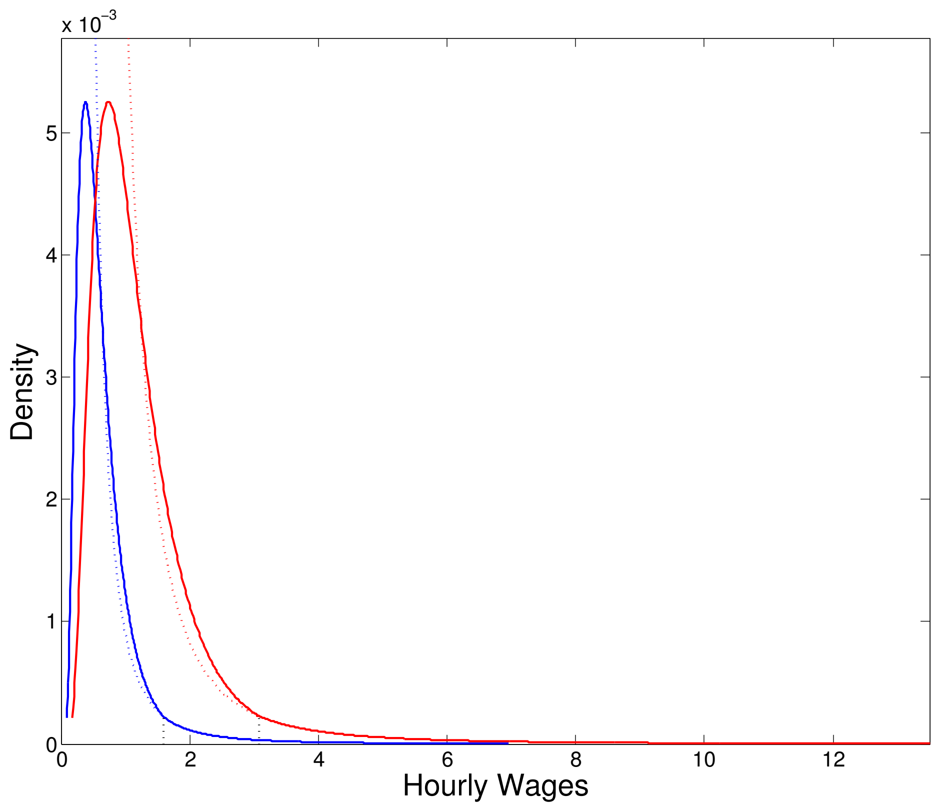


Figure 2.1: Wage Distribution

Table 2.1: Ramsey vs Mirrlees: No Type-Contingent Taxes

Tax system				Outcomes				
τ^0	τ^1	τ^2	τ^3	Welfare	Y	Mar. Tax	G/Y	τ^0/Y
0	0.188	-	-	0	0	0.188	0.188	0
0	0.053	0.108	-0.005	4.49	-5.75	0.298	0.199	0
-0.268	0.534	-	-	12.25	-14.65	0.534	0.220	0.314
-0.267	0.533	0.000	-	12.25	-14.65	0.534	0.220	0.313
-0.258	0.499	0.018	-0.001	12.36	-14.71	0.534	0.220	0.302
Second Best (Mirrlees)				16.48	-11.39	0.496	0.212	0.244
First Best				58.60	22.23	0	0.152	0.808

has large welfare gains of 16.5%. Ramsey schemes imply large output losses about 15% and the optimal marginal tax rates around 50%. The bulk of tax revenue is used for transfers, not spending. This means that to raise the redistribution, the Ramsey government needs to impose a high marginal tax rate, leading a large output loss. It is also noticeable that the lump-sum component is the key for the welfare gains; without the lump-sum component, the welfare gains are as small as 4.5% of consumption. Perhaps surprisingly, the quadratic term does not contribute to the welfare relative to the lump-sum and linear case, whereas the cubic term generates further gains.

Type-Contingent Taxes

Table 2.2 shows the results when we allow for type contingent taxes. There are significant welfare gains relative to non-contingent tax schemes. Especially, the type-specific lump-sum and linear tax scheme approximately implements the second best Mirrlees allocation in terms of the welfare gains; 96% of the welfare gains is achieved. The important implication here is that the planner wants τ^0 and τ^1 to be ε-type specific. The reason is because he/she wants to redistribute resources across ε-types via $\tau_H^1 > \tau_L^1$ for each type[3], while the planner induces higher hours from ε_H-type via $\tau_H^0 > \tau_L^0$.

Figures 2.2-2.4 illustrate the allocation and tax schedule for the baseline flat tax, the lump-sum and linear tax and the cubic tax along with the second best allocation. In each figure, the red line shows the allocation or the tax schedule for Ramsey allocation, whereas the blue line shows that for Mirrlees allocation. The dotted line is for ε_L-type and the solid line is for ε_H-type. Each panel

[3]If only linear taxes are available, the planner sets $\frac{1-\tau_H^1}{1-\tau_L^1} = \frac{\exp(\varepsilon_L)}{\exp(\varepsilon_H)}$.

Table 2.2: Ramsey vs Mirrlees: Type-Contingent Taxes

Tax system				Outcomes			
τ^0	τ^1	τ^2	τ^3	Welfare	Y	Mar. Tax	τ^0/Y
0	0.188			0	0	0.188	0
-0.258	0.499	0.018	-0.001	12.36	-14.71	0.534	0.302
-	-0.195 0.385	-	-	5.55	0.00	0.188	0
-0.413 -0.106	0.508	-	-	15.48	-12.09	0.508	0.295
-0.248	0.328 0.588	-	-	13.94	-13.03	0.501	0.285
-0.389 -0.128	0.422 0.542	-	-	15.79	-11.79	0.506	0.293
-0.367 -0.111	0.318 0.499	0.079 0.019	-0.009 -0.001	16.00	-11.82	0.505	0.271
Second Best (Mirrlees)				16.48	-11.39	0.496	0.244

plots the consumption, the hours worked, the marginal tax rate and the net tax along the log hourly wages. As is shown in Figure 2.3, the marginal tax for the Ramsey allocation with linear taxes is far from that of Mirrlees, which is indicated in the literature. Nevertheless, the equilibrium allocation for consumption and hours worked of Ramsey taxation are very close to Mirrlees allocation. This generates almost the same level of welfare gains.

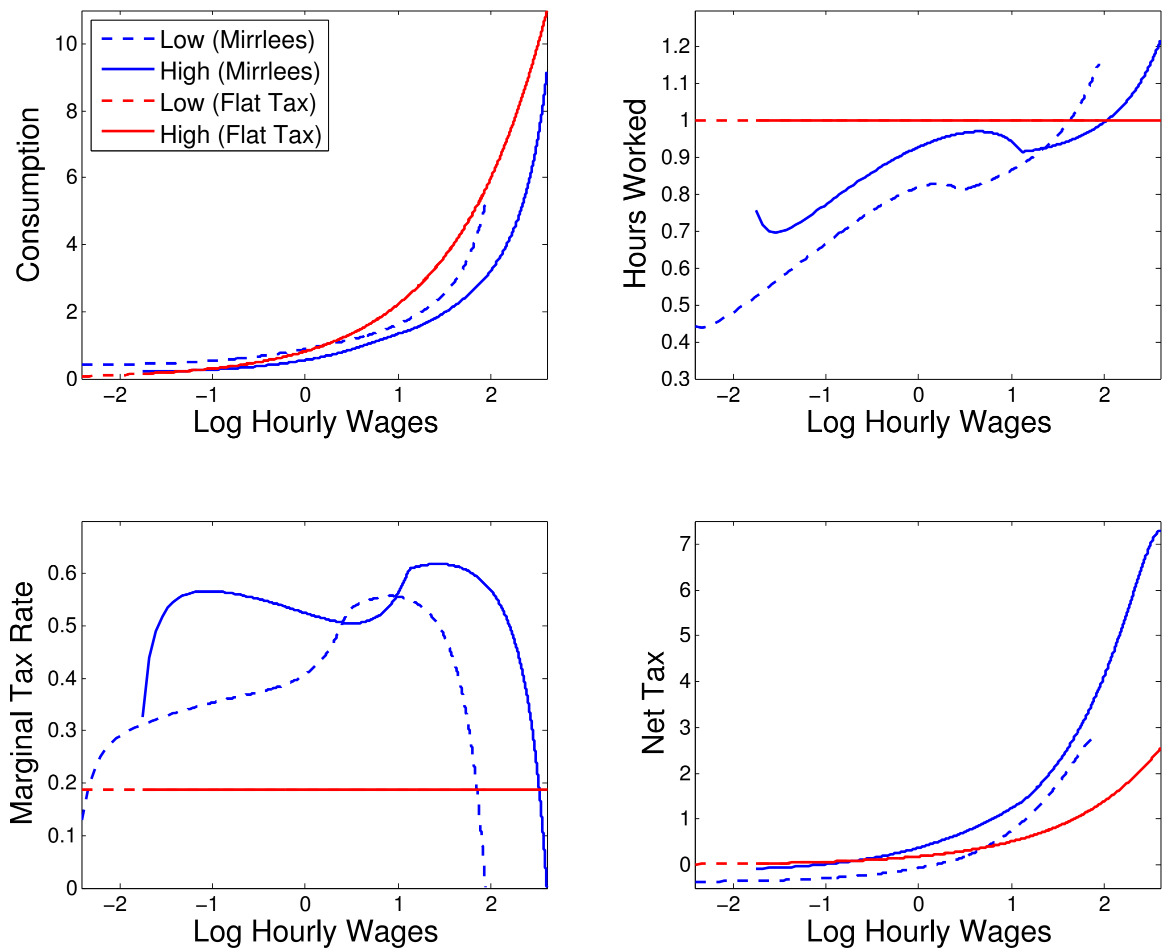


Figure 2.2: Flat tax vs Mirrlees tax

2.4.3 Sensitivity Analysis

This section examines the sensitivity of our analysis to numbers of grid points, shapes of the wage distribution, preferences and social welfare functions.

Number of Grid Points for α

This section considers the sensitivity to the number of grid points, which has gained little attention in the literature.

Table 2.3 shows the difference of the welfare gains between the Ramsey taxation and Mirrlees taxation changes depending on the number of grid points. Specifically, with coarse grid points, which is standard in the dynamic public finance literature, the Mirrlees planner can get closer to the first best. This is because if the ability of one agent is very different from that of the other

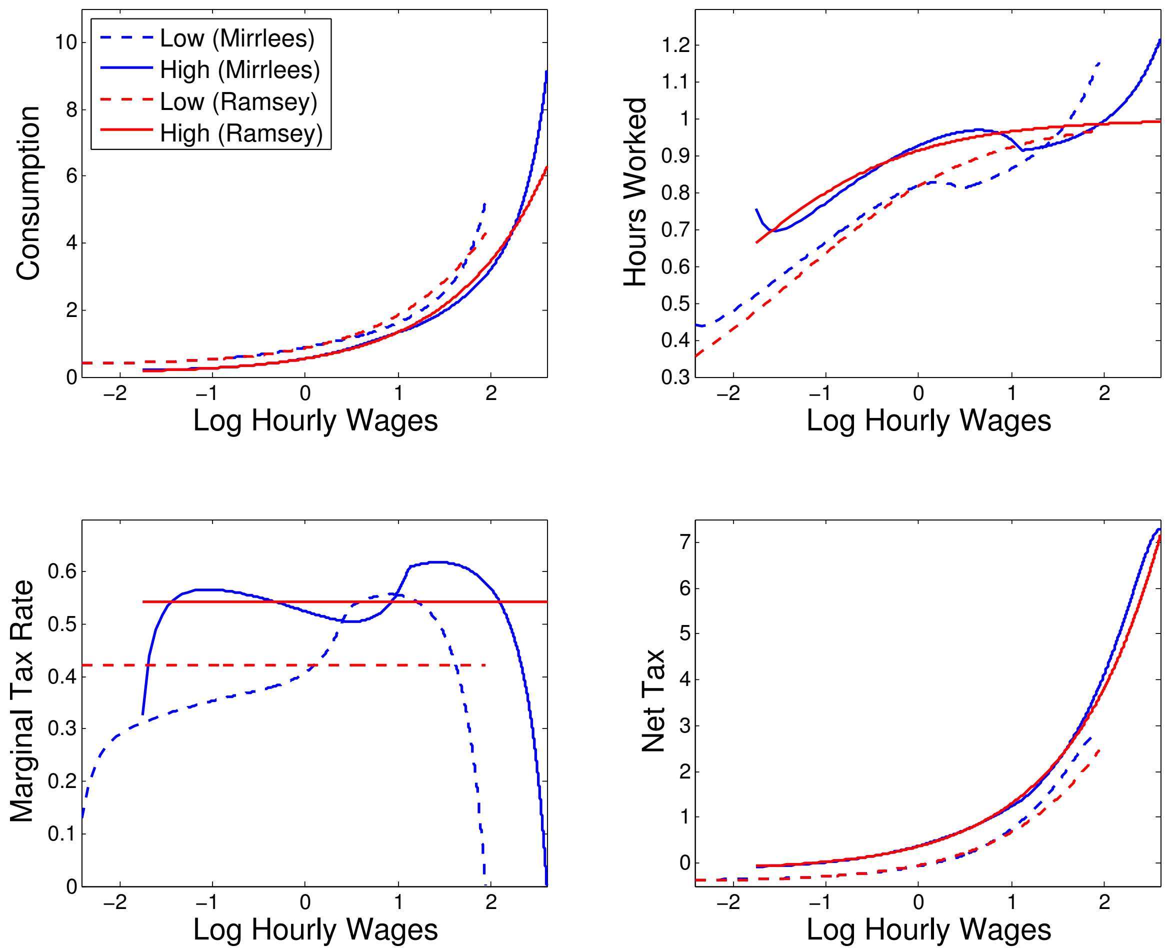


Figure 2.3: Linear tax with lump-sum transfer vs Mirrlees tax

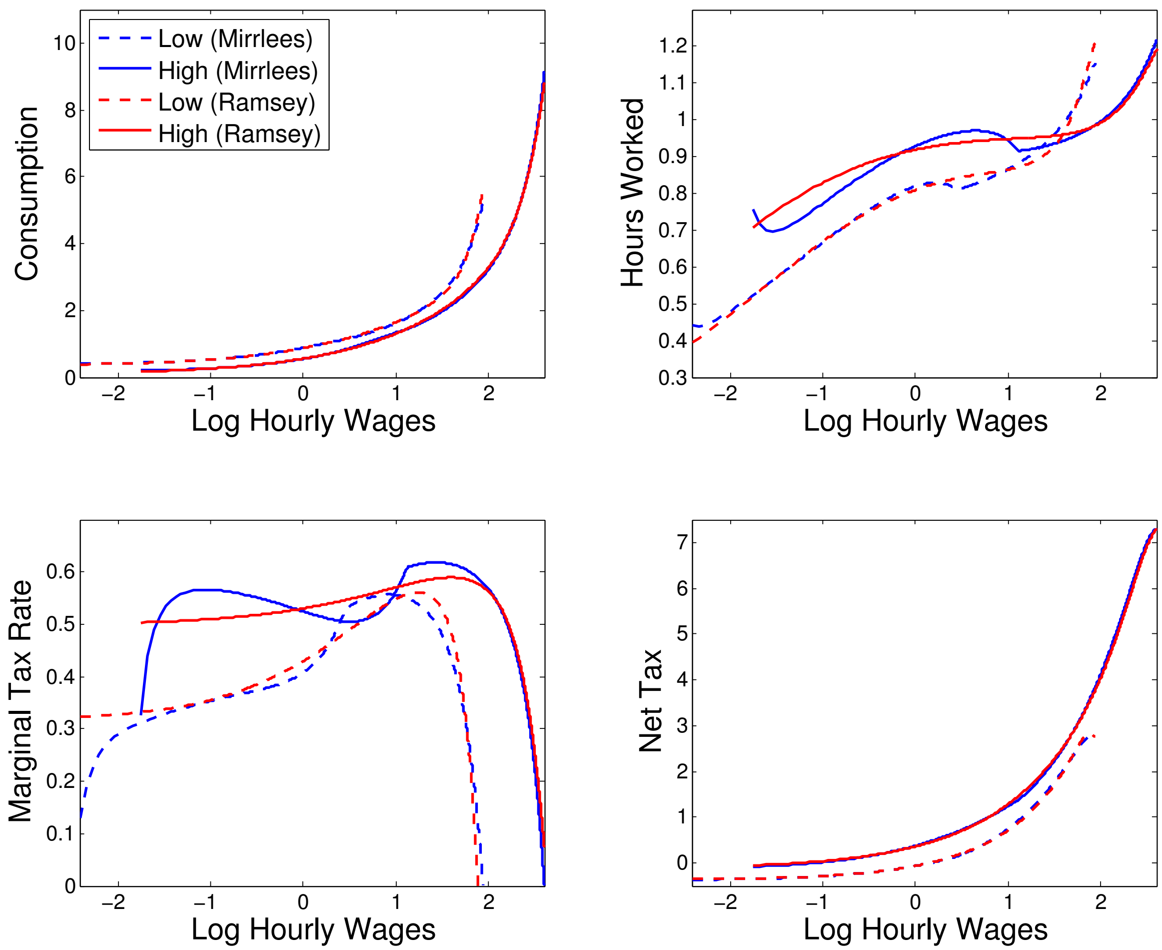


Figure 2.4: Cubic tax vs Mirrlees tax

Table 2.3: Welfare by Different Number of Grid Points

# of grid points	Welfare			
	Ramsey Linear	Ramsey Cubic	Mirrlees	First Best
2	14.40	-	36.84	43.92
15	12.99	-	26.46	36.97
100	15.78	16.00	19.53	58.60
1,000	15.79	16.00	16.48	58.60
10,000	15.79	16.00	16.13	58.62

agents, the planner can easily detect his deviation and hence induce truth-telling by giving him little compensation. Thus the lump-sum and linear taxes no longer approximate the second best allocation. This indicates that if underlying distribution is continuous but one uses coarse grid points to reduce the computational burden, then the implied welfare gains will be wrongly significant.

Shape of the Wage Distribution

Table 2.4 shows the results with alternative wage distributions. First, in contrast to the arguments in Saez (2001), the shape of the top of the productivity distribution is not quantitatively important; with the log-normal productivity distribution as in the literature, we still have a very high marginal rate and the approximation of Ramsey taxation to Mirrlees taxation in terms of welfare. This is also the case for a wider support for α. However, the planner should be raising tax rates as wage inequality increases. If the wage variance is a half of the baseline economy, then there is no need to levy a high tax rate because there is less demand for redistribution.

The last raw shows that the welfare gains are much smaller if the productivity is fully unobservable. This means that there are sizable welfare gains for the planner when a component of productivity is observable and the planner takes advantage of that information.

Preferences

Table 2.5 shows the results with different preferences. First, regarding the risk aversion, if the agents exhibit no risk aversion, $\gamma = 0$, then the planner has no motive for redistribution and just wants to finance G in the most efficient way. This quasi-linear preference in consumption results in the increase in output and hence higher welfare. In contrast, if the agents are more risk averse, $\gamma = 2$, the planner has strong incentive to redistribute and that leads much higher tax rates.

Table 2.4: Alternative Wage Distributions: Ramsey Linear vs Mirrlees

	Welfare		Output		Marg. Tax	
	Ramsey	Mirrlees	Ramsey	Mirrlees	Ramsey	Mirrlees
Baseline	15.79	16.48	-11.79	-11.39	0.506	0.496
Log-normal	14.87	15.47	-11.10	-10.34	0.490	0.474
Wider-support	16.15	17.16	-12.03	-11.54	0.511	0.500
$v_\alpha = v_\alpha/2$	9.71	10.60	-5.76	-5.31	0.380	0.363
$v_\varepsilon = 0$	12.28	12.90	-14.69	-14.36	0.534	0.526

Table 2.5: Alternative Preferences: Ramsey Linear vs Mirrlees

	Welfare		Output		Marg. Tax	
	Ramsey	Mirrlees	Ramsey	Mirrlees	Ramsey	Mirrlees
Baseline	15.79	16.48	-11.79	-11.39	0.506	0.496
$\gamma = 0$	1.04	1.16	7.53	8.77	0.058	0.037
$\gamma = 2$	29.05	29.89	-12.21	-11.82	0.620	0.610
$\sigma = 1$	15.24	15.95	-12.03	-11.41	0.450	0.439
$\sigma \to \infty$	30.94	30.94	0	0	1.00	1.00

Second, regarding the elasticity of labor supply, high Frisch elasticity, i.e., $\sigma = 1$ results in only slightly lower tax rates, inducing a larger level effect. In contrast, if we have inelastic labor supply, i.e., $\sigma \to \infty$, the planner can easily eliminate information friction. Then there are large welfare gains from equalizing consumption.

Social Welfare Function

Table 2.6 compares the outcome of Rawlsian social welfare function with the benchmark Utilitarian case. The equilibrium allocation and the forward iteration algorithm for Rawlsian social welfare function are found in Appendix A.

With the Rawlsian social welfare function, the tax rates, output losses and welfare gains are much higher than the benchmark. This is because the planner wants to redistribute the resources to the least productive agent as much as possible, inducing very high distortion to the economy. This result suggests the need of guidance on what social welfare functions to use in the analysis.

Table 2.6: Alternative Welfare Functions: Ramsey Linear vs Mirrlees

	Welfare		Output		Marg. Tax	
	Ramsey	Mirrlees	Ramsey	Mirrlees	Ramsey	Mirrlees
Baseline	15.79	16.48	-11.79	-11.39	0.506	0.496
Rawlsian	602.94	663.87	-30.27	-25.21	0.786	0.688

2.4.4 Conclusion

This chapter quantitatively studies the optimal income tax schedule, following both Mirrlees approach and Ramsey approach. We show that although the Mirrlees approach is theoretically appealing, very simple and realistic tax schemes following Ramsey approach can approximately implement the second best allocation.

This paper addresses two limitations of the Mirrlees literature. The first limitation is that in the literature it is assumed that agents' ability is fully private information. However, the productivity varies systematically with observable characteristics of workers such as education, age, gender and race, and we model this component and calibrate the relative variances of the observable and unobservable components by the wage regressions on the relative variances of between-group versus within-group wage dispersion. The second limitation is that the literature suggests highly non-linear optimal tax function, which is very different from the observed tax schedule in the U.S. This is one of the reasons why the insights of the literature have gained little attention from policy makers. We show that simple tax schemes in fact preform very well and this chapter would provide a useful communication devise with the policy makers.

In the work in progress, we are addressing other two limitations of the Mirrlees literature. One is that in the literature the tax policy is the only source of insurance and hence the resulting tax schedules are typically highly progress. We consider that a fraction of the unobservable shocks is privately insurable as is indicated in the data, and examine the properties of the optimal tax policy under such environments. The other limitation is that as shown in Section 2.4.3, the efficient tax schedule is sensitive to the choice of the social welfare functions, but we have almost no guide of how to choose one. We derive a formula in which the degree of the tax schedule is directly linked with social weights among agents and, assuming that the U.S. government is choosing the optimal progressiveness based on the social weights, we back out the underlying social weights named "empirically-motivated social welfare" Combining this with a newly calibrated model, we will reexamine the optimal income tax schedules.

References

AIYAGARI, R. (1994): "Uninsured idiosyncratic risk and aggregate saving," *Quarterly Journal of Economics*, 109, 659–684.

ATTANASIO, O., S. KITAO, AND G. L. VIOLANTE (2010): "Financing Medicare: A general equilibrium analysis," in *Demography and the Economy*, NBER Chapters, pp. 333–366. National Bureau of Economic Research, Inc.

BEWLEY, T. (1986): "Stationary monetary equilibrium with a continuum of Iidependently fluctuating consumer," in *Contributions to Mathematical Economics in Honor of Gerard Debreu*, ed. by W. Hildenbrand, and A. Mas-Colell, pp. 79–102. North-Holland.

BRGEMANN, B., AND I. MANOVSKII (2010): "Fragility: A quantitative analysis of the US health insurance systems," mimeo.

CHATTERJEE, S., D. CORBAE, M. NAKAJIMA, AND J.-V. ROS-RULL (2007): "A Quantitative Theory of Unsecured Consumer Credit with Risk of Default," *Econometrica*, 75(6), 1525–1589.

CLEMENS, J. (2012): "Regulatory redistribution in the market for health insurance," mimeo.

COLE, H. L., S. KIM, AND D. KRUEGER (2012): "Analyzing the effects of insuring health risks: On the tradeoff between short run insurance benefits vs. long run incentive costs," mimeo.

CONESA, J. C., S. KITAO, AND D. KRUEGER (2009): "Taxing Capital? Not a Bad Idea after All!," *American Economic Review*, 99(1), 25–48.

CUTLER, D. M., A. FINKELSTEIN, AND K. MCGARRY (2008): "Preference Heterogeneity and Insurance Markets: Explaining a Puzzle of Insurance," *American Economic Review*, 98(2), 157–62.

De Nardi, M., E. French, and J. B. Jones (2010): "Why do the elderly save? The role of medical expenses," *Journal of Political Economy*, 118(1), 39–75.

Diamond, P. A. (1998): "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates," *American Economic Review*, 88(1), 83–95.

Fukushima, K. (2010): "Quantifying the Welfare Gains From Flexible Dynamic Income Tax Systems," mimeo.

Glover, A., J. Heathcote, D. Krueger, and J.-V. Ros-Rull (2012): "Intergenerational redistribution in the Great Recession," mimeo.

Golosov, M., M. Troshkin, and A. Tsyvinski (2011): "Optimal Dynamic Taxes," mimeo.

Grossman, M. (1972): "On the concept of health capital and the demand for health," *Journal of Political Economy*, 80(2), 223–55.

Gruber, J. (2008): "Covering the uninsured in the United States," *Journal of Economic Literature*, 46(3), 571–606.

Hansen, G. (1993): "The cyclical and secular behaviour of the labour input: Comparing efficiency units and hours worked," *Journal of Applied Econometrics*, 8(1), pp. 71–80.

Hansen, G. D., M. Hsu, and J. Lee (2012): "Health insurance reform: The impact of a Medicare Buy-in," mimeo.

He, H., and K. Huangy (2012): "Why do Americans spend so much more in medical care than Europeans?," mimeo.

Heathcote, J., F. Perri, and G. L. Violante (2010): "Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States: 1967-2006," *Review of Economic Dynamics*, 13(1), 15–51.

Hubbard, R. G., J. Skinner, and S. P. Zeldes (1995): "Precautionary saving and social insurance," *Journal of Political Economy*, 103(2), 360–99.

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

Imrohoroglu, A. (1989): "Cost of business cycles with indivisibilities and liquidity constraints," *Journal of Political Economy*, 97, 1364–1383.

Imrohoroglu, S., and S. Kitao (2012): "Social Security reforms: Benefit claiming, labor force participation, and long-run sustainability," *American Economic Journal: Macroeconomics*, 4(3), 96–127.

Janicki, H. P. (2011): "Distributional Effects of Public Health Insurance Reform," mimeo.

Jeske, K., and S. Kitao (2009): "U.S. tax policy and health insurance demand: Can a regressive policy improve welfare?," *Journal of Monetary Economics*, 56(2), 210–221.

Jung, J., and C. Tran (2011): "Market inefficiency, insurance mandate and welfare: U.S. Health Care Reform 2010," mimeo.

Kathleen Stoll and Kim Bailey (2009): "Hidden health tax: Americans pay a premium," Report, Families USA.

Kopecky, K. A., and T. Koreshkova (2011): "The impact of medical and nursing home expenses and social insurance policies on savings and inequality," mimeo.

Kotlikoff, L. (1989): "Health expenditures and precautionary savings," in *What Determines Savings?* MIT Press.

Mankiw, N. G., and M. Weinzierl (2010): "The Optimal Taxation of Height: A Case Study of Utilitarian Income Redistribution," *American Economic Journal: Economic Policy*, 2(1), 155–76.

Mankiw, N. G., M. Weinzierl, and D. Yagan (2009): "Optimal Taxation in Theory and Practice," *Journal of Economic Perspectives*, 23(4), 147–74.

Mirrlees, J. A. (1971): "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 38(114), 175–208.

Pashchenko, S., and P. Porapakkarm (forthcoming): "Quantitative analysis of health insurance reform: Separating regulation from redistribution," *Review of Economic Dynamics*.

Piketty, T., and E. Saez (2003): "Income Inequality In The United States, 1913-1998," *The Quarterly Journal of Economics*, 118(1), 1–39.

Saez, E. (2001): "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 68(1), 205–29.

Smith, E., and R. Wright (1992): "Why Is Automobile Insurance in Philadelphia So Damn Expensive?," *American Economic Review*, 82(4), 756–72.

State Health Access Data Assistance Center (2012): "Defining "Family" for studies of health insurance coverage.," Issue brief 27, Minneapolis, MN: University of Minnesota.

Tuomala, M. (1984): "On the optimal income taxation: some further numerical results," *Journal of Public Economics*, 23, 351–366.

Weinzierl, M. (2011): "The Surprising Power of Age-Dependent Taxes," *Review of Economic Studies*, 78(4), 1490–1518.

Yogo, M. (2012): "Portfolio choice in retirement: Health risk and the demand for annuities, housing, and risky assets," mimeo.

Zhao, K. (2011): "Social security and the rise in health spending: a macroeconomic analysis," mimeo.

Appendix A

Solution with Rawlsian Social Welfare

A.1 First Best

The planner's problem:

$$\begin{cases} \max\limits_{c_{i,j}, y_{i,j}} & \min\limits_{i,j} u\left(c_{i,j}\right) - v\left(\frac{y_{i,j}}{\theta_{i,j}}\right) \\ \text{s.t.} & \sum \pi_{i,j} c_{i,j} + G \leq \sum \pi_{i,j} y_{i,j}. \end{cases}$$

This is equivalent to

$$\begin{cases} \max\limits_{c_{i,j}, h_{i,j}} & u\left(c_{1,1}\right) - v\left(\frac{y_{1,1}}{\theta_{1,1}}\right) \\ \text{s.t.} & \sum \pi_{i,j} c_{i,j} + G \leq \sum \pi_{i,j} y_{i,j}, \\ & u\left(c_{1,1}\right) - v\left(\frac{y_{1,1}}{\theta_{1,1}}\right) = u\left(c_{i,j}\right) - v\left(\frac{y_{i,j}}{\theta_{i,j}}\right), \text{ for all } (i,j) \neq (1,1). \end{cases}$$

It is straight forward to derive the system with $2N_\alpha N_\varepsilon$ unknowns, $\{c_{i,j}, y_{i,j}\}$ and $2N_\alpha N_\varepsilon$ equations:

$$\begin{cases} u'(c_{i,j}) = v'\left(\frac{y_{i,j}}{\theta_{i,j}}\right)\frac{1}{\theta_{i,j}} & : \quad N_\alpha N_\varepsilon \\ u(c_{1,1}) - v\left(\frac{y_{1,1}}{\theta_{1,1}}\right) = u(c_{i,j}) - v\left(\frac{y_{i,j}}{\theta_{i,j}}\right) & : \quad N_\alpha N_\varepsilon - 1 \\ \sum \pi_{i,j} c_{i,j} + G \leq \sum \pi_{i,j} y_{i,j} & : \quad 1 \end{cases}$$

A.2 Ramsey Allocation

The same argument applies to this case and thus we have

$$\begin{cases} \max_{\{T_\varepsilon\} \in \mathcal{T}} & \min_{i,j} u(c_{i,j}) - v\left(\frac{y_{i,j}}{\theta_{i,j}}\right) \\ \text{s.t.} & \{c_{i,j}, y_{i,j}\}_{i,j} \in \mathbb{E}(\{T_\varepsilon\}). \end{cases}$$

and the competitive equilibrium is a set of allocations $\{c_{i,j}, y_{i,j}\}$ such that it (i) solves agents' problem:

$$\begin{cases} \max_{c_{i,j}, h_{i,j}} & u(c_{i,j}) - v\left(\frac{y_{i,j}}{\theta_{i,j}}\right) \\ \text{s.t.} & c_{i,j} \leq y_{i,j} - T_{\varepsilon_j}(y_{i,j}). \end{cases}$$

and (ii) satisfies the feasibility condition: $\sum \pi_{i,j} c_{i,j} + G = \sum \pi_{i,j} y_{i,j}$. It is straightforward to derive the equilibrium conditions:

$$\begin{cases} u'(c_{i,j})\left(1 - T'_{\varepsilon_j}(y_{i,j})\right) = v'\left(\frac{y_{i,j}}{\theta_{i,j}}\right)\frac{1}{\theta_{i,j}}, \\ c_{i,j} = y_{i,j} - T_{\varepsilon_j}(y_{i,j}), \\ \sum \pi_{i,j} c_{i,j} + G = \sum \pi_{i,j} y_{i,j}. \end{cases}$$

A.3 Second Best Allocation

It is worth separating the problems by N_ε: the full private information (i.e., $N_\varepsilon = 1$) and the partial private information (i.e., $N_\varepsilon > 1$).

A.3.1 Full Private Information

Drop j for simplicity. The planner's problem:

$$\begin{cases} \max_{c_i,h_i} & \min_i \left[u\left(c_i\right) - v\left(h_i\right)\right] \\ \text{s.t.} & u\left(c_i\right) - v\left(h_i\right) \geq u\left(c_{i-1}\right) - v\left(h_{i-1}\frac{\theta_{i-1}}{\theta_i}\right) \quad \forall i = 2, \cdots, N_\alpha \\ & \sum_i \pi_i c_i + G \leq \sum_i \pi_i h_i \theta_i. \end{cases}$$

Claim The planner's problem above is equivalent to

$$\begin{cases} \max_{c_i,h_i} & u\left(c_1\right) - v\left(h_1\right) \\ \text{s.t.} & u\left(c_i\right) - v\left(h_i\right) \geq u\left(c_{i-1}\right) - v\left(h_{i-1}\frac{\theta_{i-1}}{\theta_i}\right) \quad \forall i = 2, \cdots, N_\alpha \\ & \sum_i \pi_i c_i + G \leq \sum_i \pi_i h_i \theta_i. \end{cases}$$

Proof It suffices to show that $\min_i \left[u\left(c_i\right) - v\left(h_i\right)\right] = u\left(c_1\right) - v\left(h_1\right)$. Suppose not. Then there exits $i \neq 1$ such that $u\left(c_i\right) - v\left(h_i\right) < u\left(c_1\right) - v\left(h_1\right)$. The local IC constraint tells $u\left(c_i\right) - v\left(h_i\right) \geq u\left(c_1\right) - v\left(h_1\frac{\theta_1}{\theta_i}\right)$. However this means that $u\left(c_1\right) - v\left(h_1\right) > u\left(c_1\right) - v\left(h_1\frac{\theta_1}{\theta_i}\right)$, which contradicts that v is strictly increasing. *Q.E.D.*

Then after putting the multipliers μ_i for the IC's and λ for the RC, we can take the standard FOCs:

$$\begin{cases} \text{w.r.t } c_{N_\alpha} & : & 0 = -\pi_{N_\alpha}\lambda + \mu_{N_\alpha} u'\left(c_{N_\alpha}\right) \\ \text{w.r.t. } h_{N_\alpha} & : & 0 = -\pi_{N_\alpha}\theta_{N_\alpha}\lambda + \mu_{N_\alpha} v'\left(h_{N_\alpha}\right) \\ \text{w.r.t. } c_i, i = 2, \cdots, N_\alpha - 1 & : & 0 = -\pi_i\lambda + \mu_i u'\left(c_i\right) - \mu_{i+1} u'\left(c_i\right) \\ \text{w.r.t. } h_i, i = 2, \cdots, N_\alpha - 1 & : & 0 = -\pi_i\theta_i\lambda + \mu_i v'\left(h_i\right) - \mu_{i+1} v'\left(h_i\frac{\theta_i}{\theta_{i+1}}\right)\frac{\theta_i}{\theta_{i+1}} \\ \text{w.r.t. } c_1 & : & 0 = u'\left(c_1\right) - \pi_1\lambda - \mu_2 u'\left(c_1\right) \\ \text{w.r.t. } h_1 & : & 0 = v'\left(h_1\right) - \pi_1\theta_1\lambda - \mu_2 v'\left(h_1\frac{\theta_1}{\theta_2}\right)\frac{\theta_1}{\theta_2} \end{cases}$$

$$\Leftrightarrow \begin{cases} \text{w.r.t } c_{N_\alpha} & : \quad \pi_{N_\alpha}\lambda = \mu_{N_\alpha} u'\left(c_{N_\alpha}\right) \\ \text{w.r.t. } h_{N_\alpha} & : \quad \pi_{N_\alpha}\theta_{N_\alpha}\lambda = \mu_{N_\alpha} v'\left(h_{N_\alpha}\right) \\ \text{w.r.t. } c_i, i = 2, \cdots, N_\alpha - 1 & : \quad \pi_i \lambda = u'\left(c_i\right)\left[\mu_{i,j} - \mu_{i+1}\right] \\ \text{w.r.t. } h_i, i = 2, \cdots, N_\alpha - 1 & : \quad \pi_i\theta_i\lambda = v'\left(h_i\right)\left[\mu_i - \mu_{i+1}\frac{v'\left(h_i\frac{\theta_i}{\theta_{i+1}}\right)}{v'(h_i)}\frac{\theta_i}{\theta_{i+1}}\right] \\ \text{w.r.t. } c_1 & : \quad \pi_1\lambda = u'\left(c_1\right)\left[1 - \mu_2\right] \\ \text{w.r.t. } h_1 & : \quad \pi_1\theta_1\lambda = v'\left(h_1\right)\left[1 - \mu_2\frac{v'\left(h_1\frac{\theta_1}{\theta_2}\right)}{v'(h_1)}\frac{\theta_1}{\theta_2}\right] \end{cases}$$

$$\Leftrightarrow \begin{cases} u'\left(c_{N_\alpha}\right)\theta_{N_\alpha} = v'\left(h_{N_\alpha}\right) \\ u'\left(c_i\right)\theta_i = v'\left(h_i\right)\frac{\mu_i - \mu_{i+1}\frac{v'\left(h_i\frac{\theta_i}{\theta_{i+1}}\right)}{v'(h_i)}\frac{\theta_i}{\theta_{i+1}}}{\mu_i - \mu_{i+1}} \\ u'\left(c_1\right)\theta_1 = v'\left(h_1\right)\frac{1 - \mu_2\frac{v'\left(h_1\frac{\theta_1}{\theta_2}\right)}{v'(h_1)}\frac{\theta_1}{\theta_2}}{1 - \mu_2} \end{cases}$$

A.3.2 Computation for the Full Private Economy

Assume the utility function is $u(c,h) = \frac{c^{1-\gamma}}{1-\gamma} - \varphi\frac{h^{1+\sigma}}{1+\sigma}$. The FOC w.r.t. c_i gives us

$$\begin{aligned} \pi_i\lambda &= c_i^{-\gamma}\left[\mu_i - \mu_{i+1}\right] \\ \Leftrightarrow \quad \mu_{i+1} &= \mu_i - \pi_i\lambda c_i^{\gamma}. \end{aligned} \tag{A.1}$$

Substitute this into the FOC w.r.t h_i

$$\begin{aligned} & \pi_i\theta_i\lambda = \varphi h_i^{\sigma}\left[\mu_i - \mu_{i+1}\left(\frac{\theta_i}{\theta_{i+1}}\right)^{1+\sigma}\right] \\ \Leftrightarrow \quad & \pi_i\theta_i\lambda = \varphi h_i^{\sigma}\left[\mu_i - \left(\mu_i - \pi_i\lambda c_i^{\gamma}\right)\left(\frac{\theta_i}{\theta_{i+1}}\right)^{1+\sigma}\right] \\ \Leftrightarrow \quad & h_i^{\sigma} = \frac{\pi_i\theta_i\lambda}{\varphi}\left[\mu_i - \left(\mu_i - \pi_i\lambda c_i^{\gamma}\right)\left(\frac{\theta_i}{\theta_{i+1}}\right)^{1+\sigma}\right]^{-1} \\ \Leftrightarrow \quad & h_i = \left(\frac{\pi_i\theta_i\lambda}{\varphi}\right)^{\frac{1}{\sigma}}\left\{\mu_i - \left(\mu_i - \pi_i\lambda c_i^{\gamma}\right)\left(\frac{\theta_i}{\theta_{i+1}}\right)^{1+\sigma}\right\}^{-\frac{1}{\sigma}}. \end{aligned} \tag{A.2}$$

The IC constraint for $i = 2, \cdots, N_\alpha - 1$ then tells us

$$\begin{aligned}
&\frac{c_i^{1-\gamma}}{1-\gamma} - \varphi \frac{h_i^{1+\sigma}}{1+\sigma} = \frac{c_{i-1}^{1-\gamma}}{1-\gamma} - \varphi \frac{\left(h_{i-1}\frac{\theta_{i-1}}{\theta_i}\right)^{1+\sigma}}{1+\sigma} \\
\Leftrightarrow\quad &\frac{c_i^{1-\gamma}}{1-\gamma} - \frac{\varphi}{1+\sigma}\left(\frac{\pi_i\theta_i\lambda}{\varphi}\right)^{\frac{1+\sigma}{\sigma}} \left\{\mu_i - (\mu_i - \pi_i\lambda c_i^\gamma)\left(\frac{\theta_i}{\theta_{i+1}}\right)^{1+\sigma}\right\}^{-\frac{1+\sigma}{\sigma}} \qquad \text{(A.3)} \\
&= \frac{c_{i-1}^{1-\gamma}}{1-\gamma} - \varphi \frac{\left(h_{i-1}\frac{\theta_{i-1}}{\theta_i}\right)^{1+\sigma}}{1+\sigma}.
\end{aligned}$$

Given c_{i-1}, h_{i-1}, μ_i and λ, this is a nonlinear equation for one unknown, c_i.

Algorithm for this problem**:**

1. Guess the value of the Lagrange multiplier for the resource constraint: $\lambda > 0$.

2. Choose $c_1 \in (c_{\min}, c_{\max})$.

3. Obtain μ_2 and h_1 from the FOCs:

$$\begin{aligned}
\mu_2 &= 1 - \pi_1\lambda c_1^\gamma, \\
h_1 &= \left(\frac{\pi_1\theta_1\lambda}{\varphi}\right)^{\frac{1}{\sigma}} \left[1 - \mu_2\left(\frac{\theta_1}{\theta_2}\right)^{1+\sigma}\right]^{\frac{-1}{\sigma}}.
\end{aligned}$$

4. For $i = 2, \cdots, N_\alpha - 1$, given c_{i-1}, h_{i-1} and μ_i, solve for c_i, h_i and μ_{i+1} from (A.3), (A.2) and (A.1), respectively.

5. Given μ_{N_α}, compute c_{N_α} and h_{N_α} from the FOCs:

$$\begin{aligned}
c_{N_\alpha} &= \left(\frac{\mu_{N_\alpha}}{\pi_{N_\alpha}\lambda}\right)^{\frac{1}{\gamma}}, \\
h_{N_\alpha} &= \left(\frac{\pi_{N_\alpha}\theta_{N_\alpha}\lambda}{\varphi\mu_{N_\alpha}}\right)^{\frac{1}{\sigma}}.
\end{aligned}$$

6. Check the incentive constraint for the type N_α agent:

$$\frac{c_{N_\alpha}^{1-\gamma}}{1-\gamma} - \varphi\frac{h_{N_\alpha}^{1+\sigma}}{1+\sigma} = \frac{c_{N_\alpha-1}^{1-\gamma}}{1-\gamma} - \varphi\frac{\left(h_{N_\alpha-1}\frac{\theta_{N_\alpha-1}}{\theta_{N_\alpha}}\right)^{1+\sigma}}{1+\sigma}.$$

If not satisfied, then go back to step 2 and adjust c_1.

7. Check if the aggregate resource constraint is satisfied:

$$\sum_i \pi_i c_i + G = \sum_i \pi_i h_i \theta_i.$$

If not, go back to step 1 and adjust λ. Repeat until this is satisfied under a certain tolerance.

How do we choose $c_{\min}$ and $c_{\max}$? It is natural to assume $c_{\min} = 0$. In addition, we know that $\mu_2 = 1 - \pi_1 \lambda c_1^\gamma > 0$ at optimum. This implies that $\pi_1 \lambda c_1^\gamma < 1$, leading $c_{\max} = \left(\frac{1}{\pi_1 \lambda}\right)^{1/\gamma}$.

A.3.3 Partial Private Information

The planner's problem:

$$\begin{cases} \max\limits_{c_{i,j}, h_{i,j}} & \min\limits_{i,j} \left[u\left(c_{i,j}\right) - v\left(h_{i,j}\right)\right] \\ \text{s.t.} & u\left(c_{i,j}\right) - v\left(h_{i,j}\right) \geq u\left(c_{i-1,j}\right) - v\left(h_{i-1,j}\frac{\theta_{i-1,j}}{\theta_{i,j}}\right) \quad \forall j, i = 2, \cdots, N_\alpha \\ & \sum \pi_{i,j} c_{i,j} + G \leq \sum \pi_{i,j} h_{i,j} \theta_{i,j}. \end{cases}$$

Since the analogue of the Claim in the full private information economy can be applied here, we get

$$\begin{cases} \max\limits_{c_{i,j}, h_{i,j}} & \min\limits_{j} \left[u\left(c_{1,j}\right) - v\left(h_{1,j}\right)\right] \\ \text{s.t.} & u\left(c_{i,j}\right) - v\left(h_{i,j}\right) \geq u\left(c_{i-1,j}\right) - v\left(h_{i-1,j}\frac{\theta_{i-1,j}}{\theta_{i,j}}\right) \quad \forall j, i = 2, \cdots, N_\alpha \\ & \sum \pi_{i,j} c_{i,j} + G \leq \sum \pi_{i,j} h_{i,j} \theta_{i,j}. \end{cases}$$

This is equivalent to

$$\begin{cases} \max\limits_{c_{i,j}, h_{i,j}} & u\left(c_{1,1}\right) - v\left(h_{1,1}\right) \\ \text{s.t.} & u\left(c_{i,j}\right) - v\left(h_{i,j}\right) \geq u\left(c_{i-1,j}\right) - v\left(h_{i-1,j}\frac{\theta_{i-1,j}}{\theta_{i,j}}\right), \quad \forall j, i = 2, \cdots, N_\alpha & : \ \mu_{i,j} \\ & \sum \pi_{i,j} c_{i,j} + G \leq \sum \pi_{i,j} h_{i,j} \theta_{i,j}, & : \ \lambda \\ & u\left(c_{1,1}\right) - v\left(h_{1,1}\right) = u\left(c_{1,j}\right) - v\left(h_{1,j}\right), \text{ for all } j \neq 1. & : \ \kappa_j \end{cases}$$

Note that we have additional constraints, which we call as Rawlsian constraints.

Then we can take the standard FOCs:

$$\begin{cases}
\text{w.r.t } c_{N_\alpha,j} & : \quad 0 = -\pi_{N_\alpha,j}\lambda + \mu_{N_\alpha,j}u'\left(c_{N_\alpha,j}\right) \\
\text{w.r.t. } h_{N_\alpha,j} & : \quad 0 = -\pi_{N_\alpha,j}\theta_{N_\alpha,j}\lambda + \mu_{N_\alpha,j}v'\left(h_{N_\alpha,j}\right) \\
\text{w.r.t. } c_{i,j}, i = 2,\cdots,N_\alpha - 1 & : \quad 0 = -\pi_{i,j}\lambda + \mu_{i,j}u'\left(c_{i,j}\right) - \mu_{i+1,j}u'\left(c_{i,j}\right) \\
\text{w.r.t. } h_{i,j}, i = 2,\cdots,N_\alpha - 1 & : \quad 0 = -\pi_{i,j}\theta_{i,j}\lambda + \mu_{i,j}v'\left(h_{i,j}\right) - \mu_{i+1,j}v'\left(h_{i,j}\frac{\theta_{i,j}}{\theta_{i+1,j}}\right)\frac{\theta_{i,j}}{\theta_{i+1,j}} \\
\text{w.r.t. } c_{1,1} & : \quad 0 = u'\left(c_{1,1}\right) - \pi_{1,1}\lambda - \mu_{2,1}u'\left(c_{1,1}\right) - \sum_{j\neq 1}\kappa_j u'\left(c_{1,1}\right) \\
\text{w.r.t. } h_{1,1} & : \quad 0 = v'\left(h_{1,1}\right) - \pi_{1,1}\theta_{1,1}\lambda - \mu_{2,1}v'\left(h_{1,1}\frac{\theta_{1,1}}{\theta_{2,1}}\right)\frac{\theta_{1,1}}{\theta_{2,1}} - \sum_{j\neq 1}\kappa_j v'\left(h_{1,1}\right) \\
\text{w.r.t. } c_{1,j}, j \neq 1 & : \quad 0 = -\pi_{1,j}\lambda - \mu_{2,j}u'\left(c_{1,j}\right) + \kappa_j u'\left(c_{1,j}\right) \\
\text{w.r.t. } h_{1,j}, j \neq 1 & : \quad 0 = -\pi_{1,j}\theta_{1,j}\lambda - \mu_{2,j}v'\left(h_{1,j}\frac{\theta_{1,j}}{\theta_{2,j}}\right)\frac{\theta_{1,j}}{\theta_{2,j}} + \kappa_j v'\left(h_{1,j}\right)
\end{cases}$$

$$\Leftrightarrow \begin{cases}
\text{w.r.t } c_{N_\alpha,j} & : \quad \pi_{N_\alpha,j}\lambda = \mu_{N_\alpha,j}u'\left(c_{N_\alpha,j}\right) \\
\text{w.r.t. } h_{N_\alpha,j} & : \quad \pi_{N_\alpha,j}\theta_{N_\alpha,j}\lambda = \mu_{N_\alpha,j}v'\left(h_{N_\alpha,j}\right) \\
\text{w.r.t. } c_{i,j}, i = 2,\cdots,N_\alpha - 1 & : \quad \pi_{i,j}\lambda = u'\left(c_{i,j}\right)\left[\mu_{i,j} - \mu_{i+1,j}\right] \\
\text{w.r.t. } h_{i,j}, i = 2,\cdots,N_\alpha - 1 & : \quad \pi_{i,j}\theta_{i,j}\lambda = v'\left(h_{i,j}\right)\left[\mu_{i,j} - \mu_{i+1,j}\frac{v'\left(h_{i,j}\frac{\theta_{i,j}}{\theta_{i+1,j}}\right)}{v'(h_{i,j})}\frac{\theta_{i,j}}{\theta_{i+1,j}}\right] \\
\text{w.r.t. } c_{1,1} & : \quad \pi_{1,1}\lambda = u'\left(c_{1,1}\right)\left[1 - \mu_{2,1} - \sum_{j\neq 1}\kappa_j\right] \\
\text{w.r.t. } h_{1,1} & : \quad \pi_{1,1}\theta_{1,1}\lambda = v'\left(h_{1,1}\right)\left[1 - \mu_{2,1}\frac{v'\left(h_{1,1}\frac{\theta_{1,1}}{\theta_{2,1}}\right)}{v'(h_{1,1})}\frac{\theta_{1,1}}{\theta_{2,1}} - \sum_{j\neq 1}\kappa_j\right] \\
\text{w.r.t. } c_{1,j}, j \neq 1 & : \quad \pi_{1,j}\lambda = u'\left(c_{1,j}\right)\left[-\mu_{2,j} + \kappa_j\right] \\
\text{w.r.t. } h_{1,j}, j \neq 1 & : \quad \pi_{1,j}\theta_{1,j}\lambda = v'\left(h_{1,j}\right)\left[-\mu_{2,j}\frac{v'\left(h_{1,j}\frac{\theta_{1,j}}{\theta_{2,j}}\right)}{v'(h_{1,j})}\frac{\theta_{1,j}}{\theta_{2,j}} + \kappa_j\right]
\end{cases}$$

$$\Leftrightarrow \begin{cases} u'(c_{N_\alpha,j})\,\theta_{N_\alpha,j} = v'(h_{N_\alpha,j}) \\ u'(c_{i,j})\,\theta_{i,j} = v'(h_{i,j})\,\dfrac{\mu_{i,j}-\mu_{i+1,j}\frac{v'\left(h_{i,j}\frac{\theta_{i,j}}{\theta_{i+1,j}}\right)}{v'(h_{i,j})}\frac{\theta_{i,j}}{\theta_{i+1,j}}}{\mu_{i,j}-\mu_{i+1,j}} \\ u'(c_{1,1})\,\theta_{1,1} = v'(h_{1,1})\,\dfrac{1-\mu_{2,1}\frac{v'\left(h_{1,1}\frac{\theta_{1,1}}{\theta_{2,1}}\right)}{v'(h_{1,1})}\frac{\theta_{1,1}}{\theta_{2,1}}-\sum_{j\neq 1}\kappa_j}{1-\mu_{2,1}-\sum_{j\neq 1}\kappa_j} \\ u'(c_{1,j})\,\theta_{1,j} = v'(h_{1,j})\,\dfrac{-\mu_{2,j}\frac{v'\left(h_{1,j}\frac{\theta_{1,j}}{\theta_{2,j}}\right)}{v'(h_{1,j})}\frac{\theta_{1,j}}{\theta_{2,j}}+\kappa_j}{-\mu_{2,j}+\kappa_j} \end{cases} \tag{A.4}$$

We can regard this as a system of equations: $3N_\alpha N_\varepsilon$ unknowns:

$$\begin{cases} 2N_\alpha N_\varepsilon & : & \{c_{i,j}, h_{i,j}\}_{i=1,\cdots,N_\alpha,\ j=1,\cdots,N_\varepsilon} \\ (N_\alpha-1)\times N_\varepsilon & : & \{\mu_{i,j}\}_{i=2,\cdots,N_\alpha,\ j=1,\cdots,N_\varepsilon} \\ 1 & : & \lambda \\ N_\varepsilon - 1 & : & \{\kappa_j\}_{j=2,\cdots,N_\varepsilon} \end{cases}$$

and $3N_\alpha N_\varepsilon$ equations:

$$\begin{cases} N_\alpha N_\varepsilon & : & \text{Euler equation (A.4)} \\ N_\alpha N_\varepsilon & : & \text{FOC w.r.t. } c \\ (N_\alpha-1)\times N_\varepsilon & : & \text{IC with equality} \\ 1 & : & \text{RC} \\ N_\varepsilon - 1 & : & \text{Rawlsian constraint} \end{cases}$$

Thus this system is solvable.

A.3.4 Computation for the Partial Private Economy

Assume the utility function is $u(c,h) = \frac{c^{1-\gamma}}{1-\gamma} - \varphi\frac{h^{1+\sigma}}{1+\sigma}$. We want to drop κ's. The FOC w.r.t. $c_{i,j}$ gives us

$$\begin{aligned} \pi_{i,j}\lambda &= c_{i,j}^{-\gamma}\left[\mu_{i,j}-\mu_{i+1,j}\right] \\ \Leftrightarrow \quad \mu_{i+1,j} &= \mu_{i,j} - \pi_{i,j}\lambda c_{i,j}^{\gamma}. \end{aligned} \tag{A.5}$$

Substitute this into the FOC w.r.t $h_{i,j}$

$$\begin{aligned}
&\pi_{i,j}\theta_{i,j}\lambda = \varphi h_{i,j}^{\sigma}\left[\mu_{i,j} - \mu_{i+1,j}\left(\frac{\theta_{i,j}}{\theta_{i+1,j}}\right)^{1+\sigma}\right]\\
\Leftrightarrow\quad &\pi_{i,j}\theta_{i,j}\lambda = \varphi h_{i,j}^{\sigma}\left[\mu_{i,j} - \left(\mu_{i,j} - \pi_{i,j}\lambda c_{i,j}^{\gamma}\right)\left(\frac{\theta_{i,j}}{\theta_{i+1,j}}\right)^{1+\sigma}\right]\\
\Leftrightarrow\quad &h_{i,j}^{\sigma} = \frac{\pi_{i,j}\theta_{i,j}\lambda}{\varphi}\left[\mu_{i,j} - \left(\mu_{i,j} - \pi_{i,j}\lambda c_{i,j}^{\gamma}\right)\left(\frac{\theta_{i,j}}{\theta_{i+1,j}}\right)^{1+\sigma}\right]^{-1}\\
\Leftrightarrow\quad &h_{i,j} = \left(\frac{\pi_{i,j}\theta_{i,j}\lambda}{\varphi}\right)^{\frac{1}{\sigma}}\left\{\mu_{i,j} - \left(\mu_{i,j} - \pi_{i,j}\lambda c_{i,j}^{\gamma}\right)\left(\frac{\theta_{i,j}}{\theta_{i+1,j}}\right)^{1+\sigma}\right\}^{-\frac{1}{\sigma}}.
\end{aligned} \tag{A.6}$$

The IC constraint for $i = 2, \cdots, N_\alpha - 1$ then tells us

$$\begin{aligned}
&\frac{c_{i,j}^{1-\gamma}}{1-\gamma} - \varphi\frac{h_{i,j}^{1+\sigma}}{1+\sigma} = \frac{c_{i-1,j}^{1-\gamma}}{1-\gamma} - \varphi\frac{\left(h_{i-1,j}\frac{\theta_{i-1,j}}{\theta_{i,j}}\right)^{1+\sigma}}{1+\sigma}\\
\Leftrightarrow\quad &\frac{c_{i,j}^{1-\gamma}}{1-\gamma} - \frac{\varphi}{1+\sigma}\left(\frac{\pi_{i,j}\theta_{i,j}\lambda}{\varphi}\right)^{\frac{1+\sigma}{\sigma}}\left\{\mu_{i,j} - \left(\mu_{i,j} - \pi_{i,j}\lambda c_{i,j}^{\gamma}\right)\left(\frac{\theta_{i,j}}{\theta_{i+1,j}}\right)^{1+\sigma}\right\}^{-\frac{1+\sigma}{\sigma}}\\
&\qquad = \frac{c_{i-1,j}^{1-\gamma}}{1-\gamma} - \varphi\frac{\left(h_{i-1,j}\frac{\theta_{i-1,j}}{\theta_{i,j}}\right)^{1+\sigma}}{1+\sigma}.
\end{aligned} \tag{A.7}$$

Given $c_{i-1,j}, h_{i-1,j}, \mu_{i,j}$ and λ, this is a nonlinear equation for one unknown, $c_{i,j}$.

Algorithm for this problem**:**

1. Guess the value of the Lagrange multiplier for the resource constraint: $\lambda > 0$.
2. Choose $h_{1,1} \in (h_{\min}, h_{\max})$.
3. Choose $c_{1,1}$ so that the implied sequence satisfies all the FOC's and IC's:
   i. Choose $c_{1,1} \in (c_{\min}, c^1_{\max})$.
   ii. Obtain $\mu_{2,1}$ from the FOC:

$$\frac{\pi_{1,1}\lambda}{u'(c_{1,1})} + \mu_{2,1} = \frac{\pi_{1,1}\theta_{1,1}\lambda}{v'(h_{1,1})} + \mu_{2,1}\frac{v'\left(h_{1,1}\frac{\theta_{1,1}}{\theta_{2,1}}\right)}{v'(h_{1,1})}\frac{\theta_{1,1}}{\theta_{2,1}}$$

$$\Leftrightarrow\quad \mu_{2,1} = \frac{\pi_{1,1}\theta_{1,1}\lambda\varphi^{-1}h_{1,1}^{-\sigma} - \pi_{1,1}\lambda c_{1,1}^{\gamma}}{1-\left(\frac{\theta_{1,1}}{\theta_{2,1}}\right)^{1+\sigma}}.$$

   iii. For $i = 2, \cdots, N_\alpha - 1$, given $c_{i-1,1}$, $h_{i-1,1}$ and $\mu_{i,1}$, solve for $c_{i,1}, h_{i,1}$ and $\mu_{i+1,1}$ from (A.7), (A.6) and (A.5), respectively.

iv. Given $\mu_{N_\alpha,1}$, compute $c_{N_\alpha,1}$ and $h_{N_\alpha,1}$ from the FOCs:

$$\begin{aligned} c_{N_\alpha,1} &= \left(\frac{\mu_{N_\alpha,1}}{\pi_{N_\alpha,1}\lambda}\right)^{\frac{1}{\gamma}}, \\ h_{N_\alpha,1} &= \left(\frac{\pi_{N_\alpha,1}\theta_{N_\alpha,1}\lambda}{\varphi\mu_{N_\alpha,1}}\right)^{\frac{1}{\sigma}}. \end{aligned}$$

v. Check the incentive constraint for the type N_α agent:

$$\frac{c_{N_\alpha,1}^{1-\gamma}}{1-\gamma} - \varphi\frac{h_{N_\alpha,1}^{1+\sigma}}{1+\sigma} = \frac{c_{N_\alpha-1,1}^{1-\gamma}}{1-\gamma} - \varphi\frac{\left(h_{N_\alpha-1,1}\frac{\theta_{N_\alpha-1,1}}{\theta_{N_\alpha,1}}\right)^{1+\sigma}}{1+\sigma}.$$

If not satisfied, then go back to step 3-i and adjust $c_{1,1}$. If $c_{1,1} = c_{\min}$ or $c_{\max}^1$, then go back to step 2 and adjust $h_{1,1}$.

4. For each $j = 2, \cdots, N_\varepsilon$, choose $c_{1,j}$ so that the implied sequence satisfies all the FOC's and IC's:

i. Choose $c_{1,j} \in (c_{\min}, c_{\max}^j)$.

ii. Obtain $h_{1,j}$ from the Rawlsian constraint:

$$\begin{aligned} \frac{c_{1,1}^{1-\gamma}}{1-\gamma} - \varphi\frac{h_{1,1}^{1+\sigma}}{1+\sigma} &= \frac{c_{1,j}^{1-\gamma}}{1-\gamma} - \varphi\frac{h_{1,j}{}^{1+\sigma}}{1+\sigma} \\ \Leftrightarrow \quad h_{1,j} &= \left[\frac{1+\sigma}{\varphi}\left(\frac{c_{1,j}^{1-\gamma}}{1-\gamma} - \frac{c_{1,1}^{1-\gamma}}{1-\gamma} + \varphi\frac{h_{1,1}^{1+\sigma}}{1+\sigma}\right)\right]^{\frac{1}{1+\sigma}}. \end{aligned}$$

If $h_{1,j} < 0$, then go back to step 4-i and raise $c_{1,j}$. If $h_{1,j} > 0$ and $c_{1,j} < c_{\min}$ or $c_{1,j} > c_{\max}^j$, then go back to step 2 and adjust $h_{1,1}$.

iii. Obtain $\mu_{2,j}$ from the FOC:

$$\begin{aligned} \frac{\pi_{1,j}\lambda}{u'(c_{1,j})} + \mu_{2,j} &= \frac{\pi_{1,j}\theta_{1,j}\lambda}{v'(h_{1,j})} + \mu_{2,j}\frac{v'\left(h_{1,j}\frac{\theta_{1,j}}{\theta_{2,j}}\right)}{v'(h_{1,j})}\frac{\theta_{1,j}}{\theta_{2,j}} \\ \Leftrightarrow \quad \mu_{2,j} &= \frac{\pi_{1,j}\theta_{1,j}\lambda\varphi^{-1}h_{1,j}^{-\sigma} - \pi_{1,j}\lambda c_{1,j}^{\gamma}}{1-\left(\frac{\theta_{1,j}}{\theta_{2,j}}\right)^{1+\sigma}}. \end{aligned}$$

If $\mu_{2,j} \leq 0$, then go back to step 4-i and adjust $c_{1,j}$.

iv. For $i = 2, \cdots, N_\alpha - 1$, given $c_{i-1,j}$, $h_{i-1,j}$ and $\mu_{i,j}$, solve for $c_{i,j}, h_{i,j}$ and $\mu_{i+1,j}$ from (A.7), (A.6) and (A.5), respectively.

v. Given $\mu_{N_\alpha,j}$, compute $c_{N_\alpha,j}$ and $h_{N_\alpha,j}$ from the FOCs:

$$\begin{aligned} c_{N_\alpha,j} &= \left(\frac{\mu_{N_\alpha,j}}{\pi_{N_\alpha,j}\lambda}\right)^{\frac{1}{\gamma}}, \\ h_{N_\alpha,j} &= \left(\frac{\pi_{N_\alpha,j}\theta_{N_\alpha,j}\lambda}{\varphi\mu_{N_\alpha,j}}\right)^{\frac{1}{\sigma}}. \end{aligned}$$

vi. Check the incentive constraint for the type N_α agent for each $j = 1, \cdots, N_\varepsilon$:

$$\frac{c_{N_\alpha,j}^{1-\gamma}}{1-\gamma} - \varphi\frac{h_{N_\alpha,j}^{1+\sigma}}{1+\sigma} = \frac{c_{N_\alpha-1,j}^{1-\gamma}}{1-\gamma} - \varphi\frac{\left(h_{N_\alpha-1,j}\frac{\theta_{N_\alpha-1,j}}{\theta_{N_\alpha,j}}\right)^{1+\sigma}}{1+\sigma}.$$

If not satisfied, then go back to step 4-i and adjust $c_{1,j}$. If $c_{1,j} = c_{\min}$ or $c^j_{\max}$, then go back to step 2 and adjust $h_{1,1}$.

5. Check the Euler equation is truly satisfied: (derived from the FOC w.r.t. $h_{1,j}$)

$$1 = \sum_j \left[\pi_{1,j}\theta_{1,j}\lambda\varphi^{-1}h_{1,j}^{-\sigma} + \mu_{2,j}\left(\frac{\theta_{1,j}}{\theta_{2,j}}\right)^{1+\sigma}\right].$$

If not, go back to step 2 and adjust $h_{1,1}$.

6. Check if the aggregate resource constraint is satisfied:

$$\sum_{(i,j)} \pi_{i,j}c_{i,j} + G = \sum_{(i,j)} \pi_{i,j}h_{i,j}\theta_{i,j}.$$

If not, go back to step 1 and adjust λ. Repeat until this is satisfied under a certain tolerance.

How do we choose $h_{\min}$ and $h_{\max}$? We know from the FOC that $\mu_{2,1}\left(\frac{\theta_{1,1}}{\theta_{2,1}}\right)^{1+\sigma} = 1 - \frac{\pi_{1,1}\theta_{1,1}\lambda}{\varphi}h_{1,1}^{-\sigma} - \sum_{j\neq 1}\kappa_j > 0$ at optimum. Since $\kappa_j \geq 0$, we have $\frac{\pi_{1,1}\theta_{1,1}\lambda}{\varphi}h_{1,1}^{-\sigma} < 1 - \sum_{j\neq 1}\kappa_j \leq 1$. This implies that $h_{1,1} > \left(\frac{\pi_{1,1}\theta_{1,1}\lambda}{\varphi}\right)^{1/\sigma}$, leading $h_{\min} = \left(\frac{\pi_{1,1}\theta_{1,1}\lambda}{\varphi}\right)^{1/\sigma}$. Also from the FOC w.r.t. $c_{1,1}$, we know that $\pi_{1,1}\lambda c_{1,1}^{\gamma} = 1 - \mu_{2,1} - \sum_{j\neq 1}\kappa_j > 0$. Thus we have $1 - \sum_{j\neq 1}\kappa_j > \mu_{2,1}$. It follows from the FOC w.r.t. $h_{1,1}$ that $\left(\frac{\pi_{1,1}\theta_{1,1}\lambda}{\varphi}\right)h_{1,1}^{-\sigma} + \mu_{2,1}\left(\frac{\theta_{1,1}}{\theta_{2,1}}\right)^{1+\sigma} = 1 - \sum_{j\neq 1}\kappa_j > \mu_{2,1}$. That is, $\left(\frac{\pi_{1,1}\theta_{1,1}\lambda}{\varphi}\right)h_{1,1}^{-\sigma} > \mu_{2,1}\left[1 - \left(\frac{\theta_{1,1}}{\theta_{2,1}}\right)^{1+\sigma}\right]$. This implies that $h_{1,1} < \left[\frac{\pi_{1,1}\theta_{1,1}\lambda}{\varphi\left(1-\left(\frac{\theta_{1,1}}{\theta_{2,1}}\right)^{1+\sigma}\right)}\right]^{1/\sigma}$, leading

$h_{\max} = \left[\frac{\pi_{1,1}\theta_{1,1}\lambda}{\varphi\left(1-\left(\frac{\theta_{1,1}}{\theta_{2,1}}\right)^{1+\sigma}\right)} \right]^{1/\sigma}$.

Next, how do we choose $c_{\min}$ and $c^j_{\max}$? It is natural to assume $c_{\min} = 0$. Also we know from the FOC that $\mu_{2,1} = 1 - \pi_{1,1}\lambda c^{\gamma}_{1,1} - \sum\limits_{j\neq 1} \kappa_j > 0$ at optimum. Since $\kappa_j \geq 0$, we have $\pi_{1,1}\lambda c^{\gamma}_{1,1} < 1 - \sum\limits_{j\neq 1} \kappa_j \leq 1$. This implies that $\pi_{1,1}\lambda c^{\gamma}_{1,1} < 1$. Furthermore, we need to guarantee that $\mu_{2,1} > 0$. Thus we have $\pi_{1,1}\theta_{1,1}\lambda h^{-\sigma}_{1,1} - \pi_{1,1}\lambda c^{\gamma}_{1,1} > 0$. This implies $c^{\gamma}_{1,1} < \theta_{1,1}h^{-\sigma}_{1,1}$. In summary, $c^1_{\max} = \min\left\{ \left(\frac{1}{\pi_{1,1}\lambda}\right)^{1/\gamma}, \left(\theta_{1,1}h^{-\sigma}_{1,1}\right)^{1/\gamma} \right\}$. Similarly, we know from the FOC that $\mu_{2,j} = \kappa_j - \pi_{1,j}\lambda c^{\gamma}_{1,j} > 0$ at optimum. Since $\kappa_j \leq 1$, we have $\pi_{1,j}\lambda c^{\gamma}_{1,j} < \kappa_j \leq 1$. This implies that $\pi_{1,j}\lambda c^{\gamma}_{1,j} < 1$. $c^j_{\max} = \left(\frac{1}{\pi_{1,j}\lambda}\right)^{1/\gamma}$.

Appendix B

Gains of Forward Iteration Algorithm

The algorithm described in the section 2.3.3 has gains compared with existing approaches.

The algorithm discretizes the productivity distribution and solves the allocation directly by forward iteration. Discretization enables us to examine how the solution is sensitive to the number of the grid points. This is especially informative because in the context of the dynamic Mirrlees economy, it is typical to use very coarse discretization (for example Golosov, Troshkin, and Tsyvinski (2011) and Fukushima (2010)).

Our algorithm is faster and more precise than that in Mankiw, Weinzierl, and Yagan (2009) (MWY hereafter). They discretize the wage distribution and approximate the optimal marginal tax equation given by Saez (2001). Specifically they construct a piecewise linear tax function and solve agents' maximization problem given that tax system in order to obtain the new allocation and hence a new tax system, iterating this until convergence. However, using the Saez's equation with the discrete type has fundamental flaws; with coarse grid points, the approximation error generated from the approximation of the CDF and the numerical integration is significant. Moreover, their algorithm requires to solve agents' maximization problem, which is computationally expensive. Instead, our forward iteration directly solves the allocation and hence faster. Figure B.1 compares our forward iteration algorithm (labeled by HT) with MWY's algorithm in three ways; using MWY distribution (Exercise 1), using our distribution with the same number of grids as MWY (Exercise 2), and using our distribution with finer grids (Exercise 3). We use the MWY preference for this comparison. The results in Figure B.1 show that our algorithm is faster and more precise than MWY.

Lastly, since our forward iteration algorithm is able to directly solve the allocation, it allows to solve for the allocation with Rawlsian welfare function as well (Appendix A). In contrast, the Saez's equation involves the derivative of the social welfare function. This requires to use general form of the social welfare function as suggested in Tuomala (1984):

$$W(u) = -\frac{1}{\beta} e^{-\beta u},$$

where the Utilitarian corresponds to $\beta = 0$ and the Rawlsian corresponds to $\beta = \infty$. Rawlsian welfare is not differentiable, however, and the approximation by the above function is inaccurate even when letting β be a very large number.

```
                         comparison.txt
<<<<<  Comparison : Simulation Success  >>>>>
(1a)  Is y(w) non-deacreasing?  "1" means yes.
(1b)  IC violation: Max gain from truth-telling
      [util from truth-telling] - [util from
misrepresenting]
(1c)  IC violation: Max gain from misrepresenting
      [util from misrepresenting] - [util from
truth-telling]
(1d)  Gov budget imbalance as a percent of GNI
(1e)  Max % deviation from "optimal" tax schedule
--------------------------------------------------
(2a)  Utilitarian Social Welfare
(2b)  Welfare Loss from the First Best (%,CEquiv.)
(2c)  Min optimal marginal tax rate
(2d)  Max optimal marginal tax rate
(2e)  Min net tax payment
(2f)  Max net tax payment
--------------------------------------------------
(3a)  Total number loops run
(3b)  Execution time (sec)

Exercise 1 : Use MWY's Distribution
Number of grid points =   144
          MWY                     HT
1a      1                       1
1b      0.0866461934796199      0.0000000000010749
1c     -0.0001222833000940      0.0000000000547085
1d      0.0000002937641046      0.0000000002081941
1e      0.0000089662332969      N/A
--------------------------------------------------
2a     -0.8575535286915217     -0.8004860370110234
2b    -29.0222710486049210    -25.6739478962303980
2c      0.0000000013038298      0.0000000000000003
2d      0.9992229212793968      0.7437125954715282
2e     -0.3619088568815800     -0.4143288323584117
2f      8.1187379589751831      8.1710796226604394
--------------------------------------------------
3a     28                       6
3b     33.3895968180451130     19.6148330494657690

                           Page 1
```

```
                         comparison.txt
Exercise 2 : Use HT's Distribution
Number of grid points =   146
          MWY                     HT
1a      1                       1
1b      0.0112015059332331      0.0000000000012418
1c     -0.0001630816111493      0.0000000000019369
1d      0.0000007897881085      0.0000000002036506
1e      0.0000091361553192      N/A
--------------------------------------------------
2a     -0.8505836355601993     -0.8355835493690185
2b    -25.7029229976423640    -24.8005365219014920
2c      0.0000000052154064     -0.0000000000000002
2d      0.6529944414419528      0.6410869320090952
2e     -0.3057590371118604     -0.3128802233367188
2f      1.9711222673047919      1.9879134357345816
--------------------------------------------------
3a     26                       9
3b     31.5106540359319350     32.0449511730906240


Exercise 3 : Use HT's Distribution with Finer Grids
Number of grid points =   1002
          MWY                     HT
1a      1                       1
1b      0.0002485017376224      0.0000000000021261
1c     -0.0000034287690104      0.0000000001736590
1d      0.0000007775372737      0.0000000001046033
1e      0.0000082488833463      N/A
--------------------------------------------------
2a     -0.8510362519386500     -0.8488509456238640
2b    -25.7934433159702540    -25.6631818894223470
2c      0.0000000052154183      0.0000000000000001
2d      0.6528366121410363      0.6510805183733804
2e     -0.2870375572285412     -0.2879824361065462
2f      1.9809398164898520      1.9831661770866944
--------------------------------------------------
3a     26                       6
3b     210.0862538406015000     135.5168382948951400

                           Page 2
```

Figure B.1: Gain of the Algorithm: Comparison with Mankiw, Weinzierl, and Yagan (2009)