

Heterogeneity in Labor Supply Elasticity and Optimal Taxation

Marios Karabarbounis *

January 11, 2012

Job Market Paper

Abstract

Standard public finance principles imply that workers with more elastic labor supply should face smaller tax distortions. This paper quantitatively tests the potential of such an idea within a realistically calibrated life cycle model of labor supply with heterogeneous agents and incomplete markets. Heterogeneity in labor supply elasticity arises endogenously from differences in reservation wages. I find that older cohorts are much more responsive to wage changes than younger and especially middle aged cohorts. Both a shorter time horizon and a larger stock of savings account for this difference. Since the government does not have direct information on individual labor supply elasticity it uses these life cycle variables as informative moments. The optimal Ramsey tax policy decreases the average and marginal tax rates for agents older than 50 and more so the larger is the accumulated stock of savings. At the same time, the policy increases significantly the tax rates for middle aged workers. Finally, the optimal policy provides redistribution by decreasing tax rates of wealth-poor young workers. The policy encourages work effort by high elasticity groups while targets inelastic middle aged groups to raise revenues. As a result, total supply of labor increases by 2.98% and total capital by 5.37%. These effects translate into welfare gains of about 0.85% of annual consumption.

*Department of Economics, University of Rochester, e-mail: mkarampa@z.rochester.edu. I feel deeply indebted to Yongsung Chang and Jay Hong for their continuous advice and encouragement. I am also grateful to Mark Bilal for numerous insightful comments. For helpful discussions, I would like to thank Yan Bai, Gregorio Caetano, Bill Dupor, William Hawkins, Cosmin Ilut, Loukas Karabarbounis, Guido Lorenzoni, Ellen McGrattan, Jose Victor Rios Rull, Juan M. Sanchez and seminar participants at the Federal Reserve Bank of Minneapolis and Federal Reserve Bank of St. Louis. All errors are my own.

1 Introduction

Standard public finance principles imply that workers with more elastic labor supply should face smaller tax distortions. Intuitively, the less workers decrease their hours in response to a wage reduction, the smaller the efficiency loss of taxation. Although this argument seems straightforward, the quantitative potential of such an idea is largely unexplored. This paper attempts to fill this void.

Many factors can account for individual differences in labor supply responses. These differences relate to both characteristics unobservable to policymakers, like preferences, and to characteristics that define population groups like age, gender, marital status and wealth. In this paper, I rationalize the heterogeneity in labor supply elasticity based on observables related to the life cycle. For example, a person closer to retirement is more likely to quit her job if her wage falls. The same is true if the person has accumulated a large amount of savings. The government can use these life cycle variables as informative moments to shift the tax burden away from relatively elastic groups. Parts of this idea can already be found in the current US tax system. The social security system is a form of age-dependent taxation since both the contributors and the beneficiaries belong to specific age groups.

To study this issue, I build a dynamic life cycle model of labor supply featuring overlapping generations, heterogeneous agents and incomplete markets. Individuals differ in terms of their wages (productivity), their age and the amount of assets accumulated over the life cycle. Wages have both a fixed effect, a life cycle and a transitory component. The key feature of the economy is that the labor supply decision operates both at the intensive margin, the amount of hours supplied, and at the extensive margin, the decision to participate in the labor market in the first place. A worker participates if the market wage net of taxes is higher than the minimum wage she is willing to accept, the *reservation wage*. The distribution over productivity, asset holdings and age, jointly determine a distribution of reservation wages. Small changes in the market wage will affect only those workers whose reservation wage is sufficiently close to the market wage, the marginal workers. This way, heterogeneity in labor supply elasticity arises endogenously from differences in reservation wages, with marginal workers being the most elastic group in the economy. This is true even if workers have identical preferences over consumption and work. This result goes back to Hansen (1985), Rogerson (1988) and Chang and Kim (2006) who displayed how in an economy with indivisible labor the labor supply elasticity is essentially independent of the preference parameters.

The model features both exogenous and endogenous separations. To discipline transitions between unemployment and employment I use a simple modeling device. New workers have to pay an additional cost upon labor market entry, a search cost.

In the presence of the search cost individuals try to spread employment spells as little as possible along the life cycle. Most workers will continuously work for a number of years and then retire. At the same time, young workers have higher incentive to access the labor market since they expect to work for many years. The model is calibrated to match features of the US economy both at the micro and at the aggregate level. The model is consistent first, with the inverse U-shaped life cycle profile of employment rates and especially the steep decline in participation after the age of 55. Second, the model matches the moderate variation in average hours along the life cycle conditional on participation. Third, it accounts for the very high probability of staying employed for existing workers and the declining probability over the life cycle of switching to employment for unemployed workers.

To quantify the heterogeneity in labor supply elasticity, I simulate the labor supply effects of a one time wage change.¹ The intensive margin labor supply elasticity is 0.64 while the extensive margin elasticity is 0.67. The intensive margin seems to matter more for younger and middle aged cohorts. These age groups have intensive margin labor supply elasticities around the average but approximately zero extensive margin elasticities. On the other hand, people closer to retirement respond more along the extensive margin. Older cohorts are more willing to trade employment for unemployment for two reasons. First, they can use their savings to smooth their consumption. Second, they have fewer working years ahead of them, so that giving up their job seems less costly. Decomposing the elasticities across both age and wealth groups shows that both channels are important.

As is common in optimal taxation problems, the government needs to finance a given amount of expenditures. The set of tax instruments includes a linear capital tax, a linear consumption tax and a progressive labor income tax function. The first two are exogenous in the analysis while the latter is the main subject of this study. At the benchmark economy, the functional form of the labor income tax schedule is a close approximation to the current US labor income tax code. The specification is based on Heathcote, Storesletten and Violante (2010), who show that the after-tax labor earnings are log-linear in pre-tax labor earnings. To find the optimal tax code, I follow the Ramsey approach. Specifically, I make a parametric assumption regarding the relation between labor income taxes and life cycle observables like age and wealth.² The optimal tax code picks the set of parameters that maximize the social welfare. The criterion to evaluate the different tax systems is the expected lifetime utility of the newborn at the new steady state. Since the newborn decides under the veil of ignorance, the

¹By labor supply elasticity I mean the Frisch elasticity of labor supply. This elasticity holds marginal utility of wealth constant and is larger than both the Marshallian (uncompensated) and the Hicksian (constant wealth) labor supply elasticity.

²In the Ramsey approach the tax instruments are assumed to be restricted. In the Mirrlees approach the set of tax instruments is endogenously restricted due to an informational friction, namely that the government cannot observe workers' productivity.

social welfare function places weight to both efficiency and redistribution. The welfare gains are quantified in terms of consumption equivalent variation across steady states.

The optimal tax plan tailors the average and marginal tax rates to the labor elasticity profile. The main properties are as follows. First, the plan decreases significantly the tax rates of people close to retirement. At the same time, *within* older age cohorts, wealth-rich agents face more generous tax cuts than their wealth-poor peers. This policy corresponds to our findings, namely that older *and* wealthier agents are the most sensitive groups in the economy. This leads to a large increase in working hours mainly through a participation effect. Second, the optimal policy targets middle aged groups to raise revenues. Under the new tax plan, a 45 year old worker can face an increase of as high as 7% regarding his average tax rates and 5% regarding his marginal tax rates. At first glance, this feature seems to distort heavily the working choice of relatively productive agents. However, since these groups face small intensive and approximately zero participation elasticities the government can raise revenues at a small efficiency cost. Third, as Weinzierl (2010) documents, age dependent taxation is a powerful tool for redistribution. The optimal tax plan transfers resources towards young and especially wealth-poor workers.

The effect of the reform to aggregate macroeconomic variables is substantial. Total supply of labor, measured in efficiency units, increases by 2.98%. Middle aged workers decrease their labor supply by 0.96% while older workers increase their labor supply by 9.78%. Capital increases by 5.37%. Workers who delay their retirement also delay running down their asset holdings. As a result, the wage increases by 0.84%. At the same time, consumption increases by 4.65%. This change is driven by a large increase in consumption for workers between 51 and 65 (about 5.44%) and especially for retirees (about 13.10%). This generates sizable welfare gains even for age cohorts bearing the largest part of the tax burden. The welfare gain for a newborn in terms of consumption equivalent variation is 0.85%.

To provide additional intuition regarding the results, I repeat the quantitative exercise using different versions of the benchmark model. In particular, I investigate the magnitude of efficiency gains if the tax function can only depend on age. This policy is easier to implement since some categories of assets holdings might be unobserved to the tax authorities. I find age dependent taxation to be less effective than a tax code using both age and assets. Once again, the optimal policy decreases tax rates for older more elastic cohorts and increases aggregate labor supply. However, in this case capital decreases significantly. Based on the permanent income hypothesis young people would save less in anticipation of lower tax rates closer to retirement. As a result, age dependent taxes distort savings incentives and consequently, decreases the equilibrium market wage. The optimal tax code that uses both age and wealth penalizes this behavior by specifying

high tax rates for older workers with low asset holdings. This encourages workers to keep saving during their middle ages. It is of interest to compare this last result with recent findings of the dynamic optimal taxation literature. Kocherlakota (2005) finds optimal a capital tax that decreases in labor income. That is, older people with low labor earnings face higher capital taxes. This discourages people from oversaving while young and underproviding work effort when old, while collecting the tax transfers. In our model, the negative correlation between labor income taxes and asset holdings serves two purposes: one, it encourages effort by very elastic wealthy workers and two, it encourages middle aged workers to maintain a high asset position.

The second specification I consider is a model with constant labor supply elasticity. The model used for this exercise assumes divisible labor and a Frisch utility function. In this case, labor supply elasticity is the same across agents. I simulate a tax reform which reallocates taxes away from older cohorts both at the benchmark (heterogeneous elasticity) model and at the constant elasticity model. By comparing the two economies we can assess whether heterogeneity in elasticity is the leading factor behind the efficiency gains. Indeed, I find that the same tax reform generates smaller efficiency gains in the constant elasticity model than our benchmark heterogeneous elasticity model.

Links to the Literature This paper is related to two different strands of literature. The first part of literature is the macro-labor strand which investigates the labor supply elasticity and its relevance for policy making. Saez (2002) firstly incorporated both the intensive and the extensive margin into optimal taxation theory. He demonstrates that if participation elasticities are relatively high at the bottom of the earnings distribution the optimal policy should subsidize low-income earners. Rogerson and Wallenius (2010), develop a complete markets model that also incorporates both margins of labor supply. Like their paper, I find that macro elasticities are unrelated to micro elasticities and that employment responses to a wage change are concentrated among young and old workers. Erosa, Fuster and Kambourov (2011) show how a model with nonlinear wages and heterogeneous workers can capture a rich set of life cycle labor supply facts. They report an aggregate labor supply elasticity around 1.27 and an increasing elasticity profile over age. Compared to their paper, I introduce first, exogenous and endogenous separations in life cycle labor supply, second, a search cost to discipline employment transitions and third, an initial distribution of asset holdings. I find that these modifications can capture well the participation rates along the life cycle especially for young workers.

The second strand is that of quantitative models of optimal taxation. Conesa and Krueger (2006) quantitatively characterize the optimal income tax schedule in a life cycle model that features heterogeneity in agents' skills and savings. They find that the optimal income tax system can be represented by a proportional tax code with a fixed deduction. Conesa, Krueger and Kitao (2010) expand this analysis to a framework

allowing linear capital taxes. The authors find that apart from a strong life cycle motive, a reason for high capital taxation is to implicitly tax less, very elastic older workers. Unlike their paper, I consider a richer set of tax instruments that allows to identify more clearly the elastic groups of the economy. At the same time, I find that a model with both an intensive and an extensive margin of labor supply, matches better the life cycle profile of average hours. A second close paper is the one by Weinzierl (2010). He shows that age-dependent taxes can first, redistribute income across ages and second, tailor the marginal tax rates to the wage distribution within each age cohort to avoid inefficient distortions. He finds that My paper focuses more on the relation between age and labor supply elasticity. Within the model this relation is endogenously determined through a combination of life cycle savings and search frictions. Fukushima (2010) revisits the problem posed by Conesa et al. (2010) using a dynamic model that allows arbitrary tax instruments. He considers an intensive margin model with uniform elasticities. However a model with no extensive margin misses the very high participation elasticity for people close to retirement. Both his paper and Kitao's (2011) verify the results of Kocherlakota (2005) regarding the negative correlation of optimal taxes between capital and labor income. As mentioned above, in my model this negative correlation encourages middle aged workers to maintain a high asset position as they approach retirement. Another very close paper is the one by Guner, Kaygusuz and Ventura (2011) who exploit heterogeneity in labor supply elasticity across genders. They find that a differential tax rate on married females can increase welfare compared to the current progressive US system but it is suboptimal compared to a case of equal *proportional* tax rates across genders. My paper focuses on the life cycle dimension of labor supply elasticity. I find that exploiting this margin can lead to significant gains.

This paper is organized as follows. Section 2 constructs a simple example to develop intuition regarding the main results of the paper. Section 3 sets up the model. Section 4 describes the quantitative specification of the model. Section 5 examines the implications of the model for reservation wages and labor supply elasticities. Section 6 describes the main quantitative experiment as well as different specifications. Section 7 builds a simple exercise to test the paper's main argument. Finally, Section 8 concludes.

2 Intuition in a Static Framework

This section builds a simple static model of labor supply. I explain how to compute the labor supply elasticity both at the intensive and extensive margin for a specific agent. The former depends mostly on preferences while the latter on the relative density of marginal workers. In this example all heterogeneity is generated by differences in initial

asset holdings a_i .³ Finally I show how a simple policy reform can increase participation in the labor market. Each agent i is endowed with asset holdings a_i and has preferences over consumption, c and hours worked, h :

$$U = \max_{c,h} \left\{ \log c_i + \psi \frac{(1 - h_i)^{1-\theta}}{1 - \theta} \right\} \quad (1)$$

subject to

$$c_i = w(1 - \tau)h_i + (1 + r)a_i \quad (2)$$

where w is the wage rate per effective unit of labor, τ is the proportional tax rate, r is the real interest rate and a_i is i 's initial asset holdings. The parameter ψ defines the preference towards leisure and θ the intertemporal substitution of labor supply.

Intensive Margin Adjustments The intensive margin is defined by how much existing workers change the amount of hours they supply in response to wage variations. Worker i equates the marginal rate of substitution between consumption and leisure to the real wage rate.

$$\psi(1 - h(a_i))^{-\theta} = \frac{w(1 - \tau)}{c(a_i)} \quad (3)$$

The optimal supply of hours $h(a_i)$ depends on initial asset holdings. If worker i has a lot of assets she will buy more leisure and work less (income effect). The (intensive) Frisch elasticity of labor supply for i :

$$\varepsilon_i^{\text{Int}} = \frac{1}{\theta} \frac{(1 - h(a_i))}{h(a_i)} \quad (4)$$

This preference specification makes the intensive margin labor supply elasticity endogenous to working hours. Agents working many hours will respond more inelastically than those working a few number of hours. Hence the amount of heterogeneity in the intensive margin elasticity of labor supply will depend on the distribution of hours across workers. If the initial asset holding distribution is concentrated we would expect people to supply equal amount of hours and respond at the same way to wage changes.

Extensive Margin Adjustments The extensive margin of labor supply is defined by how many people enter or exit the labor market in response to wage variations. To make the extensive margin operational, I assume that workers have to pay a fixed cost FC every working period. This cost will not affect the optimal choice of hours but will affect the decision to be employed in the first place. Worker i with initial asset holdings a_i will participate if the value of employment $V^E(a_i)$ is at least as large as the value of being unemployed $V^U(a_i)$. These two are given by

³The full model in Section 3 assumes heterogeneity both in productivity, asset holdings and age.

$$V^E(a_i) = \log(w(1 - \tau)h(a_i) + (1 + r)a_i) + \psi \frac{(1 - h(a_i))^{1-\theta}}{1 - \theta} - FC \quad (5)$$

$$V^U(a_i) = \log((1 + r)a_i) + \psi \frac{1^{1-\theta}}{1 - \theta} \quad (6)$$

The reservation wage is the *wage* net of taxes that makes the agent indifferent between working and not. It is given by

$$w^R(a_i) = \frac{(1 + r)a_i}{h(a_i)} \left[\exp \left\{ -\psi \frac{(1 - h(a_i))^{1-\theta}}{1 - \theta} + const \right\} - 1 \right] \quad (7)$$

where $const = \psi \frac{1^{1-\theta}}{1-\theta} + FC$. Participation amounts to $w(1 - \tau) > w_i^R$. Ceteris paribus, a rich agent will demand a higher wage to enter the labor market. The participation schedule is a step function and consists of three parts. If $w(1 - \tau) < w_i^R$ the worker is not participating. If $w(1 - \tau) = w_i^R$ the worker is indifferent between working and not and if $w(1 - \tau) > w_i^R$ the worker enters the labor market. Worker's i extensive margin elasticity depends on the distance between her reservation wage and the market net wage. If her reservation wage is much lower or higher than the market net wage, small variations in the market wage will leave the worker unaffected. If her reservation wage is sufficiently close to the market wage she is very elastic to wage variations. Workers whose reservation wage is sufficiently close to the market wage are the *marginal workers*. Taking into account both the intensive and the extensive margin we can construct the labor supply decision

$$l_i^s(w^R(a_i)) = \begin{cases} h(a_i) & \text{if } w(1 - \tau) \geq w_i^R(a_i) \\ 0 & \text{if } w(1 - \tau) < w_i^R(a_i) \end{cases} \quad (8)$$

Aggregate Response of Labor Supply The aggregate labor supply at the market wage w equals total amount of hours supplied by people who are working: $L^s(w) = \int_0^w l^s(w^R) d\phi(w^R)$. Differentiating with respect to the market wage and using the Leibnitz rule, we can decompose the aggregate labor supply elasticity ε^{Tot} to its intensive margin ε^{Int} and extensive margin ε^{Ext} components.

$$\underbrace{\frac{L'^s(w)w}{L^s(w)}}_{\text{Total Elasticity}} = \underbrace{\frac{\int_0^w l'(w^R) d\phi(w^R)w}{L^s(w)}}_{\text{Intensive Margin Elasticity}} + \underbrace{l^s(w)w \frac{\phi(w)}{L^s(w)}}_{\text{Extensive Margin Elasticity}} \quad (9)$$

In a heterogeneous agents framework, the adjustment in total hours equals the adjust-

ment in the intensive and the extensive margin. The first term at the right hand side of equation (9) is the aggregate intensive margin elasticity. The magnitude of the response depends on the curvature of the labor supply function l' . The second term at the right hand side of equation (9) is the aggregate extensive margin elasticity. Its value depends *mostly* on the distribution of the reservation wages around the market wage $\phi(w)$. If the reservation wage distribution is very concentrated, the ratio $\frac{\phi(w)}{L^s(w)}$ increases and hence the labor supply elasticity increases. The Hansen-Rogerson limit of infinite elasticity is reached if the reservation wage distribution is degenerate. On the other hand, a dispersed reservation wage distribution will imply a small aggregate labor supply elasticity.

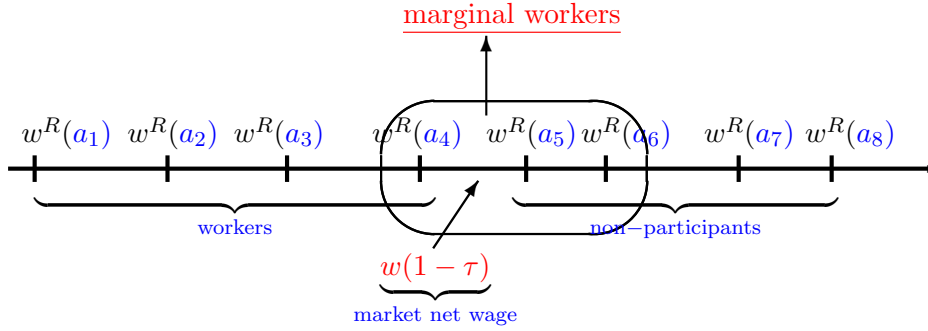


Figure 1: Reservation wages and marginal workers.

Figure 1 displays how the model economy works. In this simple example there are 8 agents. Each is endowed with initial asset holdings a_i where $a_i < a_j$ with $i < j$. The initial asset holdings distribution will imply a distribution of reservation wages $\phi(w^R(a))$. Low number agents participate in the labor market since their reservation wages are lower than the net market wage. High number, wealthy agents will stay out of the labor market since the net market wage is not high enough. The intensive margin decision for working agent i is based on the function $h(a_i)$. In this example, the employment rate is equal to 50%. A wage variation will affect mostly agents 4, 5, and 6 whose reservation wage is sufficiently close to the net market wage. These marginal workers have very high labor extensive margin elasticities. The larger the density of workers around the market wage, the larger the aggregate response of the economy to a wage change. Agents 1, 2 and 3 will respond only at the intensive margin. This group features zero extensive margin elasticity. Finally, agents 7 and 8 have very large assets so they cannot be affected by small variations in the market wage. Hence, differences in reservation wages generate heterogeneity in labor supply elasticity.

Optimal Taxation To improve the efficiency of the tax system the government should tax less, elastic workers. The government cannot identify directly who is more elastic but can use asset holdings as a proxy for labor supply elasticity. An example of such a tax code is the following.

$$\tau(a) = \begin{cases} \tau_H & \text{if } a \leq a_3 \\ \tau_L & \text{if } a > a_3 \end{cases}$$

The new tax code uses assets to differentiate labor income taxes between low and high elasticity groups. Low assets-low elasticity groups, pay higher labor income taxes. Figure 2 describes the outcome. Agents 1, 2 and 3 with low level of asset holdings pay taxes τ_H and receive a lower net wage $w(1 - \tau_H)$. However their reservation wages are low enough to keep them employed. Adjustment will take place only at the intensive margin. Marginal worker 4 continues to work and pays lower taxes. Marginal workers 5 and 6 enter the labor market in response to the tax cuts. Under the new system they receive a higher net wage $w(1 - \tau_L)$. Agents 7 and 8 are indifferent to this policy. The new policy increases employment. However, several issues arise. First, the policy effect on total hours is ambiguous since agents 1, 2 and 3 will decrease their labor supply at the intensive margin. Second, the policy raises equity concerns as wealth-poor people will bear a higher tax burden. Lastly, this static example cannot capture the significance of time horizon in determining both the reservation wages and the labor supply elasticity. These are all issues that I am going to discuss in the full model.

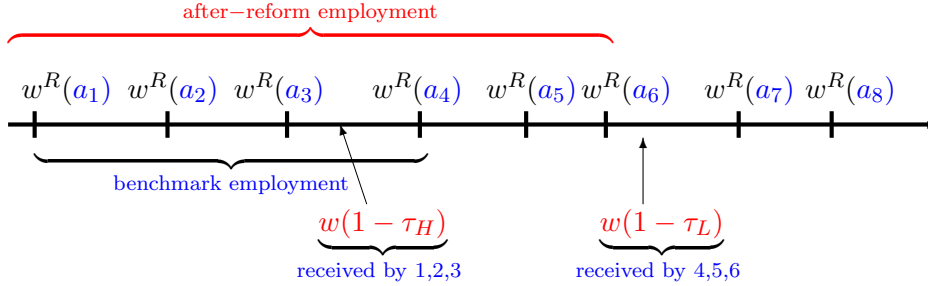


Figure 2: Effects of new tax system on employment.

3 Model

The model is an overlapping generations economy with production and endogenous labor supply decision. The focus is only on steady state equilibria, so I will abstract from any time subscript.

Timing The timing of events can be summarized as follows.

1. At the beginning of the period exogenous separations occur. A fraction λ of previously employed agents, is excluded from the labor market.
2. Idiosyncratic productivity x is realized.

3. All agents make consumption and savings decisions. Previously employed agents who didn't lose their job (the fraction $1 - \lambda$) as well as unemployed from the previous period, make working decisions.

Demographics The economy is populated by J overlapping generations. Generation j is of measure μ_j . In each period a continuum of new agents is born, whose mass is $(1 + n)$ times larger than the previous generation. Conditional on being alive at period $j - 1$ the probability of surviving at year j is s_j . Hence, $\frac{\mu_{j+1}}{\mu_j} = \frac{s_j}{1+n}$. The weights μ_j are normalized so that the economy is of measure one. Agents that reach age j_R have to retire. Retirees receive social security benefits ss financed by proportional labor taxes τ_{ss} . Agents have the option to exit the labor market early but if they do so they will not receive Social Security benefits before the age of j^R .⁴

Preferences Agents derive utility from consumption (c) and leisure. They are endowed with one unit of productive time which they split between work (h) and leisure. Preferences are assumed to be representable by a time separable utility function of the form

$$U = \mathbf{E}_0 \left[\sum_{j=1}^J \beta^{j-1} \prod_{j=1}^J s_j \left\{ \log c_j + \psi_j \frac{(1 - h_j)^{1-\theta}}{1 - \theta} \right\} \right] \quad (10)$$

where β is the discount factor and θ affects the Frisch elasticity of labor supply. I allow the preference parameter ψ_j to depend on age. This assumption helps matching some features of average working hours for people who participate in the labor market. However, the main results of the paper regarding employment rates and the distribution of labor supply elasticity across workers do not depend on this feature (see Section 4.3 and Appendix B for a detailed explanation of this assumption).

Productivity The economy features a nondegenerate distribution of wages. Individuals face permanent differences in productivity and similar life cycle income profiles. At the same time they are subject to persistent idiosyncratic shocks. The natural logarithm of wages for agent i of age j is given by

$$\log \hat{w}_{ij} = \log w + \log z_i + \log \epsilon_j + \log x_j \quad (11)$$

The first component of individual wages is the stationary market wage w which is going to clear the market. Permanent ability is denoted z and is distributed as: $\log(z) \sim N(0, \sigma_z^2)$. The age-specific productivity profile $\{\epsilon_j\}_{j=1}^J$ captures differences in average wages between workers of different ages. This profile evolves deterministically along the life cycle and

⁴If such a case was allowed wealth-poor workers would have a higher incentive to retire early and claim the benefit in case of a bad labor income shock. In addition, this option would deter many workers to save much in the first place. Though interesting I abstract from these modifications for simplicity.

peaks around the age of 50. Finally workers experience idiosyncratic wage shocks. These follow an AR(1) process in logs:

$$\log x_j = \rho \log x_{j-1} + \eta_j, \quad \text{with } \eta_j \sim \text{iid } N(0, \sigma_\eta^2) \quad (12)$$

I assume that newborns enter the life cycle having the lowest level of productivity. As usual the autoregressive process is approximated using Tauchen's method (1986). Appendix C describes the method in detail. The transition matrix which describes the autoregressive process is given by $\Gamma_{xx'}$.

Asset Market and Borrowing Constraints The asset market has two distinct features. The first is that markets are incomplete. Within the set of heterogeneous agents life cycle models such an assumption is standard. From an empirical standpoint incomplete markets support the evidence that consumption responds to income changes. At the same time, in the absence of state-contingent assets, agents use labor effort to insure against negative labor income shocks. This mechanism lowers the correlation between hours and wages, a pattern well documented in the data (Pijoan-Mas, 2006). With this in mind, I restrict the set of financial instruments to a risk-free asset. In particular, agents buy physical claims to capital in the form of an asset a , which costs 1 consumption unit at time t , and pays $(1 + r)$ consumption units at time $t + 1$. r is the real interest rate and will be determined endogenously in the model by the intersection of aggregate savings to aggregate demand for investment. The second feature is a zero borrowing limit.⁵ This assumption can affect greatly labor supply responses.⁶ In the model savings takes place for three reasons. Agents wish to smooth consumption across time (intertemporal savings motive), to insure against labor market risk (precautionary savings motive) and to insure against retirement (life-cycle savings motive).

Initial Assets A robust feature of the data is the increasing employment rate early at the life cycle. Young people enter gradually the labor market until the age of thirty. To generate this pattern I assume that newborns are endowed with an initial level of assets. This asset is a random draw from a lognormal distribution with mean $(\bar{a}_{\{j=1\}})$ and standard deviation $(\sigma_{a_{\{j=1\}}})$. Total initial assets for the newborns are denoted as $\int a_{\{j=1\}}$.

Production There is a representative firm operating a Cobb-Douglas production function. The firm rents labor efficiency units and capital from households at rate w (the wage rate per effective unit of labor) and r (the rental rate of capital), respectively. Capital depreciates at rate $\delta \in (0, 1)$. The aggregate resource constraint is given by

⁵The reason the limit is zero instead of a small negative value is the presence of stochastic mortality. If borrowing was allowed some net borrowers would die (unexpectedly) without having paid their debt.

⁶According to Domeij and Floden (2006) borrowing constrained individuals can smooth their consumption only by increasing their labor supply. Hence, on the presence of borrowing constraints the labor supply elasticity is downward biased.

$$C + (n + \delta)K + G = f(K, L) + \int a_{\{j=1\}} \quad (13)$$

where C is aggregate consumption, K is aggregate capital and L is aggregate labor, measured in efficiency units. G represents government expenditures. Equation (14) equalizes total demand and total supply. The latter equals output produced by the technology production $f(K, L)$ and the initial endowment $\int a_{\{j=1\}}$.

Government The government operates a balanced pay-as-you-go social security system. Each beneficiary receives social security benefits ss that are independent of his contributions and are financed by proportional labor taxes τ_{ss} . This payroll tax is taken as exogenous in the analysis. In addition, the government needs to collect revenues in order to finance the given level of government expenditures G . To do so it taxes consumption, capital and labor. Consumption and capital income taxes τ_c, τ_k are proportional and exogenous. At the same time the government taxes labor earnings using a nonlinear tax schedule:

$$T^L(\hat{w}h) = \hat{w}h - (1 - \tau_0)(\hat{w}h)^{1-\tau_1} \quad (14)$$

where $\hat{w} = wz\epsilon_jx$. If $\tau_1 = 0$ the tax function becomes a proportional tax schedule. For $\tau_1 > 0$ the system becomes progressive since high earners pay a higher fraction of their earnings in taxes. The parameter τ_0 affects the average and the marginal taxes rates in the same way. Higher values of τ_0 imply that working agents face both higher average and marginal tax rates. This specification is used by Heathcote et al. (2010). Finally, the government distributes uniformly the accidental bequests to all living agents. These transfers are denoted Tr .

Fixed Cost and Search Cost To make the participation margin operational I assume that workers have to pay a fixed cost every time they work. The fixed cost is measured in utility terms. The fixed cost can take two values corresponding to young and old working cohorts $FC_j = \{FC_y, FC_o\}$. In addition, I assume that *new workers* have to pay an extra utility cost, rationalized as a search cost sc . This way people who were unemployed at age $j - 1$ must pay a larger total cost at age j in order to work. The search cost also takes two values corresponding to young and old working cohorts $sc_j = \{sc_y, sc_o\}$. I denote the total fixed cost

$$\zeta_j(S_{-1}) = \begin{cases} FC_j + sc_j & \text{if } S_{-1} = u \\ FC_j & \text{if } S_{-1} = e \end{cases} \quad (15)$$

I index ζ_j because both the fixed cost and the search cost are a function of age.

Worker's problem There are five dimensions of heterogeneity: asset holdings a ,

stochastic productivity x , fixed effect z , lagged employment status S_{-1} , and age j . A working agent of age j has pre-tax labor income $\hat{w}h = wzx\epsilon_j h$ and pre-tax capital income $r(a + Tr)$. The worker will decide to participate in the labor market if the value of being employed evaluated at the optimal hours level is higher than the value of being unemployed. Workers decision is constrained by the limited borrowing constraint $a' \geq 0$ and the nonnegative consumption constraint $c \geq 0$. In the following problems I take these constraints as given. The value function for employment is given by :

$$V_j^E(a, x, z, S_{-1}) = \max_{c, a'} \left\{ \log(c) + \psi_j \frac{(1 - \mathbf{h})^{1-\theta}}{1 - \theta} - \zeta(S_{-1}) + \beta s_{j+1} \sum_{x'} \Gamma_{xx'} [(1 - \lambda)V_{j+1}(a', x', z, S) + \lambda V_{j+1}^U(a', x', z)] \right\} \quad (16)$$

s.t.

$$(1 + \tau_c)c + a' = (1 - \tau_{ss})\hat{w}\mathbf{h} - T^L(\hat{w}\mathbf{h}) + (1 + r(1 - \tau_k))(a + Tr) \quad (17)$$

$$\mathbf{h} \text{ solves the first order condition } \psi_j(1 - h)^{-\theta}c(1 + \tau_c) = \hat{w}(1 - T^{L'}(\hat{w}h)) \quad (18)$$

$$x' \sim \Gamma_{xx'} \text{ and } S = e \quad (19)$$

The value function is the sum of current and future utility evaluated at the maximum choices. The continuation value includes the small probability of exogenous unemployment. Equation (17) is the worker's budget constraint. As usual consumption and savings equal after-tax labor and capital income. Transfers from accidental bequests are part of the budget constraint. Equation (18) is the static first order condition between consumption and hours. Equation (19) describes the evolution of the state variables. Productivity x evolves according to the autoregressive process. In addition, next period's employment status will be e . The value function for the unemployed is given by the following equation.

$$V_j^U(a, x, z) = \max_{c, a'} \left\{ \log(c) + \frac{\psi_j}{1 - \theta} + \beta s_{j+1} \sum_{x'} \Gamma_{xx'} V_{j+1}(a', x', z, S) \right\} \quad (20)$$

s.t.

$$(1 + \tau_c)c + a' = (1 + r(1 - \tau_k))(a + Tr) \quad (21)$$

$$x' \sim \Gamma_{xx'} \quad \text{and} \quad S = u \quad (22)$$

The value function for the unemployed does not depend on previous employment status so that S_{-1} is not a state variable. However, if the worker decides to work next year she will have to pay the additional search cost. The continuation value includes this period's employment status, $S = u$. The participation decision is based on the relative values of employment and unemployment.

$$\textbf{Participation Decision:} \quad V_{j+1} = \max_{h \in \{0, \textcolor{red}{h}\}} \{V_{j+1}^E, V_{j+1}^U\} \quad (23)$$

The problem for the retirees is similar to the unemployed with the exception of the social security benefit received every period. It is not displayed for convenience.

Distribution of states Agents are heterogeneous in their state vectors $\omega \in \Omega = A \times X \times Z \times \Sigma$, where $A = [0, \bar{a}]$ is the asset space. The lower bound of zero is based on our no-borrowing assumption. Since the agents cannot save more than what they earn over their lifetime we can safely assume an upper bound $\bar{a}$. The productivity state space is given by $X = Z = R$ and $\Sigma = \{e, u\}$ is the set of possible values for the previous employment status. The policy function for savings, consumption and hours is given by $g_j^a(\omega)$, $g_j^c(\omega)$ and $g_j^h(\omega)$ respectively. Let $\Phi_j(a, x, z, S_{-1})$ denote the cumulative probability distribution of the individual states $(a, x, z, S_{-1}) \in \Omega$ across agents of age j . The marginal density is denoted by $\phi_j(a, x, z, S_{-1})$.

Equilibrium The model is solved in general equilibrium. The equilibrium is described in a recursive way. I focus on a stationary equilibrium where prices and aggregate variables are constant. Specifically, given a tax structure $\{\tau_c, T^L(\cdot)\tau_k, \tau_{ss}\}$ and an initial distribution $\Phi_1(a, 1, z, u)$, a *stationary competitive equilibrium* consists of **functions** $\{V_j^E, V_j^U, g_j^a, g_j^c, g_j^h\}_{j=1}^J$, **prices** $\{w, r\}$, **inputs** $\{K, L\}$, **benefits** $\{ss\}$, **transfers** $\{Tr\}$ and **distributions** $\{\Phi_j(a, x, z, S_{-1})\}_{j=2}^J$ s.t.

- given prices $\{w, r\}$, benefits $\{ss\}$ and transfers $\{Tr\}$ the functions solve the household's problem;
- the prices satisfy the firm's optimal decisions, $r = F_K(K, L) - \delta$ and $w = F_L(K, L)$;

- capital and labor markets clear:

$$K = \sum_{j=1}^{J-1} \mu_{j+1} \int_{\Omega} g_j^a \phi_j \quad \text{and} \quad L = \sum_{j=1}^J \mu_j \int_S z x \epsilon_j g_j^h \phi_j$$

.

- the social security system clears: $\tau_{ss} w L = ss \sum_{j=j^R}^J \mu_j$;
- the transfers are given by: $Tr = \int_{\Omega} \mu_j (1 - s_j) g_j^a$;
- the government balances its budget: $G = \tau_c C + \tau_k r K + \int_{\Omega} T^L(.) d\phi$;
- the distribution of states for people with fixed effect z who are currently working evolves based on the following rule:

$$\phi_{j+1}(a', x', z, e) = \sum_{S_{-1}=\{e,u\}} \sum_x \Gamma_{xx'} \phi_j(g_a^{-1}(a', .), x, z, S_{-1})$$

I explain the last condition in more detail. $\phi_{j+1}(a', x', z, e)$ is the density of people with assets a' , productivity x' and fixed effect z who were working at age j . This measure will consist of people who saved $a' = g^a(a, x, z, S_{-1})$. The inverse function $g_a^{-1}(a', x, z, S_{-1})$ gives the amount of assets a needed to save a' given productivity x . From people with states a, x that lead to savings a' only $\Gamma_{xx'}$ will move to (a', x') . The sum is taken all over possible values of x . The outer sum denotes that this rule holds for age j workers either employed at $j - 1$ or unemployed at $j - 1$. We can construct similar rules for the currently unemployed.

4 Quantitative Analysis

4.1 Data-Facts

I use data from the PSID waves from 1970 to 2005 and restrict the sample to male head of households who are the primary earners (see Appendix A for a detailed description of the data). An agent is regarded as employed if she works more than 800 hours annually (15 hours per week). I briefly describe key patterns regarding males' labor supply. These patterns are consistent with other studies focusing on the labor supply decision of males (Prescott, Rogerson and Walenius, 2009 and Erosa et al., 2011).

1. Annual working hours are roughly hump shaped over the life cycle. On average annual hours increase from around 1850 hours at age 21 to 2250 hours at age 35. At middle ages the hours profile stays roughly constant around 2200 hours. After the age of 50 the profile declines at an increasing rate. Average hours fall from 1950 at the age of 55 to 1650 at the age 60 and to 900 at the age of 65.
2. Conditional on participation, males vary very little their lifetime labor supply. Middle aged cohorts work around 2350 hours per year while cohorts close to retirement work around 2100 hours. Hence life cycle variations in average hours are mainly driven from the participation margin.
3. The probability of being employed (working more than 800 hours annually) at time $t + 1$ is very high - around 95% for employed males at time t . The probability decreases only after the age of 60. The probability of switching to employment at time $t + 1$ for unemployed males at time t is decreasing along the life cycle. This implies that unemployment becomes an absorbing state.

4.2 Calibration

This section describes the calibration of the model. I first calibrate exogenously a subset of parameters. Then I choose the remaining parameters so that the associated stationary equilibrium is consistent with U.S. data along several dimensions. Essentially this calibration strategy can be seen as an exactly identified method of moments estimation. The parameter estimates are summarized in the Appendix D.

Externally Calibrated Parameters The model period is set to one year. The agents are born at real life age of 21 (model period 1) and live up to a maximum real life age of 101 (model period 81). Agents become exogenously unproductive and hence retire at real life age of 65 (model period 46). The survival probabilities are taken from the life table (Table 4.C6) in Social Security Administration (2005). I use the corresponding probabilities for males.

The population growth rate is set to $n = 1.1\%$, the long-run average population growth in the US. The deterministic age-dependent productivity profile is taken from Hansen (1993). The production function is Cobb-Douglas, $f(K, L) = K^\alpha L^{1-\alpha}$, where $\alpha = 0.36$ is chosen to match the capital share. As already noted, preferences are separable in consumption and leisure. Parameter θ elasticity which determines the Frisch labor supply elasticity is set to 2. This is based on Erosa et al. (2011). The time endowment equals 5200 hours per year (Prescott et al., 2009). I set the standard deviation of the initial

asset distribution equal to $\sigma_{a_{\{j=1\}}} = 1.96$, based on Alan (2006).

For the tax rates I use values based on Imrohoroglu and Kitao (2009). The consumption tax is set at $\tau_c = 5\%$ and the capital tax rate to $\tau_k = 30\%$. The social security tax is set at $\tau_{ss} = 10.6\%$ based on Kitao (2010). This gives a replacement ratio around 45%. To pin down the parameter τ_1 I use the estimates by Heathcote et al. (2010). The authors show that the after tax earnings is log-linear in pre-tax earnings. The rate of progressivity τ_1 defines the slope. The authors use data from CPS for the time period of 1980-2005 and estimate $\tau_1 = 0.26$.

Endogenous Calibration There are a total of 15 parameters to be estimated.

In a general equilibrium framework all parameters affect all moments. However, it is possible to associate a specific parameter with a given moment.

- Discount factor (β): The discount factor affects directly the level of aggregate savings. Discounting the future at higher rates leads to more savings and a higher capital-output ratio. The discount factor targets a capital-output ratio equal to 3.2.

- Depreciation rate (δ): Using the steady state relationship $I = (n + \delta)K$, we can easily pin down the depreciation rate as $\delta = \frac{\frac{I}{Y}}{\frac{K}{Y}} - n$. Targeting an investment-output ratio of 0.25 leads to a value of $\delta = 0.0816$.

- Utility parameter (ψ_j): This parameter captures the relative preference towards work. Higher values of ψ decrease the amount of work supplied by workers. To pin down ψ_j I target the slightly hump-shaped profile of hours conditional on participation along the life cycle. I assume that $\psi_j = \alpha_0 + \alpha_1 j$. To find α_0 I use the average working hours conditional on participation for ages 21-40 and for α_1 the average between 41-60. The first group works on average 43.92% while the second 40.22% of their time endowment. For the last 5 years I specify a new profile $\psi_j = \psi_{60} + \alpha_2 j$. I calibrate α_2 to match the average hours during those last five years equal to 37.6%. The choices about these specifications are explained in detail in the next section.

- Initial assets ($\bar{a}_{j=1}$): To determine the mean of the initial asset holdings distribution I use that young people below 30 have around 10% of average asset holdings of all agents below 65.

- Fixed costs FC_y, FC_o : The fixed cost discourages agents from participating in the labor market. To find the two values, I use the average employment rate between ages 21-42 equal to 0.93 and between 43-65 equal to 0.82.

- Separation rate (λ): Higher separation rate increases these transitions from employment to unemployment. The average life cycle transitions between these states, equal to 6.09%, serves as a target.
- Search costs (sc_y, sc_o): Both parameters discipline the transitions between unemployment and employment. Larger search cost limits the transitions from unemployment to employment. To pin down sc_y, sc_o I will use the average transition probability between ages 21-42 equal to 0.48, and the average between ages 43-65 equal to 0.17. The search cost helps creating a decreasing life cycle profile.
- Tax parameter (τ_0): The labor income tax is pinned down so that in equilibrium the government spending to output ratio equals 0.20.
- Productivity parameters ($\sigma_z, \rho, \sigma_\eta$): To pin down the last three parameters I follow the identification strategy of Storesletten, Telmer and Yaron (2004). My main target is the life-cycle profile of the variance of log labor earnings. Storesletten et al. (2004) report that this variance is close to 0.3 at age 22 and increases linearly to 0.9 by the age of 60. In this model all agents start off their lives having the same transitory shock x . As a result, any dispersion in labor earnings is caused by the dispersion in the fixed effect z , i.e. by the parameter σ_z . As the cohort ages the distribution of transitory shocks converges towards its invariant distribution. The variance of log labor earnings at the stationary distribution is pinned down by the variance of the transitory shock, σ_η . Lastly, the persistence of the transitory shock determines how fast we get to the invariant distribution. The slower the rate the flatter the slope of the life cycle variance. This helps pin down ρ .

4.3 Model's Performance

Our exactly identified estimation strategy left a rich set of statistics untargeted. A good way to test the model is to examine how the model performs with respect to these out-of-sample predictions. This is equivalent to an informal over-identification test. Good performance builds confidence to use the model for policy recommendations.

Life Cycle Profiles of Employment and Hours The average participation rate between 21-43 equal to 0.94 and between 43-65 equal to 0.82 were explicitly targeted. The right panel of Figure 3 examines how well the model fits the whole life cycle profile. In the model, employment features the three phases observed in the data. Firstly, an increasing profile up to the age of 30. Agents start their life at the lowest productivity level. Gradually some people start getting better wage offers (higher productivity shocks)

and enter the market. This mechanism resembles a standard job search model where agents receive randomly offers and accept if the wage is higher than their reservation wage. Agents also experience higher wages on average due to an increasing life cycle component of earnings. At the same time positive initial asset holdings allow the workers to stay out of employment during the first unproductive years. Gradually, as productivity increases and as their assets run out, they enter the labor market. The second feature of the data, captured by the model is a flat, very persistent profile at middle ages. There are two reasons why agents at this age are very strongly attached to their labor market status. The first is very high productivity. The second is the search cost, which deters people from going in and out of employment, at regular time intervals. Finally the model replicates the steep decline in employment rates after the age of 50 generated by a large stock of accumulated savings and a declining average life cycle productivity. Note that the model can match very well the participation profiles even in the absence of age dependent preference parameters ψ_j (see the following discussion as well as Figure 9 in Appendix B).

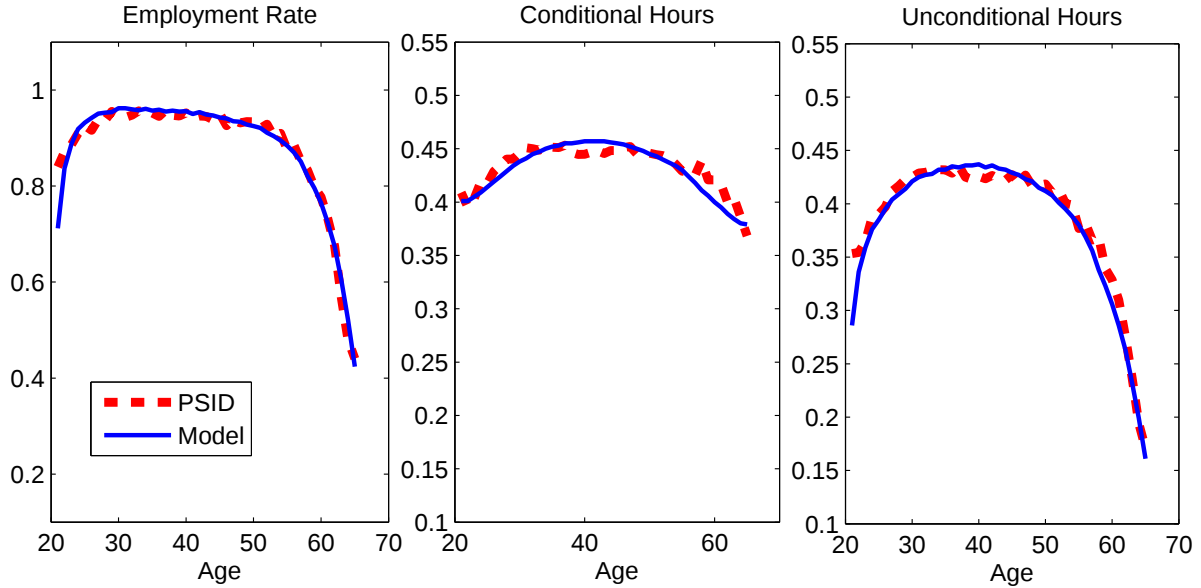


Figure 3: *Left Panel.* Participation Rates over the Life Cycle. *Middle Panel.* Average Hours conditional on Participation. *Right Panel.* Average Hours for all agents.

The middle panel of Figure 3 plots average working hours conditional on participation. Many factors affect this profile. To build intuition we write the Euler equation for hours.

$$\left(\frac{1 - h_{j+1}}{1 - h_j}\right)^\theta = \frac{\psi_j}{\psi_{j+1}} \frac{\epsilon_j}{\epsilon_{j+1}} \beta s_{j+1} (1 + r(1 - \tau_k)) \quad (24)$$

The profile depends on the life cycle productivity $\frac{\epsilon_j}{\epsilon_{j+1}}$. Life cycle wages are in general

increasing which induces the agent to work more. In addition, the profile depends on the calibrated value of $\beta s_{j+1}(1 + r(1 - \tau_k))$. This value is approximately 1.02, which decreases the average hours over the life cycle. To match better the profile I use the preference parameters $\frac{\psi_j}{\psi_{j+1}}$. The middle upper panel of Figure 9 in Appendix B shows how average hours look like if ψ is uniform across ages. Two features stand out. Firstly, although average hours are hump shaped the profile peaks very early. Calibration finds a negative value for $\alpha_1 = -0.0066$ in the specification $\psi_j = \alpha_0 + \alpha_1 j$ to induce workers to increase average hours up to the age of 45. The second feature is the increase in average hours after the age of 60. This is a selection effect. By this age low skill workers have retired leaving only high productivity - high hours workers as part of the workforce. To deal with this problem I specify a new profile $\psi_j = \psi_{60} + \alpha_2 j$ with $\alpha_2 = 0.005$. This way even productive agents who wish to be employed decrease the amount of hours they provide. Lastly, the profile depends on θ . Higher values of θ imply smaller intensive margin labor supply elasticity and smaller response of hours to wage and interest rate changes. Hence, low values of θ imply a flatter hours profile.

Life Cycle Transitions Figure 4 plots average transitions over the life cycle: employment to employment (left panel) and unemployment to employment (right panel). The separation rate targeted the average transitions between employment and employment. The model is able to match the very flat probability of staying employed within a year, and the decreasing part after the age of 60. The model also matches a decreasing life cycle probability of switching from unemployment to employment. The search cost helps to discipline this profile. The right lower panel of Figure 9 in Appendix B shows how transitions between unemployment and employment look like without the search cost. The profile increases a lot until middle ages by workers who lost exogenously their job and wish to re-enter to the labor market. The search cost limits these transitions. At the same time, given the search cost the strategy of switching too often between employment and unemployment becomes suboptimal. Most workers will continuously work for a number of years and then retire.

Wealth Inequality It is important to test if the model can generate a realistic amount of wealth heterogeneity not only at the aggregate level, but also within age-cohort. Table 1 reports wealth Gini coefficients across age groups as found in the PSID and in the model.

Table 1: Wealth Gini by Age

Age Group	21-30	31-40	41-50	51-60	61-65
Gini (PSID)	0.9455	0.8728	0.8045	0.7814	0.7860
Gini (Model)	0.8661	0.6296	0.5003	0.4312	0.3958

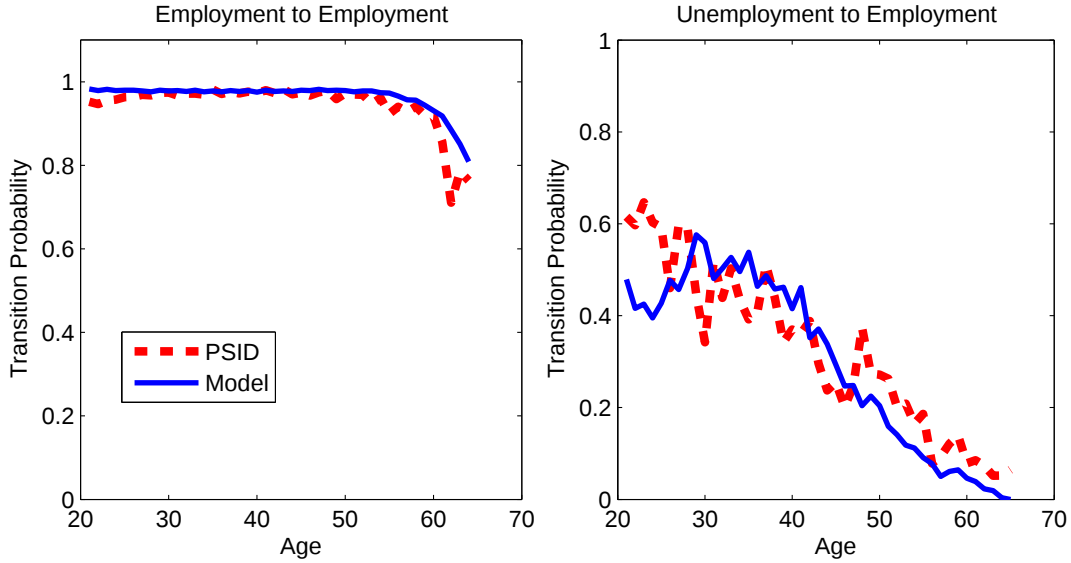


Figure 4: Transition probabilities over the life cycle. *Left Panel.* Employment to Employment. *Right Panel.* Unemployment to Employment.

I find that in the PSID the coefficients are highest for households in their twenties and weakly decreasing between the ages of 30 and 65. The model is able to generate a concentrated wealth distribution, albeit not as much as in the data. The average wealth Gini for ages 21-25 is 0.9455, while 0.8661 in the model. The average Gini for agents at their 40's is 0.8045 in PSID while 0.5003 in the model. This failure to generate high concentration is a standard property of incomplete markets models with idiosyncratic risk. In spite of this, the model can still produce sufficient heterogeneity both across and within age cohorts.

4.4 Life Cycle Profiles

Figure 5 summarizes the average life cycle profiles in the benchmark economy. The left panel displays average asset holdings which feature the usual hump shape. Agents build up their (life-cycle) savings to prepare for retirement. At the same time, agents save precautionary to insure against negative income shocks. These are the two main savings motives in the model. The assets profile increases slowly at first. This happens because young workers expect higher life cycle wages and delay savings for some periods. As agents approach retirement the life cycle motive becomes stronger and the profile increases steeply. After retirement, the retirees use their stock of savings to boost their consumption. The very high probability of dying after the age of 90 explains why people hold basically zero wealth.

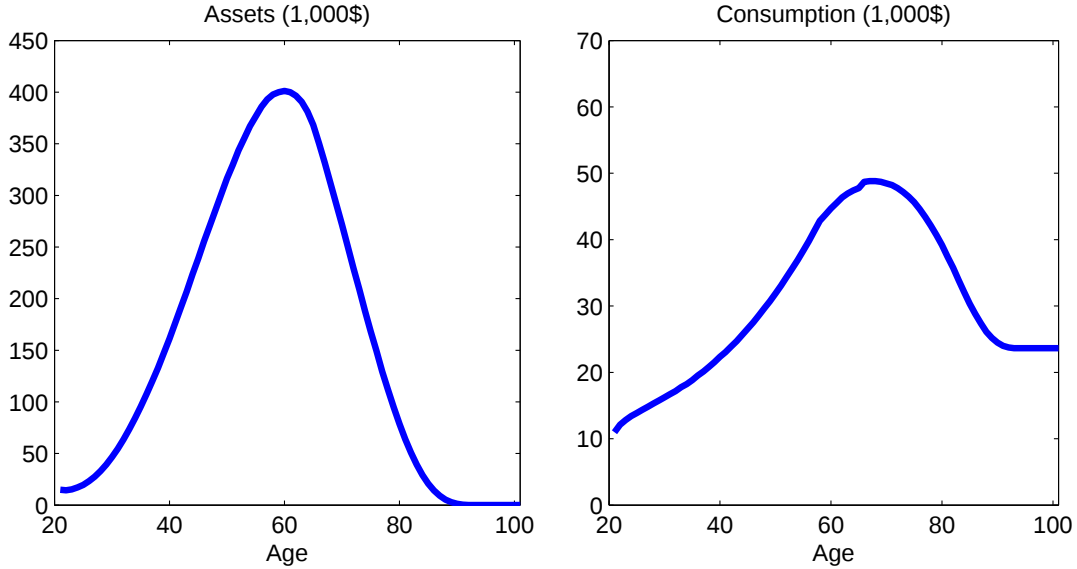


Figure 5: Life cycle profiles *Left Panel.* Average assets. *Right Panel.* Average consumption.

The right panel features average life-cycle consumption. The profile is increasing up to retirement. This has to do with three reasons. First, the combination of borrowing constraints and an increasing life-cycle productivity. If capital markets were perfect (borrowing was allowed) agents would borrow against higher future wages and the consumption path would be flatter. Borrowing constraints make consumption track productivity at least early in the life cycle. Second, as Gourinchas and Parker (2002) emphasize, precautionary savings lead to early asset accumulation. Third, the parameter $\beta s_{j+1}(1 + r(1 - \tau_k)) = 1.02$ makes consumption growth optimal. After retirement the profile decreases as high mortality risk decreases peoples' willingness to save.

5 Labor Supply Elasticity

This section quantifies the heterogeneity in labor supply elasticity. I find significant heterogeneity across skill, wealth and age groups. The extensive margin accounts mostly for these differences. I firstly analyze the determinants of reservation wages and then compute the labor supply elasticity across age and wealth groups.

Reservation wages The reservation wage is the minimum wage an agent would be willing to work for. Naturally, people of different skill, wealth and age would demand different minimum wages. In Figure 6, I plot reservation wage schedules as a function of assets, for people previously employed. The left panel displays schedules for a low skilled worker at two different ages 45 and 55. The right panel displays the same

schedules for a high skilled worker. The schedule represents the wage that makes the agent indifferent between working and not working. People whose reservation wage is below the market wage will participate in the labor market. In both figures, reservation wages are increasing in *assets*. Intuitively, wealth-rich people have higher outside options than wealth-poor and thus demand higher wages. Alexopoulos and Gladden (2006) find that wealth increases reservation wages and decreases the probability of moving into employment. Secondly, reservation wages decrease in *productivity*. The same reservation schedule is lower for the high skill types (right panel) compared to the low skill types (left panel). Productive workers have very high effective wages. Thus, they don't mind a lower market wage. Lastly, conditional on asset holdings, reservation wages increase in *age*. In the model economy, workers have to pay a one-time search cost to find a job. Paying the search cost is equivalent to buying an asset which allows access to the labor market. Young people who expect to work for many years are more reluctant to switch to unemployment (sell this asset). Hence, younger cohorts have on average lower reservation wages.

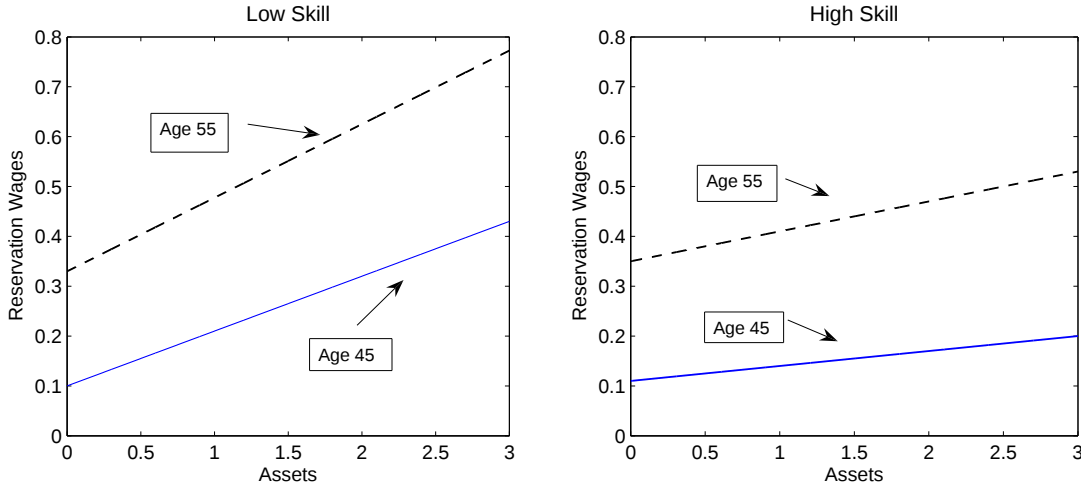


Figure 6: Reservation Wages and Assets. *Left Panel.* Low skill worker. *Right Panel.* High skill worker.

Labor Supply Elasticity The intensive margin labor supply elasticity is computed based on equation (4). For each group I use average working hours. Table 2 reports our findings. The intensive margin labor supply elasticity is high around 0.64. This value depends crucially on parameter θ which is calibrated at the value of 2. On average older cohorts are more elastic since they work fewer hours. Intensive elasticities across age groups range from 0.62 to 0.71. At the same time, the elasticity decreases on wealth. The variation is insignificant though, with all groups ranging between the values of 0.63 and 0.65. In general the dispersion of intensive elasticities is small.

Table 2: Labor Supply Elasticity

Intensive ϵ^{Int}	Q1	Q2	Q3	Q4	Aggregate
Age 21-30	0.66	0.66	0.61	0.56	0.62
Age 31-40	0.65	0.61	0.59	0.56	0.60
Age 41-50	0.64	0.63	0.63	0.61	0.63
Age 51-60	0.71	0.70	0.69	0.68	0.70
Age 61-65	0.71	0.71	0.70	0.69	0.71
Aggregate	0.65	0.62	0.63	0.63	0.64
Extensive ϵ^{Ext}	Q1	Q2	Q3	Q4	Aggregate
Age 21-30	0.00	0.00	0.00	0.53	0.12
Age 31-40	0.00	0.00	0.00	0.07	0.02
Age 41-50	0.00	0.00	0.05	0.36	0.11
Age 51-60	0.19	0.31	1.12	3.39	1.30
Age 61-65	0.07	0.51	3.37	6.28	3.20
Aggregate	0.10	0.32	0.75	1.52	0.67

Conditional on participation people work more or less the same amount of hours. More interesting are the findings regarding the extensive margin labor supply elasticity. This elasticity is computed by using the relative density of marginal workers around the market wage. The density represents the measure of people going into or out of employment in case of a small variation in the market wage. Since the wealth distribution is assumed constant this elasticity approximates the Frisch elasticity of labor supply. To compute group level elasticities I use the relative density of marginal workers at each group. The aggregate extensive margin elasticity is 0.67, which is very close to the intensive margin elasticity. Hence, the extensive margin accounts for almost 50% of the aggregate value. What sets the extensive margin apart is the significant variation in labor supply elasticity across groups. Young cohorts feature essentially zero participation elasticities while older cohorts around 2.5 on average. Even after controlling for wealth, older cohorts are more elastic. This is consistent with the reservation schedules plotted in Figure 6. The most elastic group in the economy are wealthy agents close to retirement. These agents can easily switch to retirement because first, they can use their assets to boost their consumption and second, they place smaller value on their employment status as they have a shorter working time horizon. In conclusion, the intensive margin seems to matter more for younger and middle aged cohorts than the extensive margin. The opposite is true for people close to retirement. Adding both the intensive and extensive margin we find a U-shape with respect to age.

Links to the Literature There is an extensive literature and a wide range of methodologies regarding the measurement of labor supply elasticity. Most of the

evidence from the micro-labor side, point to relatively small labor supply elasticities, especially for males. For example, MaCurdy (1981) and Altonji (1986) find values equal to 0.15 and 0.172 respectively for the Frisch labor supply elasticity. Pistaferri (2003) reports a higher value of 0.70. An alternative approach, supported by researchers of the macro-labor side, uses the extensive margin to break the tight link between the individual preference parameters and the aggregate response of the economy. Chang and Kim (2006) show that an indivisible labor economy calibrated to match heterogeneity in wages and participation rates gives a labor supply elasticity around 0.9 for males (and even larger for females). In their framework the value of the individual preference parameter becomes irrelevant. Rogerson and Wallenius (2010) simulate a panel data set of wages and hours using a life cycle economy with fixed cost of working. They find a discrepancy between the micro value and the true value of labor supply elasticity. While the former ranges between 0.05 and 1.25 the latter ranges between 2.3 and 3. Erosa et al. (2011) extend this analysis by modeling incomplete markets and nonlinear wages. The aggregate labor supply elasticity in their paper is 1.27 with the extensive margin accounting for almost 50% of the aggregate value. Imai and Keane (2004) argue that a model with human capital can correctly estimate the intertemporal substitution of labor. They find a value of 3.82 in their model. My value of 1.31 is in general consistent with the macro approach.

Much less work has been conducted on the issue of group level, life cycle elasticities. Erosa et al. (2011) find elasticities of 1.0 for agents around 25-35 and 1.98 for individuals aged 55-64. Gourio and Noual (2010) focus on younger cohorts and report a decreasing pattern of labor supply elasticity with younger people being more elastic than middle aged. Jaimovich and Siu (2009) report that young and old cohorts experience much greater cyclical volatility in hours than the prime-aged. Lastly, French (2005) simulates a life cycle model and finds that at the age of 40 the labor supply elasticity is around 0.25 while at age 60 is around 1.15. My findings are consistent with the U shape reported by most studies. What is missing in the literature in my view, is a better understanding of the factors responsible for this life cycle pattern. I find that both a larger stock of savings and a shorter time horizon account for the life cycle profile of elasticity. This result has direct policy implications. By using information on both assets and age, the policymakers can identify the marginal workers with a higher level of accuracy.

6 Optimal Taxation

This section sets up the main quantitative experiment. The problem is to choose the best possible tax code from a given set of tax instruments, with the objective to collect a necessary amount of revenues. I state the problem in terms of an optimal Ramsey

problem and discuss the results.

Social welfare function The social planner's objective is to maximize the ex ante expected lifetime utility of the newborn at the new steady state. Employing this social welfare function is common among optimal policy models. Since the newborn is unaware of her initial assets and her fixed effect, the function corresponds to a Rawlsian veil of ignorance. This welfare function captures two main concerns. The first is the efficiency of the economy. The newborn would like to be born in a high wage - high consumption economy. The second is the insurance provided in the economy. The newborn would like to be protected in case she faces a series of bad labor income shocks. Formally, the function is written as

$$W = \int V_1(a, 1, z, u) \Phi_1(a, 1, z, u). \quad (25)$$

The integral is taken over possible types z and possible asset holdings a . The newborn always starts her life cycle having the lowest transitory shock $x = 1$ and having unemployment status.

Tax instruments The benchmark labor income tax schedule

$$\pi^1 = T^L(\hat{w}h) = \hat{w}h - (1 - \tau_0)(\hat{w}h)^{1-\tau_1} \quad (26)$$

depends on earnings. The main experiment is to introduce a new set of tax instruments π^2 which incorporates all available information about the agent, namely her earnings, her age and her asset position. The idea is to use these life cycle variables as a proxy for labor supply elasticity. The new tax code can be written as follows.

$$\pi^2 = T^L(\hat{w}h, \textcolor{red}{j}, \textcolor{blue}{a}) = \hat{w}h - (1 - \tau_0(\textcolor{red}{j}, \textcolor{blue}{a}))(\hat{w}h)^{1-\tau_1} \quad (27)$$

I assume that the new tax function depends on life cycle observables through the parameter τ_0 . This parameter affects both the average and the marginal taxes rates in the same way. The average tax rates affect the participation decision while the marginal tax rates affect mostly the intensive margin decision. I give a specific functional form to $\tau_0(\textcolor{red}{j}, \textcolor{blue}{a})$.⁷

Ramsey Problem The Ramsey problem is that of maximizing the social welfare

⁷ The parametrization of the labor income tax function takes the form $\tau_0(\textcolor{red}{j}, \textcolor{blue}{a}) = \tau_{00} + \frac{\tau_{01}}{\textcolor{blue}{a}} + (\tau_{02} + \frac{\tau_{03}}{\textcolor{red}{a}})\textcolor{red}{j} + \tau_{04}\textcolor{red}{j}^2 + (\tau_{05} + \tau_{06}\textcolor{blue}{a})\textcolor{red}{j}^3$ Although the parametrization of the labor income tax function seems complicated it is designed to capture the differences in elasticities across age and wealth groups while at the same time respecting the need for redistribution. With respect to age, the function is a polynomial of degree three. This specification can capture well the inverse U-shaped profile of labor supply elasticity. At the same time, since assets are an important determinant of labor supply elasticity I added interaction terms in wealth. This means that different wealth groups will face different life cycle profile of taxes. The functional form can be either an inverse function (first two components) or a linear term (last component). I found this to work better than having only included linear interaction terms. To find the optimum I maximize over $\{\tau_{01}, \tau_{02}, \tau_{03}, \tau_{04}, \tau_{05}, \tau_{06}\}$ while using τ_{00} to keep the government constraint balanced.

function with respect to the given set of policy instruments. The allocations have to respect the government budget constraint and to consist a competitive equilibrium. The problem is written as follows.

$$\max_{\pi^2} W(\pi^2) \quad \text{s.t.} \quad G = \tau_c C(\pi^2) + \tau_k r K(\pi^2) + \int T^L(\pi^2) \quad (28)$$

The problem is solved in two stages. For a *given* set of tax instruments I calculate the competitive equilibrium and make sure that the government budget constraint is satisfied. I then iterate over all possible tax parameters to find the one that maximizes the social welfare function.

Properties of the Optimal Tax Function Figure 7 simulates tax rates paid at the benchmark and the optimal economy. The thick solid line at the left panel, plots the average tax rates paid at the benchmark economy averaged across wealth groups. The line is increasing in age since older cohorts have on average higher earnings. The same panel also plots average tax rates paid at the optimal economy for every wealth quartile. The optimal tax code features three key properties compared to the benchmark.

1) The life cycle path of taxes is hump shaped. At the benchmark economy people between 21 and 35 receive *on average* 12% of their labor income as transfers. People between 35 and 50 pay 6% of their annual earnings in taxes and people between 51 and 65 pay 13.58%. At the optimal economy these age groups pay on average 0.03%, 12% and 2% respectively. The optimal system transfers resources across ages by taxing less groups at the tails of the life cycle. Middle aged cohorts receive on average higher wages than older and especially younger cohorts. This policy is efficient as middle aged groups have a high incentive to stay employed both because of high productivity and because they expect to work for many more years. In contrast, older cohorts with smaller average wages and fewer years of work are more responsive to wage changes. As a result, the optimal tax decreases their tax rates.

2) Wealth conveys important information both for the labor supply elasticity and for the current wage of the agent. If labor income shocks are persistent, wealthy agents are more likely to experience high wages in the current period than low wages. The optimal tax code recognizes this relation and prescribes a positive correlation between labor income and assets early at the life cycle. At the same time, wealthier people can switch to unemployment more easily. This is particularly true *within* older cohorts where elasticities across wealth groups exhibit high dispersion (Table 3). As a result, at the second stage of the life cycle the optimal tax plan lowers the correlation between labor income taxes and asset holdings.

3) The optimal tax plan assigns a unique life cycle path of taxes to every wealth group. Surprisingly, the tax rates for wealth-poor agents increase or weakly decrease along the life cycle. From an efficiency standpoint there is no need to decrease tax rates much, since wealth-poor agents are very inelastic. More importantly, this property provides good incentives to save. If agents faced tax deductions at older ages independently of their assets, they would not save a lot during middle ages. The optimal tax system penalizes this behavior by setting relatively high labor income tax rates for workers who didn't save much along their life cycle.

Links to the Literature It is of interest to compare the last result with recent findings of the dynamic optimal taxation literature. Kocherlakota (2005) finds a negative correlation between the capital tax rate and labor income. In his model, high earners pay a smaller capital tax rate than low earners. This property discourages people from oversaving while young and underproviding work effort when old while collecting generous tax transfers. Thus, capital taxes are negatively correlated with labor income in order to encourage work effort. Here, labor income taxes are negative correlated with asset holdings in order to encourage work effort by workers with very elastic labor supply. This happens to encourage agents to save during middle ages, in spite of expecting lower taxes closer to retirement.

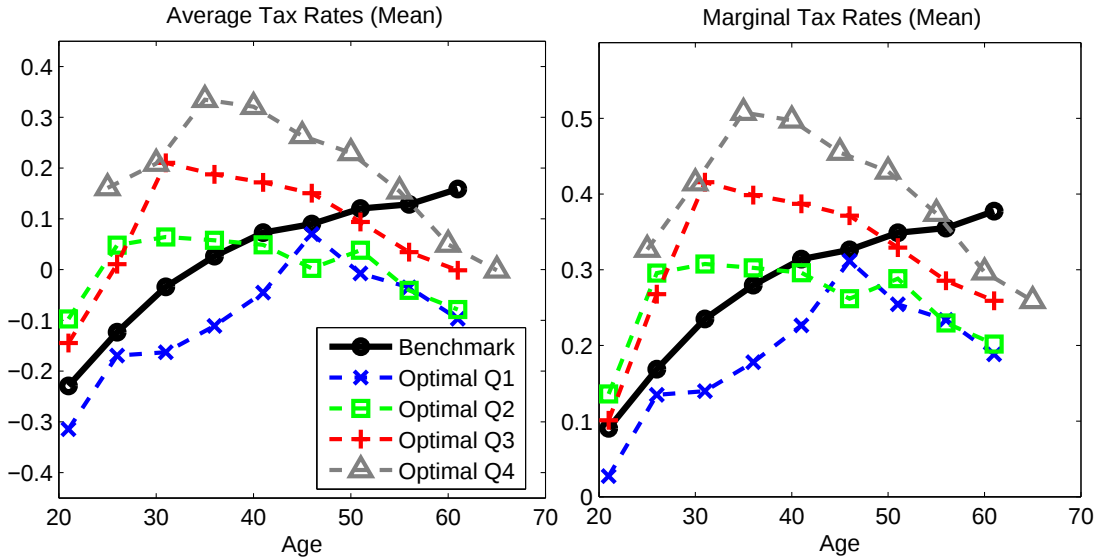


Figure 7: Benchmark and optimal tax system. *Left Panel.* Average tax rates paid at the benchmark economy averaged across wealth groups and average tax rates paid at the optimal economy for every wealth quartile. *Right Panel.* Marginal tax rates paid at the benchmark economy averaged across wealth groups and marginal tax rates paid at the optimal economy for every wealth quartile.

The right panel of Figure 7 plots the marginal tax rates paid at the benchmark economy

averaged across wealth groups and at the optimal economy for every wealth quartile. Both marginal and average tax rates have the same life cycle properties. This is because parameter τ_0 affects both rates in the same way.

Table 3: Aggregate Effects of Policy

Variable	Benchmark	Optimal	21-35	35-50	51-65	65+
Capital	–	+5.37%	+1.44%	-5.69%	+2.45%	+32.65%
Labor	–	+2.98%	+0.01%	-0.96%	+9.78%	–
Hours	–	+1.18%	-0.28%	-1.01%	+7.78%	–
Consumption	–	+4.65%	-0.19%	-5.08%	+5.44%	+13.10%
Wage Rate	–	+0.84%	–	–	–	–
Interest Rate	2.86%	2.70%	–	–	–	–
Consumption Gini	0.2396	+0.0141	-0.0025	+0.0135	+0.0265	+0.0371
Consumption Equivalent	–	+0.85%	+0.64%	+1.07%	+1.81%	+1.95%

Results Table 3 reports the percentage change in key macro aggregates between the benchmark and the optimal economy. Both economies are revenue neutral. Capital increases by **5.37%**. Total labor supply as measured in efficiency units increases by **2.98%**. Hours increase by **1.18%** reflecting a positive composition effect. The optimal tax code encourages work effort by relatively productive individuals. Aggregate consumption increases by **4.65%**. The wage also increases by **0.84%**. On the one hand, labor supply increases which depresses the wage. On the other hand, capital stock also increases making workers more productive. This increases labor demand and consequently the wage. At the optimal economy the interest rate decreases a little. Table 3 also reports the dispersion in consumption as measured by the consumption Gini. The Gini coefficient increases from 0.2396 to 0.2537. To measure the welfare gains we compute the uniform percentage in consumption at each date and each state needed to make a newborn indifferent between the benchmark and the optimal economy, provided that labor effort is the same. If the consumption equivalent is positive, the new economy is preferable since the agent would have to be compensated in order to accept being born in the initial economy. At the new steady state, welfare increases by **0.85%** annual consumption.

Intuition about Results To understand how the optimal tax codes works, I report in Table 3 how these changes are decomposed across age groups. In addition, Figure 8 plots the whole life cycle path of average assets, average consumption and average hours, for both the benchmark and the optimal economy. The most notable feature of the new tax code is the very large increase in market hours by people close to retirement. This is mostly a participation effect. For example, at the benchmark economy, 76% of 60 year old agents are participating in the labor market. This number goes to 81% at the new economy. At the same time, existing workers also increase their effort as a response to

lower marginal tax rates. Unconditional hours for people between 50 and 65 increase by 7.78% (Table 3). More importantly, this large increase in hours does not occur at someone's expense. Although, middle aged workers face larger tax distortions, they still work as much as before. For example, 95.6% of 40 year old workers participate at the benchmark economy. Under the new plan, participation decreases a little, to 94.7%. The total decrease in labor supply for people between 35 and 50 is 1.01% (Table 3). These efficiency gains are a result of the new policy which taxes heavily groups whose labor supply is relatively inelastic.

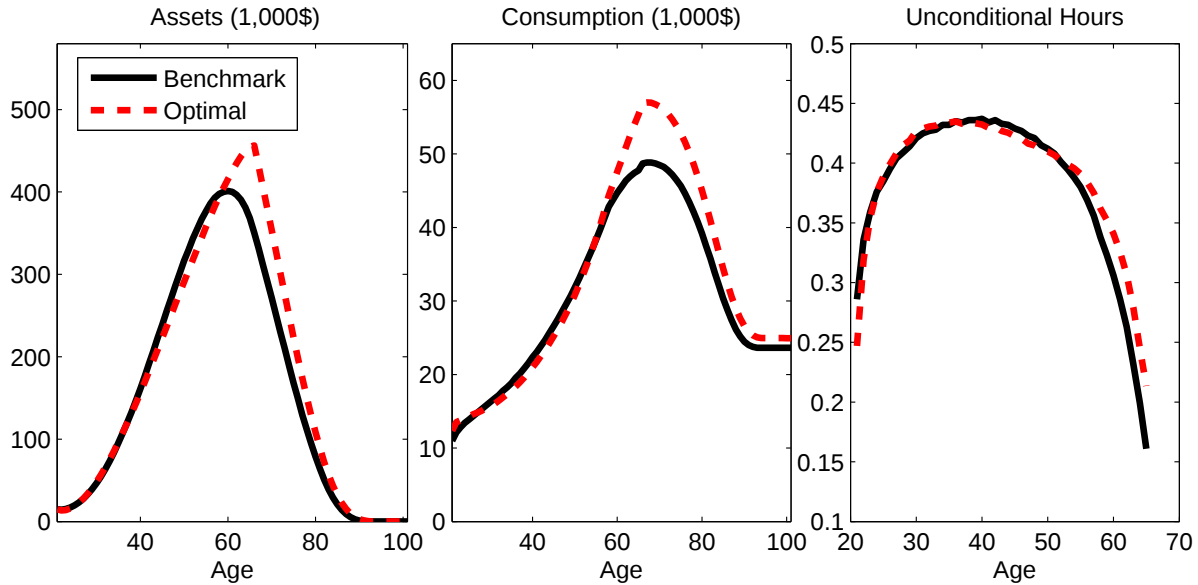


Figure 8: Life cycle profiles: benchmark and optimal system. *Left Panel.* Average assets. *Middle Panel.* Average consumption. *Right Panel.* Average hours.

Efficiency gains are also reflected in the increase in average asset holdings especially at later ages. At the benchmark economy the profile peaks at the age of 60 while at the optimal economy around retirement. This increase in savings occurs for two reasons. First, workers delay their retirement and continue to build up their life cycle savings up to the age of 65. Secondly, the optimal tax system *increases* tax rates for wealth-poor workers who approach retirement. As explained, this property induces workers to keep saving during middle ages in spite of expecting lower tax rates later on. As Table 3 reports, asset holdings for retirees increase by 32.6%. The middle panel of Figure 8 shows the effects of the reform on consumption. Consumption increases significantly, especially for ages close to retirement. At the optimal economy, agents enter retirement having on average a much larger stock of savings. At the same time higher wage and higher labor supply implies a higher social security benefit. All these contribute to a large increase in

consumption of about 13.1% after the age of 65.

Table 3 also decomposes the welfare gains across age groups. Specifically, a newborn would demand 0.85% increase in annual consumption in order to be born at the old steady state instead of the new steady state. The welfare gain is even larger for older cohorts. Being a random middle aged individual (age 35-50) in the new economy is worth a 1.07% increase in annual consumption at the old steady state. Being a random retiree (age 65+) in the new economy is worth a 1.95% increase in annual consumption at the old steady state. Hence the new system produces welfare gains even for age cohorts who pay the highest burden under the new tax plan. A reason the new system fails to produce even larger gains, is that these gains are not allocated uniformly. The optimal tax code decreases taxes on relatively wealthy productive individuals. These groups would have enjoyed high consumption in the first place. As a result, consumption Gini increases by 1.41 percentage points. Nonetheless, the system still produces sizable welfare gains.

Age Dependent Taxation It is of interest to measure the welfare gains if wealth is not part of the information set of the government. This exercise is useful for two reasons. First, because some type of asset holdings cannot be observed. A policy that taxes only age would certainly be easier to implement. Second, we can evaluate better the importance of asset holdings in shaping the optimal tax code. The policy instruments are now

$$\pi^3 = T^L(\hat{w}h, \textcolor{red}{j}) = \hat{w}h - \tau_0(\textcolor{red}{j})(\hat{w}h)^{1-\tau_1} \quad (29)$$

Again the dependence of the tax function on age takes a specific form.⁸ Once again, the optimal life cycle tax path is hump shaped. The main idea is to tax heavily inelastic middle aged workers while decreasing tax rates for older workers and provide redistribution for younger cohorts. Table 4, reports the results of this exercise under the column "Only Age". For comparison Table 4 reports the aggregate effects of the optimal tax code using both age and wealth. Tax cuts towards older individuals increase labor supply by **0.4%**. However, in this case, capital decreases by **1.23%**. According to permanent income hypothesis, the young have less incentive to save in anticipation of lower tax rates closer to retirement. This leads to large distortions at the savings margin. Consumption increases only by **0.82%**. Wages also decrease at the new steady state by **0.59%**, both because labor supply increases and because labor demand decreases. As a result, welfare increases by only **0.11%** annual consumption. A tax code that depends both in age and wealth proves to be a better policy tool. The code specifies high tax rates for older workers with low asset holdings. This encourages workers to keep saving during their middle ages.

⁸ The parametrization is now $\tau_0(\textcolor{red}{j}) = \tau_{00} + \tau_{01}\textcolor{red}{j} + \tau_{02}\textcolor{red}{j}^2 + \tau_{03}\textcolor{red}{j}^3$.

Table 4: Aggregate Effects of Different Reforms

Variable	Benchmark	Age and Wealth	Only Age
Capital	–	+5.37%	-1.23%
Labor	–	+2.98%	+0.40%
Hours	–	+1.18%	+0.68%
Consumption	–	+4.65%	+0.82%
Wage Rate	–	+0.84%	-0.59%
Interest Rate	2.86%	2.70%	2.98%
Consumption Gini	0.2396	0.2537	0.2393
Consumption Equivalent	–	+0.85%	+0.11%

7 Discussion on Heterogeneity in Elasticity and Efficiency Gains

I have argued so far that a policy which uses information on age and wealth can generate large efficiency gains by shifting the tax burden away from elastic workers. I now consider an experiment that shows that heterogeneity in labor supply elasticity is crucial in generating these gains. To test this assumption I compare two models. The first is a “constant elasticity model” (CEM) which is a divisible labor economy with a Frisch utility function.

$$U = \log c_j + \psi \frac{h_j^{1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}} \quad (30)$$

With this specification all agents have the same labor supply elasticity which is given by parameter γ .⁹ The second model is the benchmark “heterogenous elasticity model” (HEM). To perform the experiment I assume that both economies experience a reform which decreases tax rates for older cohorts by taxing more younger generations. The magnitude the reform is given by the ratio T_o^L/T_y^L which measures total taxes paid by older generations relative to total taxes paid by younger generations.¹⁰ If heterogeneity is not important, the same policy will have the same effects on the two economies. Table 5 reports the statistics of this exercise. At the HEM the policy reduces the ratio by 2.05 (from 4.13 to 2.08) while at the CEM the policy decreases the ratio by 2.18 (from 5.58 to 2.40). At the HEM the new tax system increases labor and hours by 1.59% and 2.18% respectively. Consistent with our main experiments younger cohorts decrease their labor supply and hours a little (2.59% and 0.71%) while older cohorts respond more elastically

⁹In this experiment I used a value of γ equal to one.

¹⁰I consider an age dependent policy for simplicity. Results would not change if we had used a tax code that depends on both age and wealth.

(4.88% and 7.27%). In contrast, the policy has minor aggregate effects on the CEM. Young cohorts decrease their labor supply by 1.38% and hours by 0.62% while the old increase their labor by 2.01% and hours by 2.08%. These two offset each other so that at the aggregate the reform increases total labor supply by 0.17% and total hours by 0.65%. Both economies experience a decrease in capital. However the HEM experiences an increase in consumption since older cohorts work and produce more. This exercise shows that heterogeneity in elasticity drives the welfare gains reported in our main experiment.

Table 5: HEM vs CEM

Variable	T_o^L/T_y^L	HEM _y	HEM _o	HEM	T_o^L/T_y^L	CEM _y	CEM _o	CEM
Capital	-2.05	+1.64%	-14.5%	-6.25%	-2.18	+0.85%	-11.3%	-5.16%
Labor	-2.05	-2.09%	+4.88%	+1.59%	-2.18	-1.38%	+2.08%	+0.17%
Hours	-2.05	-0.71%	+7.27%	+2.18%	-2.18	-0.62%	+2.01%	+0.65%
Consumption	-2.05	-7.03%	+1.34%	+0.56%	-2.18	-7.27%	-0.56%	-1.36%

8 Conclusion

This paper develops a dynamic life cycle model with heterogeneous agents and endogenous working choice. Agents make decisions about their participation on the market (extensive margin) and conditional on participating, about the amount of hours they will supply (intensive margin). The model produces significant dispersion in wage elasticities, variation which originates from the extensive margin. I find that wealthy agents and people close to retirement are the most sensitive groups in the economy. The main policy recommendations from an optimal Ramsey tax exercise is 1) to cut taxes for relatively wealthy people close to retirement 2) to raise taxes for middle-aged groups and 3) to decrease taxes for young cohorts. This policy leads to large gains. Total supply of labor as measured in efficiency units increases by 2.98%, capital increases by 5.37% and consumption increases by 4.65%. Welfare increases by 0.85% in terms of consumption equivalent variation.

It is important to discuss now several points that can motivate future research related to this paper. The first is the analysis of the transition path between the steady states. Analyzing the behavior of the economy along the transition allows to determine which groups will support the new tax plan. Agents who under the new plan, have higher life-time utility relative to the old system, will be in favor of the reform. Second, the tax proposed in this exercise is optimal given the restricted set of tax instruments. Ideally

we would like to construct a tax schedule that is an arbitrary function of age, wealth and earnings. This kind of exercise follows the Mirrlees tradition of optimal taxation problems. However, in both approaches the government would have an incentive to differentiate tax rates across elasticity groups. Hence, the main recommendations of our simpler Ramsey tax exercise would remain intact. Third, the model abstracted from bequests and inter-generational transfers. Incorporating such an element would more likely lower the labor supply elasticity of older cohorts provided these groups cared about the utility of younger generations.

References

- Aiyagari, S. Rao (1994).** “Uninsured Idiosyncratic Risk and Aggregate Saving.” *Quarterly Journal of Economics*, 109(3): 659-684.
- Alan, Sue (2006).** “Entry costs and stock market participation over the life cycle.” *Review of Economic Dynamics*, 9(4): 588-611.
- Alesina, Alberto, Andrea Ichino and Loukas Karabarbounis (2011).** “Gender based taxation and the division of family chores”. *American Economic Journal: Economic Policy*, Vol. 3(2): pp. 1-40.
- Alexopoulos, Michelle and Tricia Gladden (2006).** “Wealth, reservation wages, and labor market transitions in the US: Evidence from the SIPP.” University of Toronto, mimeo.
- Altonji, Joseph (1986).** “Intertemporal substitution in labor supply: Evidence from micro data.” *Journal of Political Economy* 94: 176-285.
- Blundell, Richard, Antoine Bozio and Guy Laroque (2011).** “Extensive and Intensive Margins of Labour Supply: Working Hours in the US, UK and France”, *Working paper IFS*.
- Blundell, Richard and Andrew Shephard (2011).** “Hours of Work and the Optimal Taxation of Low Income Families.” *Forthcoming in Review of Economic Studies*.
- Chang, Yongsung and Sun-Bin Kim (2006).** “From Individual to Aggregate Labor Supply: a Quantitative Analysis Based on a Heterogeneous Agent Macroeconomy.” *International Economic Review* 47(1), February: 1-27.
- Chang, Yongsung, Sun-Bin Kim and Frank Schorfheide (2010).** “Labor-Market Heterogeneity, Aggregation, and the Lucas Critique.” NBER Working Paper No. 16401.
- Cho, Jank-Ok and Richard Rogerson (1988).** “Family Labor Supply and Aggregate Fluctuations.” *Journal of Monetary Economics*, 21: 233-245.
- Chon, Philippe and Laroque, Guy (2011).** “Optimal taxation in the extensive model.” *Journal of Economic Theory*, 146(2): 425-453.
- Conesa, J. C., S. Kitao, and D. Krueger (2009).** “Taxing capital? not a bad idea after all!” *American Economic Review*, 99(1): 25-48.

Conesa, Juan Carlos, and Dirk Krueger (2006) On the Optimal Progressivity of the Income Tax Code.” *Journal of Monetary Economics*, 53(7): 1425-1450.

Davila, Julio, Jay Hong, Per Krusell, and Jose-Victor Rios-Rull (2005). “Constrained Efficiency in the Neoclassical Growth Model with Uninsurable Idiosyncratic Shocks.” PIER Working Paper No. 05-023.

Domeij, David, and Martin Floden (2006). “The Labor Supply Elasticity and Borrowing Constraints: Why Estimates are Biased.” *Review of Economic Dynamics*, 9(2): 242-262.

Erosa, Andres, and Martin Gervais (2002). “Optimal Taxation in Life Cycle Economies.” *Journal of Economic Theory*, 105(2): 338-369.

Erosa, Andres, Gueorgui Kambourov and Luisa Fuster (2011). “Towards a Micro-Founded Theory of Aggregate Labor Supply.” University of Toronto, mimeo.

Farhi, Emmanuel and Ivan Werning (2010). “Insurance and Taxation over the Life Cycle.” NBER Working Paper No. 16749.

French, Eric (2005). “The Effects of Health, Wealth, and Wages on Labor Supply and Retirement Behavior.” *Review of Economic Studies* 72: 395-427

Fukushima, Kenichi (2010). “Quantifying the welfare gains from flexible dynamic income tax systems.” University of Minnesota, mimeo.

Gomme, Paul, Richard Rogerson, Peter Rupert and Randall Wright (2004). “The Business Cycle and the Life Cycle.” *NBER Macroeconomics Annual*: 415-461.

Gourinchas, Pierre-Olivier and Jonathan Parker (2002). “Consumption over the Life Cycle.” *Econometrica*, 70(1): 4791.

Gourio, Francois and Pierre-Alexandre Noual (2009). “The Marginal Worker and the Aggregate Elasticity of Labor Supply.” Boston University, mimeo.

Guner, Nezih, Remzi Kaygusuz and Gustavo Ventura (2011). “Taxing Women: A Macroeconomic Analysis.” *Journal of Monetary Economics*, forthcoming.

Hansen, Gary (1985). “Indivisible Labor and the Business Cycle.” *Journal of Monetary Economics*, 16(3): 309-327.

Hansen, Gary (1993). “The Cyclical and Secular Behavior of the Labor Input: Comparing Efficiency Units and Hours Worked.” *Journal of Applied Econometrics*, (8): 71-80.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante (2010). “Redistributive Taxation in a Partial-Insurance Economy.” Federal Reserve Bank of Minneapolis, Working Paper.

Heckman, James and Thomas MaCurdy (1980). “A Life Cycle Model of Female Labor Supply.” *Review of Economic Studies*, 47(1): 47-74.

Imrohoroglu, Selahattin and Sagiri Kitao (2009). “Labor Supply Elasticity and Social Security Reform.” *Journal of Public Economics*, 93(7-8): 867-878.

Imrohoroglu, Selahattin and Sagiri Kitao (2009). “Social Security, Benefit Claiming and Labor Force Participation: A Quantitative General Equilibrium Approach.” *Federal Reserve Bank of New York*, Working Paper.

Jaimovich, Nir and Henry E. Siu (2009). “The Young, the Old, and the Restless: Demographics and Business Cycle Volatility.” *American Economic Review*, 99(3): 804-826.

Jesper, Linde (2001). “Testing for the Lucas Critique: A Quantitative Investigation.” *American Economic Review*. 91, 986-1005.

Kaplan, Greg (2011). “Inequality and the Life Cycle”. PIER Working Paper No. 11-014.

Kitao, Sagiri (2010). “Labor-dependent capital income taxation.” *Journal of Monetary Economics*, 57(8): 959-974.

Kocherlakota, Narayana (2005). “Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation.” *Econometrica* 73(5): 1587-1621.

MaCurdy, Thomas (1981). “An empirical model of labor supply in a life cycle setting.” *Journal of Political Economy*. 88: 1059-85.

Michau, Jean-Baptiste (2011). “Optimal Redistribution with Intensive and Extensive Labor Supply Margins: A Life Cycle Perspective”. *Mimeo*.

Mirrlees, James A. (1971). “An exploration in the theory of optimum income taxation.” *Review of Economic Studies*. 38, 175,208.

Mulligan, Casey B. (2001) “Aggregate Implications of Indivisible Labor.” *Advances in Macroeconomics* 1 (1).

Pijoan-Mas, Joseph (2006). “Precautionary savings or working longer hours?” *Review of Economic Dynamics*. 12: 23-36.

Pistaferri, Luigi (2003). “Anticipated and Unanticipated Wage Changes, Wage Risk, and Intertemporal Labor Supply.” *Journal of Labor Economics*, Vol. 21, No. 3, July.

Prescott, Edward, Richard Rogerson and Johanna Wallenius (2009). “Lifetime Aggregate Labor Supply with Endogenous Workweek Length.” *Review of Economic Dynamics*. 12: 23-36.

Rogerson, Richard (1988). “Indivisible Labor, Lotteries and Equilibrium.” *Journal of Monetary Economics*. 21(1): 1-16.

Rogerson, Richard and Johanna Wallenius (2009). “Micro and Macro elasticities in a Life Cycle Model with Taxes.” *Journal of Economic Theory*, 144: 2277-2292.

Saez, Emmanuel (2001). “Using Elasticities to Derive Optimal Income tax Rates.” *Review of Economic Studies*. 68, 205-229.

Salanie, Bernard (2003). *The Economics of Taxation*. MIT Press.

Social Security Administration (2005), Annual Statistical Supplement to the Social Security Bulletin, Washington, DC: Social Security Administration.

Storesletten, Kjetil, Chris Telmer and Amir Yaron (2004). “Consumption and risk sharing over the life cycle.” *Journal of Monetary Economics*, vol. 51, pp. 609-633, 144: 2277-2292.

Weinzierl, Matthew (2011). “The Surprising Power of Age-Dependent Taxes.” *Review of Economic Studies*, (4): 1490-1518.

Appendix A: Data Source

I use data from the PSID. I use a wide range of waves from 1970 to 2005. The survey was conducted annually up to 1997 and biannually from 1999 to 2005. For each year data are collected about the age and sex of the head of the household, the total amount of hours supplied and his labor income. For hours I use the variable "Head Annual Hours of Work". This variable represent the total annual work hours on all jobs including overtime. For the labor income I use the variable "Head Wage" which is wages and salaries. Apart from the Head I also collect information for the wife of the head if the head is not single. The variable of interest is "Wife Wage". I restrict the sample to only 1) head of households 2) males 3) head of households that are the primary earners. The last condition requires that the wife either is making less than the median annual wage over all wives for a given year or that she is making less than half of the head's annual wages and salaries. My measure of wealth is the variable WEALTH2 as found in specific waves of PSID. This variable is constructed as sum of values of several asset types (family farm business, family accounts, assets, stocks, houses and other real estate etc.) net of debt value.

Appendix B: Preference Heterogeneity and Search Cost

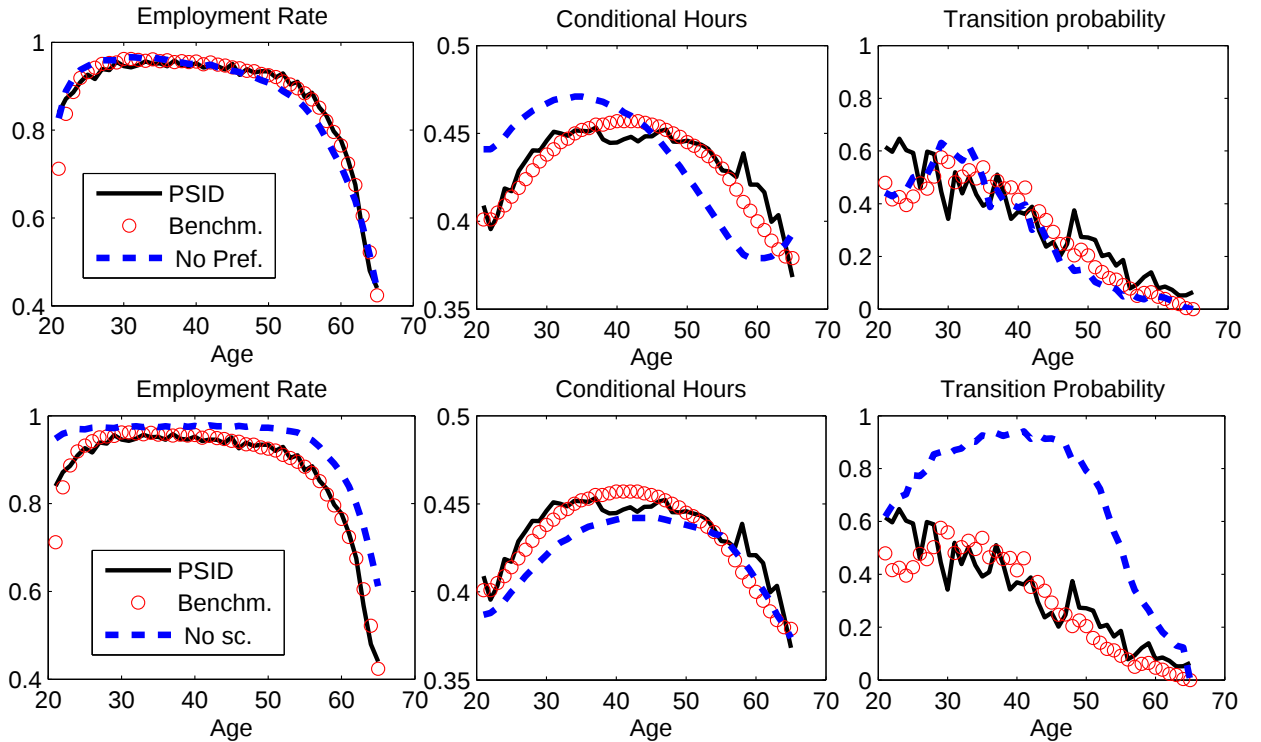


Figure 9: Upper Panel: Model without preference heterogeneity. Lower Panel: Model without search cost.

Appendix C: Solution Algorithm (Benchmark)

This is a general equilibrium problem. I search for the market prices $\{w, r\}$ which clear the markets and the transfers Tr that equal to the total amount of savings by the deceased. To solve this problem I guess prices w^0, r^0 and transfers Tr^0 . The dynamic program is solved by backwards induction.

1. *Grid Construction:* I specify a grid of 200 points for the assets. I make sure that the upper bound is large enough. More grid points are assigned to lower values. The continuous process of transitory labor income shock x is discretized into a six state Markov chain using the methodology described by Tauchen (1986). The unconditional variance of the process is equal to $\sigma_x^2 = \frac{\sigma_\eta^2}{1-\rho^2}$. I set the grid's bounds to $[-\lambda\sigma_x, \lambda\sigma_x]$ and $\lambda = 1.2 \times \log(6)$. The space is divided into 6 equally distanced points. The corresponding transition matrix is

$$Q(\eta' | \eta) = \begin{pmatrix} 0.8904 & 0.1096 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0427 & 0.8690 & 0.0883 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0551 & 0.8747 & 0.0702 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0702 & 0.8747 & 0.0551 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0883 & 0.8690 & 0.0427 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.1096 & 0.8904 \end{pmatrix}$$

The transition process implies an invariant distribution equal to $\Pi^* = [0.0671, 0.1722, 0.2758, 0.2758, 0.1722, 0.0671]$. Lastly I transform the grid into consumption units by taking the exponential and I normalize by using the invariant distribution. The grid used in the simulation is the following: $x = [0.1662, 0.3001, 0.5418, 0.9783, 1.7662, 3.188]$. The permanent component of labor income $\log z$ is distributed normally with mean zero and variance σ_z^2 . Setting the grid for the log component to $\log z = [-0.4, 0.4]$ and assuming that the population is divided equally among the two types gives a variance of $\sigma_z^2 = 0.16$.

2. *Guessing prices:* To solve the problem I guess a set of firm inputs K_d^0, L_d^0 . Using the first order conditions these imply a set of prices $\{w, r\}$. I also guess a value for transfers Tr^0 .
3. *Solving for the Retirees:* The problem is solved by backwards induction. Using that $a'_{81} = 0$ we can easily back out the value function $V_{81}(a)$. To find $V_{80}(a)$ I solve the one dimensional optimization problem over a' . I use golden search and spline interpolation to approximate the value function for out of the grid points. Using this method we can get a series of value functions $\{V_j(a)\}_{j=66}^{81}$ and policy functions $\{g^a(a)\}_{j=66}^{81}$.
4. *Solving for Workers:* The problem for working cohorts requires calculating two different value functions V_j^E, V_j^U . To calculate V_j^E we need to optimize over both a' and h . I proceed as follows: for every state vector $\omega = (a, z, x, S_{-1})$ and potential savings choice

a' , I use bisection to solve the static first order condition $\psi(1-h)^{-\theta} = \frac{\dot{w}(1-T^{L'}(\dot{w}h))}{c(1+\tau_c)}$ to get $h(a'; \omega)$. The problem is now reduced into a one dimensional problem. Finding $g_j^a(\omega)$ allows to back out $g_j^h(\omega)$. Using both we can find the value of being employed $V_j^E(\omega)$. The value for the unemployed V_j^U is easier to obtain since it requires a one optimization problem. Participation is found by comparing the two functions: $V_j = \max \{V_j^E, V_j^U\}$. Using this method we can get a series of value functions $\{V_j^E(\omega), V_j^U(\omega), V_j(\omega)\}_{j=21}^{65}$ and policy functions $\{g^a(\omega), g^h(\omega)\}_{j=21}^{65}$.

5. *Simulation:* At this stage I generate a cross section of 10,000 individuals and track them over their lifetime. Exogenous variables (productivity) evolve based on the Markov process. Endogenous variables are consistent with the decision rules. Aggregating gives K^s, L^s and Tr .
6. The new guess is found by $K_d^1 = \chi K_d^0 + (1 - \chi)K_s$, $L_d^1 = \chi L_d^0 + (1 - \chi)L_s$ and $Tr^1 = \chi Tr^0 + (1 - \chi)Tr$. To guarantee convergence I set χ very close to 1. Using the new guesses I go back and solve the problem again. This process stops when all our guesses are sufficiently accurate.

Appendix D: Tables

Table 6: Exogenous Calibration:

Parameter	Description	Reference
$n = 1.1\%$	Population growth	US long-run average
$\alpha = 0.36$	Technology parameter	Capital share
$\theta = 2$	Preference parameter	EFK (2010)
$\tau_1 = 0.26$	Tax parameter	HSV(2010)
$\tau_{ss} = 0.106$	Social security tax	Kitao (2010)
$\tau_c = 0.05$	Consumption Tax	Kitao and Imrohoroglu (2010)
$\tau_k = 0.30$	Capital Tax	Kitao and Imrohoroglu (2010)
$\sigma_{\bar{a}_{j=1}} = 1.96$	Standard Deviation of Initial Assets	Allen (2006)
$\{\epsilon_j\}$	Life cycle productivity	Hansen (1993)
$\{s_j\}$	Conditional survival probabilities	Social security administration (2005)

Table 7: Endogenous Calibration

Parameter	Description	Target
$\beta = 1.005$	Discount Factor	$K/Y = 3.2$
$\delta = 0.0816$	Depreciation rate	$I/Y = 0.25$
$\alpha_0 = 0.75$	Utility parameter	Average Hours=0.42
$\alpha_1 = -0.066$	Utility parameter	Average Hours=0.438
$\alpha_2 = 0.005$	Utility parameter	Average Hours=0.393
$\bar{a}_{j=1} = -4.0$	Mean of lognormal distribution	$\frac{\bar{a}_{21-30}}{\bar{a}_{21-65}} = 0.103$
$FC_y = 0.29$	Fixed cost of working (young)	Employment ₂₁₋₄₂ = 0.94
$FC_o = 0.26$	Fixed cost of working (old)	Employment ₄₃₋₆₅ = 0.82
$\lambda = 0.045$	Probability of separation	$p(E \rightarrow U) = 0.06$
$sc_y = 7.1$	Search cost	$p(U \rightarrow E)_{21-42} = 0.48$
$sc_o = 4.6$	Search cost	$p(U \rightarrow E)_{43-65} = 0.17$
$\tau_0 = 0.32$	Tax rate	$G/Y = 0.2$
$\sigma_z^2 = 0.16$	Variance of permanent shock	$\text{Var}(y_{22}) = 0.27$
$\rho = 0.96$	Persistence of AR(1)	Linear Slope of profile
$\sigma_\eta^2 = 0.04$	Variance of AR(1)	$\text{Var}(y_{60}) = 0.9$