Quantifying the Welfare Gains From Flexible Dynamic Income Tax Systems[1]

Kenichi Fukushima
University of Minnesota and FRB Minneapolis
fuku0028@umn.edu

Job Market Paper

October 22, 2009

[1]Special thanks to Narayana Kocherlakota and Chris Phelan for their advice and support. Thanks also to workshop participants, Laura Kalambokidis, Ctirad Slavik, and Yuichiro Waki for their comments. The Minnesota Supercomputing Institute provided valuable computational resources. The views expressed here are not necessarily those of the Federal Reserve System.

Abstract

This paper sets up an overlapping generations general equilibrium model with incomplete markets similar to Conesa, Kitao, and Krueger's (2009) and uses it to simulate a policy reform which replaces an optimal flat tax with an optimal non-linear tax that is allowed to be arbitrarily age and history dependent. The reform shifts labor supply toward productive households and thereby increases aggregate productivity. This leads to higher per capita consumption and shorter per capita hours. Under a utilitarian social welfare function that places equal weight on all current and future cohorts, the implied welfare gain is worth more than 10% in lifetime consumption equivalents.

1 Introduction

In modern societies, income taxation by the government plays two beneficial roles: it raises revenue for funding public goods and provides social insurance by redistributing from the fortunate to the unfortunate. The associated cost is that taxes negatively affect current and future production possibilities by discouraging labor supply and investment. An important goal in macroeconomics and public finance is to understand how these forces are best balanced given a well-defined notion of social welfare.

In a recent series of papers, Conesa and Krueger (2006) and Conesa, Kitao, and Krueger (2009) take a quantitative approach to this question using a dynamic general equilibrium model that incorporates many of the relevant ingredients, such as endogenous labor supply, capital accumulation, life cycles, and uninsurable idiosyncratic wage risk whose structure is consistent with the empirical findings of Storesletten, Telmer, and Yaron (2004) and others. In doing so, Conesa, Kitao, and Krueger (CKK hereafter) solve for the optimal tax system under the restriction that taxes cannot depend on income histories or age. Their findings broadly support Hall and Rabushka's (1995) proposal that labor and asset income be taxed at a moderate, flat rate with a fixed deduction per household.

The restrictions that CKK impose on the set of tax instruments, however, are not quite ideal. There is an obvious, general sense in which they cannot be—restricting the choice set in an optimization problem can never help—but there is also a specific theoretical reason to suspect that they create a positive and possibly significant loss in this instance. The latter concern derives from several recent studies, collectively referred to as the New Dynamic Public Finance (NDPF) by Kocherlakota (2009), which theoretically examine the optimal structure of labor and asset income taxes when they are allowed to be arbitrarily non-linear and age/history dependent. Two general lessons that have emerged from this literature are that optimal taxes are: (i) necessarily non-separable between labor and asset income; and (ii) most likely history dependent as well when wages are random and persistent as in CKK's model (Albanesi and Sleet, 2006, Golosov and Tsyvinski, 2006, Kocherlakota, 2005). The flat tax whose optimality obtains under CKK's restrictions has neither property.

To address this concern, this paper sets up a model similar to CKK's and uses it to quantify the welfare gain from replacing CKK's optimal flat tax with an optimal non-linear tax that is allowed to be arbitrarily age and history dependent. The gain turns out to be large: under a social welfare function that places equal weight on all current and future cohorts, it is worth more than a 10 percent increase in consumption for every household at all dates and contingencies. This gain mostly comes from higher per capita consumption and shorter per capita hours. These improvements are supported by a massive shift of labor supply toward productive households, which effectively increases aggregate productivity.

The main technical challenge in carrying out this analysis is computational, and CKK in fact cite this as their primary reason for formulating the problem the way they did:

> Ideally one would impose no restrictions on the set of tax functions the government can choose from. Maximization over such an unrestricted set is computationally infeasible, however. (Conesa, Kitao, and Krueger, 2009, p. 34)

This paper overcomes this challenge by analytically simplifying the unrestricted optimal tax problem before resorting to numerical methods. The procedure has three steps: The first step follows the NDPF by using mechanism design and Kocherlakota's (2005) implementation result to reduce the problem to a fictitious social planning problem which maximizes social welfare subject to resource and incentive constraints. The second step then establishes a theoretical result which further reduces this planning problem to a "partial equilibrium" dynamic mechanism design problem without capital. This eliminates the intractability of the former that comes from the model's general equilibrium structure. The third step wraps up by applying a recursive method devised by Fukushima and Waki (2009) to manage the computational intensity that comes from wage persistence.

There are several recent papers that also use mechanism design to address quantitative questions on optimal taxation, but do so using partial equilibrium models without capital and with stylized forms of wage risk.[1] An early paper by Golosov and Tsyvinski (2006) studies the optimal structure of disability insurance using a model in which agents are subject to a two-state shock sequence (disability or not), where disability is an absorbing state. A more recent paper by Huggett and Parra (2009) speaks to the optimal structure of tax systems more generally, but they are able to use mechanism design only when households experience no wage risk after entering the labor market. Weinzierl (2008) employs a richer specification of wage risk, but only in a two-period setting. This paper therefore expands the technological frontier of this literature by making it possible to handle general equilibrium models with capital accumulation and richer, empirically better motivated specifications of wage risk. This bridges a gap between this literature and the quantitative incomplete markets literature, and is, in my view, intrinsically valuable as well given the plausible importance of these elements in assessing how tax systems are best structured.

[1] An interesting outlier is Farhi and Werning (2009), who use a model with a general structure that allows for capital accumulation and arbitrary forms of labor market risk. They focus on a partial reform which keeps the labor allocation intact and find that it generates a modest welfare gain (relative to a benchmark allocation that resembles what is currently observed in the U.S.). This paper considers a "full" reform which allows for labor reallocations and finds that there are potentially large gains from doing so. On the other hand, this conclusion is more model-dependent than Farhi and Werning's.

2 Model

The model is almost identical to CKK's, except for: (i) the fact that the government is given access to a richer set of tax instruments; and (ii) several technical differences that make the model mathematically better behaved.

Environment. Time flows $t = 1, 2, 3, ...,$ and in each period a measure $(1 + \eta)^{t-1}$ of households is born. Each household lives for at most J periods and its lifetime utility is the expected value of

$$\sum_{j=1}^{J} \beta^{j-1} U(c_j, l_j)$$

where c_j and l_j are its consumption and hours of work at age j, respectively. Here, $U(c, l) = u(c) - v(l)$, where u', $-u''$, v', and v'' are all non-negative and v is isoelastic.

At each age j, a household draws an idiosyncratic skill shock θ_j from a finite set $\Theta_j \subset \mathbb{R}_{++}$, which enables it to transform l_j units of labor into $n_j = \theta_j l_j$ units of effective labor. For technical reasons I assume that n_j is bounded from above by a large constant $n_{\max}$. The skill shock process is first order Markov and has strictly positive transition probabilities. Households also face skill-independent mortality risk, and ψ_j denotes the probability of survival between ages $j - 1$ and j. The distribution of both shocks across households is i.i.d. and satisfies the law of large numbers. Let $\theta^j \equiv (\theta_1, ..., \theta_j) \in \Theta^j \equiv \Theta_1 \times \cdots \times \Theta_j$ and $\theta_i^j \equiv (\theta_i, ..., \theta_j) \in \Theta_i^j \equiv \Theta_i \times \cdots \times \Theta_j$, and let π_j denote the joint density of survival and skill draws. The measure of age j households in period t with skill history θ^j is then $\mu_{jt}(\theta^j) = (1 + \eta)^{t-j} \pi_j(\theta^j)$.

The technology is described by the aggregate resource constraint

$$C_t + K_{t+1} - (1 - \delta)K_t + G_t \leq F(K_t, N_t) \tag{1}$$

for each t, where the initial capital stock K_1 is given. Here, C_t is aggregate consumption, K_t is the capital stock, N_t is aggregate effective labor, $G_t = (1+\eta)^{t-1}G$ is an exogenous expense on public goods, δ is the depreciation rate of capital, and $F : \mathbb{R}_+^2 \to \mathbb{R}_+$ is a constant-returns-to-scale (CRS) aggregate production function which is increasing, concave, and continuously differentiable. Using CRS, let $\hat{r}(K/N) \equiv F_K(K, N) - \delta$ and $\hat{w}(K/N) \equiv F_N(K, N)$. The Inada conditions $\lim_{\kappa \to 0} \hat{r}(\kappa) = \infty$ and $\lim_{\kappa \to \infty} \hat{r}(\kappa) = -\delta$ hold.

Allocations. An allocation is a sequence $x = ((c_{jt}, n_{jt})_{j=1}^J, K_t)_{t=1}^\infty$, where $c_{jt} : \Theta^j \to \mathbb{R}_+$, $n_{jt} : \Theta^j \to [0, n_{\max}]$, and $K_t \in \mathbb{R}_+$ for each j and t. Here, $c_{jt}(\theta^j)$ is the consumption of an age

j household at calendar time t whose skill history up to that point is θ^j. This household's date of birth is the end of period $t-j$. The interpretation of $n_{jt}(\theta^j)$ is analogous.

Thus under allocation x, a household from cohort $t \geq 0$ obtains lifetime utility:

$$V_t(x) = \sum_{j=1}^{J} \sum_{\theta^j} \beta^{j-1} U(c_{j,t+j}(\theta^j), n_{j,t+j}(\theta^j)/\theta_j)\pi_j(\theta^j)$$

whereas one from cohort $t = 1-i < 0$ with skill history θ^{i-1} at date $t=1$ obtains:

$$V_{1-i}(x;\theta^{i-1}) = \sum_{j=i}^{J} \sum_{\theta_i^j} \beta^{j-i} U(c_{j,1-i+j}(\theta^j), n_{j,1-i+j}(\theta^j)/\theta_j)\pi_j(\theta_i^j|\theta_{i-1}).$$

Abusing notation, let $V_{1-i}(x) = \sum_{\theta^{i-1}} V_{1-i}(x;\theta^{i-1})\pi_{i-1}(\theta^{i-1})$.

An allocation is stationary if each (c_{jt}, n_{jt}) is independent of t and K_t grows at constant rate $(1+\eta)$.

Markets and Tax Policies. Commodity and factor markets operate as usual: a number of privately-held firms own the production technology; households rent labor and capital services to the firms and use the income they receive in return to purchase goods for consumption and investment; and all market transactions are competitive. Let r_t denote the interest rate and w_t the price of effective labor.

Insurance markets for skill risk are assumed to be missing however, and this creates room for the government to enhance social welfare by providing social insurance through income taxation (broadly defined, so as to include such functionally related arrangements as social security). Annuity markets are missing as well.

Given the goal of this paper, I allow the government to choose from a very rich set of tax instruments. Thus, taxes are allowed to be arbitrary non-linear functions of calendar time, age, income history, and any other messages received (such as statements pertaining to unemployment, disability, or retirement). The government can also issue debt, commit to future actions, and confiscate any bequests (all of which are accidental in this model). Following Mirrlees (1971), however, I do not allow taxes to depend directly on households' skill levels that realize after date $t=1$. I take an agnostic stand on why this restriction may be difficult to overcome in reality, given its irrelevance for my analysis.

Thus a tax policy is formally a sequence $T = ((M_{jt}, \tau_{jt})_{j=1}^{J}, B_t)_{t=1}^{\infty}$, where M_{jt} is the set of messages that an age j household is allowed to send to the government at date t, τ_{jt} describes the tax obligation of an age j household at time t as a function of its history h_{jt} (a complete record of the household's income and messages sent to the government up to that date), and B_t is the amount of debt issued by the government in period t. Let $\mathcal{T}^*$ denote the set of all tax policies T.

Equilibrium. An equilibrium given a tax policy T and an initial wealth distribution $(k_{i,1}, b_{i,1})_{i=2}^J$ is a sequence of household-level quantities $((c_{jt}, n_{jt}, k_{jt}, b_{jt}, m_{jt})_{j=1}^J)_{t=1}^\infty$, aggregate quantities $(C_t, N_t, K_t)_{t=1}^\infty$, and factor prices $(w_t, r_t)_{t=1}^\infty$ that satisfy the following conditions.

1. The marginal product conditions $r_t = F_K(K_t, N_t) - \delta$ and $w_t = F_N(K_t, N_t)$ hold for each t.

2. The quantities $(c_{j,t+j}, n_{j,t+j}, k_{j+1,t+j+1}, b_{j+1,t+j+1}, m_{j,t+j})_{j=1}^J$ for cohort $t \geq 0$ households maximize $V_t(x)$ subject to the flow budget constraints

$$\begin{aligned} &c_{j,t+j}(\theta^j) + k_{j+1,t+j+1}(\theta^j) + b_{j+1,t+j+1}(\theta^j) \\ &\qquad \leq w_{t+j}n_{j,t+j}(\theta^j) + (1 + r_{t+j})(k_{j,t+j}(\theta^{j-1}) + b_{j,t+j}(\theta^{j-1})) - \tau_{j,t+j}(h_{j,t+j}(\theta^j)) \end{aligned} \quad (2)$$

and

$$h_{j,t+j}(\theta^j) = (w_{t+i}n_{i,t+i}(\theta^i), r_{t+i}(k_{i,t+i}(\theta^{i-1}) + b_{j,t+j}(\theta^{j-1})), m_{i,t+i}(\theta^i))_{i=1}^j \quad (3)$$

$$(c_{j,t+j}(\theta^j), n_{j,t+j}(\theta^j), k_{j,t+j}(\theta^{j-1}) + b_{j,t+j}(\theta^{j-1}), m_{j,t+j}(\theta^j)) \in \mathbb{R}_+ \times [0, n_{\max}] \times \mathbb{R}_+ \times M_{j,t+j} \quad (4)$$

for each j and θ^j, given the initial condition $k_{1,t+1}(\theta^0) = b_{1,t+1}(\theta^0) = 0$.

3. The quantities $(c_{j,1-i+j}(\theta^{i-1}, \cdot), n_{j,1-i+j}(\theta^{i-1}, \cdot), k_{j+1,2-i+j}(\theta^{i-1}, \cdot), m_{j,1-i+j}(\theta^{i-1}, \cdot))_{j=i}^J$ for cohort $t = 1 - i < 0$ households with initial skill history θ^{i-1} maximize $V_{1-i}(x; \theta^{i-1})$ subject to (2),

$$\begin{aligned} &h_{j,1-i+j}(\theta^j) \\ &\quad = (\theta^{i-1}, (w_{1-i+s}n_{s,1-i+s}(\theta^s), r_{1-i+s}(k_{s,1-i+s}(\theta^{s-1}) + b_{s,1-i+s}(\theta^{s-1})), m_{s,1-i+s}(\theta^s))_{s=i}^j), \end{aligned}$$

and (4) for each $j \geq i$ and θ^j, where $k_{i,1}(\theta^{i-1})$ and $b_{i,1}(\theta^{i-1})$ are given values which aggregate to K_1 and B_1, respectively.

4. Markets clear. That is, (1) and

$$(C_t, N_t, K_{t+1}, B_{t+1}) = \sum_{j=1}^J \sum_{\theta^j} (c_{jt}(\theta^j), n_{jt}(\theta^j), k_{j+1,t+1}(\theta^j), b_{j+1,t+1}(\theta^j))\mu_{jt}(\theta^j)$$

hold for each t.

5. The government's budget balances for each t:

$$G_t + (1+r_t)B_t = B_{t+1} + \sum_{j=1}^{J}\sum_{\theta^j}\tau_{jt}(h_{jt}(\theta^j))\mu_{jt}(\theta^j) + (1+r_t)\sum_{j=2}^{J}\sum_{\theta^j}(1-\psi_j)\left(k_{jt}(\theta^{j-1})+b_{jt}(\theta^{j-1})\right)\mu_{j-1,t-1}(\theta^{j-1}),$$

where the final term is revenue from bequest taxation.

Call $x = ((c_{jt}, n_{jt})_{j=1}^{J}, K_t)_{t=1}^{\infty}$ the equilibrium allocation. An equilibrium is stationary if its allocation is stationary.

3 Question and Approach

Let us now consider a class of optimal tax problems of the form:

$$\max_{T,x} W(x), \quad \text{subject to} \quad T \in \mathcal{T},\ x \in \mathcal{E}(T) \tag{5}$$

where $\mathcal{T} \subset \mathcal{T}^*$ is a set of tax instruments under consideration, $\mathcal{E}(T)$ is the set of equilibrium allocations under tax policy T, and W is a utilitarian social welfare function that places equal weight on all cohorts:

$$W(x) = \liminf_{H\to\infty} \frac{1}{H+J}\sum_{t=1-J}^{H} V_t(x). \tag{6}$$

In their analysis, CKK focus on a particular set $\mathcal{T}^{CKK} \subsetneq \mathcal{T}^*$ under which taxes depend only on current income as:

$$\tau_{jt}(h_{jt}) = \tau^n(w_t n_{jt};\varphi_t) + \tau^a r_t(k_{jt}+b_{jt}), \tag{7}$$

where $\tau^n(y;\varphi_t) \equiv \varphi_0(y-(y^{-\varphi_1}+\varphi_{2t})^{-1/\varphi_1})$ is the Gouveia and Strauss (1994) tax function. Each $T \in \mathcal{T}^{CKK}$ is therefore indexed by three parameters $(\varphi_0, \varphi_1, \tau^a)$, and φ_{2t} adjusts in each period so that the government's budget constraint holds. The level of per capita government debt is given and no messages are collected. They then solve for the optimal $T^{CKK} \in \mathcal{T}^{CKK}$, and find that the optimal τ^n is essentially a flat tax with a fixed deduction and that τ^a is significantly positive.[2]

[2]This description differs somewhat from CKK's, but the two are mathematically equivalent under a technical convergence assumption which I will assume throughout: For any $T \in \mathcal{T}^{CKK}$, there exists an

There are theoretical reasons to expect the performance of T^{CKK} to be less than ideal, however. A general point of course is that setting $\mathcal{T} = \mathcal{T}^{CKK}$ instead of $\mathcal{T} = \mathcal{T}^*$ in (5) imposes a restriction on the choice set and hence cannot be welfare-enhancing. But more specifically, several recent papers have studied the theoretical solution properties of (5) with $\mathcal{T} = \mathcal{T}^*$ and have concluded that an optimal tax system is necessarily: (i) non-separable in labor and asset income, and (ii) most likely history dependent as well when skills are serially dependent (Albanesi and Sleet, 2006, Golosov and Tsyvinski, 2006, Kocherlakota, 2005). Because none of the tax systems in $\mathcal{T}^{CKK}$ are allowed to have these properties, the loss from CKK's restrictions is strictly positive.

But the question stands: Is the loss from restricting attention to $\mathcal{T}^{CKK}$ small or large in a quantitative sense? If it is small, it would make sense to ignore the above concern for all practical purposes, given that adding complexity to the tax system will no doubt increase costs of administration and compliance (neither of which are explicitly modelled here). If it is large, however, it may make sense to give it due consideration.

To address this question, I perform the following computational experiment. I first solve for T^{CKK} and let the economy start in period $t = 1$ from the associated stationary equilibrium. Then I consider two policy scenarios. Under the first, the government keeps T^{CKK}. Under the second, the government switches to the optimal unrestricted tax system $T^* \in \mathcal{T}^*$. I ask how much better the latter scenario is compared according to W, and interpret it as an answer to the question above.

Of course, implementing this plan requires solving (5) with $\mathcal{T} = \mathcal{T}^*$—which I call the unrestricted optimal tax problem hereafter—and it is not possible to do so by conducting a direct numerical search over $\mathcal{T}^*$. My approach is therefore to simplify the problem analytically before resorting to numerical methods.

The first step in this simplification is to take a mechanism design approach to the problem following the NDPF, and it is useful to introduce the relevant terminology. Thus, let us say that an allocation $x = ((c_{jt}, n_{jt})_{j=1}^{J}, K_t)_{t=1}^{\infty}$ is feasible if it satisfies the following two conditions. The first condition is resource feasibility, which requires that (1) hold with

$$(C_t, N_t) = \sum_{j=1}^{J} \sum_{\theta^j} (c_{jt}(\theta^j), n_{jt}(\theta^j))\mu_{jt}(\theta^j).$$

The second condition is incentive compatibility for each household. An allocation is incentive

allocation that maximizes $W(x)$ subject to $x \in \mathcal{E}(T)$ and converges to a stationary allocation. Under this assumption, one can solve the optimal tax problem (5) under $\mathcal{T}^{CKK}$ by choosing a tax system in $\mathcal{T}^{CKK}$ so as to maximize the lifetime utility of a household who is born in the associated stationary equilibrium. CKK define their welfare criterion in terms of this procedure. A proof of this easily follows from Lemma 2 in appendix A. See Aiyagari and McGrattan (1998) for a closely related discussion.

compatible for a cohort $t \geq 0$ household if:

$$V_t(x) \geq \sum_{j=1}^{J} \sum_{\theta^j} \beta^{j-1} U(c_{j,t+j}(\sigma^j(\theta^j)), n_{j,t+j}(\sigma^j(\theta^j))/\theta_j)\pi_j(\theta^j) \tag{8}$$

for all reporting strategies $(\sigma_j)_{j=1}^J$, where $\sigma_j : \Theta^j \to \Theta_j$ and $\sigma^j = (\sigma_1, ..., \sigma_j)$. Analogously, an allocation is incentive compatible for a cohort $t = 1 - i < 0$ household with initial skill history θ^{i-1} if:

$$V_{1-i}(x; \theta^{i-1}) \geq \sum_{j=i}^{J} \sum_{\theta_i^j} \beta^{j-i} U(c_{j,1-i+j}(\theta^{i-1}, \sigma_i^j(\theta_i^j)), n_{j,1-i+j}(\theta^{i-1}, \sigma_i^j(\theta_i^j))/\theta_j)\pi_j(\theta_i^j|\theta_{i-1}), \tag{9}$$

for all reporting strategies $(\sigma_{i,j})_{j=i}^J$, where $\sigma_{i,j} : \Theta_i^j \to \Theta_j$, and $\sigma_i^j = (\sigma_{i,i}, ..., \sigma_{i,j})$. The planning problem is then to choose an allocation x so as to maximize social welfare W subject to feasibility.

Now because any tax-distorted market arrangement is a particular mechanism, it follows from the revelation principle that no such arrangement can do better than an optimal direct mechanism, namely a solution x^* to the planning problem. And because Kocherlakota's (2005) implementation result is readily adapted to this setup, we can conclude that x^* together with a tax system T^* constructed following his approach solves the unrestricted optimal tax problem.

The remaining task is then to compute x^*. Doing so directly is difficult however due to the model's general equilibrium structure, and a further simplification is necessary. Fortunately, it is possible to obtain one as follows. The starting point is to make the educated guess that the capital-labor ratio under x^* will satisfy the golden rule in the long run, which would pin down the long-run intertemporal shadow price. If so, this would enable us to characterize the long-run behavior of x^* as a solution to a collection of "partial equilibrium" problems that treat each household separately taking this price as given (Atkeson and Lucas, 1992). And because W effectively places all "weight" on the long run, this is plausibly all we need to know about x^*. This reasoning suggests the following result, whose formal proof is given in appendix A:

Proposition 1. *Let the capital-labor ratio κ^* satisfy the golden rule $\hat{r}(\kappa^*) = \eta$ and let the consumption-labor profile $(c_j^*, n_j^*)_{j=1}^J$ solve the dynamic mechanism design problem:*

$$\max_{(c_j, n_j)_{j=1}^J} \sum_{j=1}^{J} \sum_{\theta^j} \beta^{j-1} U(c_j(\theta^j), n_j(\theta^j)/\theta_j)\pi_j(\theta^j) \tag{10}$$

subject to

$$\sum_{j=1}^{J}\sum_{\theta^j}\left(\frac{1}{1+\hat{r}(\kappa^*)}\right)^{j-1}\left\{c_j(\theta^j)-\hat{w}(\kappa^*)n_j(\theta^j)\right\}\pi_j(\theta^j)+G\leq 0 \tag{11}$$

and

$$\sum_{j=1}^{J}\sum_{\theta^j}\beta^{j-1}\left\{U(c_j(\theta^j),n_j(\theta^j)/\theta_j)-U(c_j(\sigma^j(\theta^j)),n_j(\sigma^j(\theta^j))/\theta_j)\right\}\pi_j(\theta^j)\geq 0 \tag{12}$$

for all reporting strategies $(\sigma_j)_{j=1}^J$. Then any feasible allocation $x^*=((c^*_{jt},n^*_{jt})_{j=1}^J,K^*_t)_{t=1}^\infty$ such that $(c^*_{jt},n^*_{jt})_{j=1}^J\to(c^*_j,n^*_j)_{j=1}^J$ as $t\to\infty$ together with some tax system T^* solves the unrestricted optimal tax problem, and the maximum value of (10) is the welfare level after the reform to T^*.

The final challenge is to solve the dynamic mechanism design problem (10). Although this can be rather difficult when skills are serially dependent, it is possible to ameliorate this difficulty considerably using a recursive method due to Fukushima and Waki (2009), and that is what I will do.

4 Calibration

This section describes the functional forms and parameter values I use in the simulations. My overall approach is similar to CKK's: I first posit a tax policy that resembles the current U.S. system and then choose the parameters so that the associated stationary equilibrium is consistent with U.S. data along several dimensions. Unless indicated, the empirical targets are average values for years 1980-2007 computed from the data sources listed in appendix B. In the discussion I quote all numbers in annualized terms, and associate parameters with empirical targets in the usual heuristic fashion. Throughout, I identify l with hours worked and $w\theta$ with wages.

Demographics. A model period stands for 10 years, and households can live from ages 25 to 85. (Thus $J=6$, where $j=1$ stands for ages 25-35, $j=2$ for ages 35-45, and so on.) I set the population growth rate to its data counterpart $\eta=0.012$, and take the survival rates ψ_j from the U.S. life tables (Arias, Curtin, Wei, and Anderson, 2008).

Technology. The aggregate production function is Cobb-Douglas $F(K,N)=K^\alpha N^{1-\alpha}$ with capital share $\alpha=0.382$, and I set the depreciation rate $\delta=0.072$ so as to hit the 20.6% investment-output ratio in the data.

Preferences. Household utility takes the form:

$$U(c,l) = \frac{c^{1-\gamma}-1}{1-\gamma} - \phi \frac{l^{1+1/\epsilon}}{1+1/\epsilon}.$$

As a benchmark I use $\gamma = 1$ for the relative risk aversion coefficient and $\epsilon = 0.5$ for the Frisch labor supply elasticity. These are on the conservative side of values used in the literature. I also report results for $\gamma = 2$ and $\epsilon = 1$ because these values come closer to CKK's specification in terms of the implied elasticities. I choose the discount factor β to hit the capital-output ratio of 3.16 in the data, and set the share parameter ϕ so that hours $l = 0.33$ on average in the population.

Skill Process. The skill/wage process has the representation $\theta_j = e_j \exp(z_j)$, where $(e_j)_{j=1}^J$ is a deterministic age-dependent sequence taken from Hansen (1993) and $(z_j)_{j=1}^J$ follows a 5-state Markov chain. To specify the z_j process, I first define a parametric class of Markov chains indexed by three parameters $(\rho, \sigma_\nu^2, \sigma_{z_1}^2)$ as follows: (i) discretize the continuous state model

$$\begin{aligned} &z_j = \rho z_{j-1} + \nu_j, \quad \nu_j \sim N(0, \sigma_\nu^2), \quad j = 2, ..., J \\ &z_1 \sim N(0, \sigma_{z_1}^2) \end{aligned}$$

using Tauchen's (1986) method; and (ii) construct an approximation of the resulting process such that the transition probabilities have the representation:

$$\Pr(z_j | z_{j-1}) = p_{1j}(z_j)\omega_j(z_{j-1}) + p_{2j}(z_j)(1 - \omega_j(z_{j-1})), \tag{13}$$

where p_{1j} and p_{2j} are densities over z_j and $\omega_j(z_{j-1})$ is a weight between 0 and 1. Here, step (ii) follows Fukushima and Waki (2009), and the representation (13) makes it possible to solve the dynamic mechanism design problem (10) using their method. Then, I choose $(\rho, \sigma_\nu^2, \sigma_{z_1}^2)$ so that the implied Markov chain fits the age profile of cross-sectional log wage variance estimated by Heathcote, Storesletten, and Violante (2005) as close as possible.[3] (In view of Heathcote, Storesletten, and Violante's argument, I use the estimates they obtain controlling for time effects.) The resulting process is persistent—the annualized second largest eigenvalues of the transition matrices are above 0.92—and attains a close fit with the empirical targets (figure 1).

[3] The identification strategy here is essentially that of Storesletten, Telmer, and Yaron (2004): the profile's value at age 25 pins down $\sigma_{z_1}^2$, its slope pins down σ_ν^2, and its curvature pins down ρ.

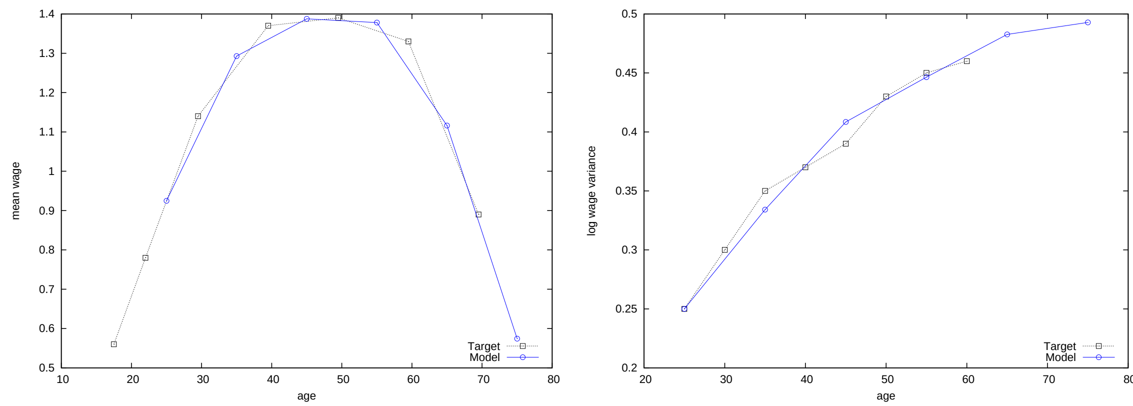

Figure 1: Skill process's fit with empirical targets.

Government Policy. The tax system has two components. The first is a social security system which imposes a linear tax on labor income and pays out a constant benefit to those above age 65. I set the payroll tax rate to 10.6% and choose the benefit level so that the GDP share of social security benefit payments is 3.5%, both as in the data. The second component is a progressive federal income tax which levies $\varphi_0(y - (y^{-\varphi_1} + \varphi_2)^{-1/\varphi_1})$ as a function of current taxable income y, defined as labor income plus asset income less one half of social security tax payments. Here, I take the values $(\varphi_0, \varphi_1) = (0.258, 0.768)$ from Gouveia and Strauss (1994) and let φ_2 adjust so that the government's budget constraint holds. I assume $B_t = (1+\eta)^{t-1}B$ and choose G and B so that the GDP shares of government expenditures and government debt hit the data values 17.8% and 50.1% respectively.

5 Results

5.1 Welfare Gains

I now simulate the policy reform in question and quantify its impact on welfare. The status quo is of course the stationary equilibrium under the optimal $T^{CKK} \in \mathcal{T}^{CKK}$, but I depart from CKK's original analysis by choosing the level of government debt optimally as well. By abstracting from the standard long-run effects of government debt on capital accumulation (Diamond, 1965), this procedure places a lower bound on the welfare gain of interest. The status quo policy under the benchmark calibration consists of an 20% flat tax on labor income with a deduction of about 0.6 times median income per household, zero taxes on asset income, and sizable government asset holdings (negative debt) which account for about 83% of the capital stock.

	Welfare			Per capita aggregates					Transitions	
(γ, ϵ)	W	W_a	W_d	C	L	N	K	Y	NTC	$\hat{K}_1/\bar{K}_1$
(1.0, 0.5)	15.6%	13.4%	2.2%	10.5%	-3.4%	7.5%	3.2%	5.8%	2	0.55
(2.0, 0.5)	17.9%	14.9%	3.0%	7.7%	-7.8%	5.6%	2.2%	4.3%	2	0.56
(1.0, 1.0)	18.9%	20.6%	-1.7%	14.3%	-6.4%	11.2%	9.9%	10.7%	2	0.64
(2.0, 1.0)	18.9%	18.4%	0.5%	6.2%	-12.7%	4.7%	4.7%	2.9%	2	0.57

Table 1: Impact of the tax reform.

Table 1 summarizes the impact of the policy reform. Column W reports the welfare gain in terms of the percentage increase in consumption for all households at all dates and contingencies needed to generate an equivalent welfare increase (keeping labor supply constant). The numbers, which are all well above 10%, are large by conventional standards.

To highlight the source of this gain, columns C through Y report the long-run percentage changes in per capita aggregates. For each case there is a large increase in consumption (column C) and a moderate decline in hours (column L). Column W_a reports the welfare gain that is attributable to these two effects at the aggregate level, namely the gain that would obtain if households in the status quo were to have their consumption and hours shifted by these amounts at all dates and contingencies. As we can see, this accounts for most of the total gain; the contribution of improved insurance/redistribution, as measured by the residual $W_d \equiv W - W_a$, is small and possibly negative. Distributional effects are critical for physically supporting these improvements in per capita aggregates, however. Indeed, column N shows that effective labor input per capita increases significantly after the reform, and this is compatible with the decline in per capita hours only because of an effective increase in aggregate productivity that comes from a massive shift of labor supply toward productive households. Column K shows that the indirect effect of this through capital accumulation is significant as well.

5.2 Transitions

Because the policy reform induces significant capital accumulation, there is a transition phase during which heavy investment takes place and capital accumulates at a rapid rate. The welfare analysis above did not take this into account, however.

From a formal, mathematical point of view there is no problem with this: using a balanced growth path comparison for welfare calculations is justified by Proposition 1. But if we think through the economics behind this result, we can see that its validity depends on a peculiar (and in fact mathematically non-generic) property of of the social welfare function W, namely that it places zero Pareto weight on any finite number of cohorts. This makes the transition phase irrelevant for welfare and the "optimal transition path" indeterminate. Thus, there are infinitely many transition paths that attain the same welfare gain, some of which treat cohorts born at early dates better than others.

Given this, it would seem useful to ask if there is a transition path that treats all cohorts in a respectable fashion, say without making any of them worse off than they were under the status quo, and if so, how long it will take. In the following I address these questions by directly constructing a such a path.

My starting point is an allocation $\tilde{x}$ under which cohorts born before the reform are given the status quo consumption-labor profile $(\bar{c}_j, \bar{n}_j)_{j=1}^J$, all newborns are given the profile $(c_j^*, n_j^*)_{j=1}^J$ from Proposition 1, and the capital stock sequence equals that under the post-reform balanced growth path, $(K_t^*)_{t=1}^\infty$. This allocation satisfies all of the desired condition except for resource feasibility—the initial capital stock $\bar{K}_1$ is insufficient to support it (i.e., $\bar{K}_1 < K_1^*$). But because $\tilde{x}$ makes those cohorts born over the first several periods strictly better off than they were under the status quo, it is possible to convert some of their consumption into investment while securing their pre-reform welfare. So a way to proceed is to check if doing so will suffice to make up for the shortage of initial capital.

To this end, I construct a new allocation $\hat{x}$ by perturbing $\tilde{x}$ as follows. First fix $H(\geq J)$ which indexes the length of the transition, and choose $((\Delta_{jt})_{j=1}^J, K_t)_{t=1}^H$ so as to minimize K_1 subject to the constraints:

$$\sum_{j=1}^{J} \sum_{\theta^j} c_{jt}^{\Delta}(\theta^j)\mu_{jt}(\theta^j) + K_{t+1} - (1-\delta)K_t + G_t = F(K_t, \tilde{N}_t), \quad \forall t = 1, ..., H \tag{14}$$

$$c_{jt}^{\Delta}(\theta^j) = \begin{cases} u^{-1}\left(u(c_j^*(\theta^j)) - \Delta_{jt}\right) & \text{if } 0 \leq t-j \leq H-J \\ c_j^*(\theta^j) & \text{if } t-j > H-J \\ \bar{c}_j(\theta^j) & \text{if } t-j < 0 \end{cases} \tag{15}$$

$$\sum_{j=1}^{J} \beta^{j-1} \Delta_{j,t+j} \left(\prod_{i=1}^{j} \psi_i \right) \leq W^* - \bar{W}, \quad \forall t = 0, ..., H-J \tag{16}$$

where $K_{H+1} = K_{H+1}^*$, $(\tilde{N}_t)_{t=1}^H$ is the effective labor sequence under $\tilde{x}$ and W^* $(\bar{W})$ is the post-reform (pre-reform) welfare level. Let $((\hat{\Delta}_{jt})_{j=1}^J, \hat{K}_t)_{t=1}^H$ denote a solution to this problem. Then define $\hat{x}$ by taking $\tilde{x}$ and replacing the consumption for cohorts $0, ..., H-J$ by $\hat{c}_{jt} = u^{-1}(u(c_j^*(\theta^j)) - \hat{\Delta}_{jt})$ and the capital stock for periods $1, ..., H$ by $(\hat{K}_t)_{t=1}^H$.

In words, this perturbation designates cohorts $t = 0, ..., H-J$ as the "heavy investors," whose consumption is reduced relative to $(c_j^*)_{j=1}^J$ for the sake of investment. The consumption reduction takes the form (15) so as to preserve incentive compatibility (Rogerson, 1985), while the constraint (16) insures that none of these cohorts are made worse off than under the status quo. Hence $\hat{x}$ satisfies all of the desired conditions as long as $\hat{K}_1 \leq \bar{K}_1$.

Given this, I compute the minimum H for which $\hat{K}_1 \leq \bar{K}_1$, and report the results in the final part of table 1. As we can see, a desired transition indeed exists for all cases, and it takes $NTC \equiv H - J + 1 = 2$ model cohorts—cohorts born over a span of 20 years—to accomplish the required investment in capital. The low values of $\hat{K}_1/\bar{K}_1$ imply that it is possible to further Pareto improve upon $\hat{x}$ by distributing a significant fraction of the initial capital stock in an arbitrary fashion.

6 Conclusion

This paper showed that in the model economy it considers, there are large gains from introducing the right kinds of non-linearities and age/history dependence into the income tax code. Doing so enlarges the social pie by motivating talented people to work harder.

In future work, it would be useful to revisit this assessment using models that feature richer and more realistic formulations of the labor market. It is not obvious how this would affect the results obtained here. On the one hand, allowing for long-term labor contracts would at least partially obviate the need for government-provided social insurance (cf. Golosov and Tsyvinski, 2007), and this may lead to a downward revision of the gains. On the other hand, accounting for the extensive margin of labor supply in a reasonable fashion may increase the sensitivity of labor supply to tax distortions (cf. Chang and Kim, 2006; Rogerson and Wallenius, 2007), and thereby lead to an upward revision of the gains. Much remains to be sorted out on this front.

A Proof of Proposition 1

We first observe the following property of W:

Lemma 2. *If $V_t(x) \to V_\infty$ as $t \to \infty$, $W(x) = V_\infty$.*

Proof. Write:

$$\frac{1}{H^2+J}\sum_{t=1-J}^{H^2} V_t(x) = \left(\frac{H+J}{H^2+J}\right)\frac{1}{H+J}\sum_{t=1-J}^{H} V_t(x) + \left(\frac{H^2-H}{H^2+J}\right)\frac{1}{H^2-H}\sum_{t=H+1}^{H^2} V_t(x).$$

As $H \to \infty$, the first term on the right hand side converges to zero, while the second term converges to V_∞. □

To proceed, let us reformulate the planning problem recursively following Fernandes and Phelan (2000) by introducing a new variable υ representing continuation utilities. Formally, a continuation utility as of age j given $(c_i, n_i)_{i=j}^J$, where $(c_i, n_i) : \Theta^i \to \mathbb{R}_+ \times [0, n_{\max}]$ for each i, is $\upsilon_j : \Theta^{j-1} \to \mathbb{R}^{\Theta_{j-1}}$ such that

$$\upsilon_j(\theta^{j-1})(\theta'_{j-1}) = \sum_{i=j}^{J} \sum_{\theta_j^i} \beta^{i-j} U(c_i(\theta^i), n_i(\theta^i)/\theta_i) \pi_i(\theta_j^i | \theta'_{j-1})$$

for all $(\theta^{j-1}, \theta'_{j-1})$, where $\Theta_0 = \Theta^0 \equiv \emptyset$. This defines a mapping $\Upsilon_j : (c_i, n_i)_{i=j}^J \mapsto \upsilon_j$. Also define a sequence of functions $(D_j^I, D_j^P)_{j=1}^J$ by:

$$\begin{aligned} D_j^I(c_j, n_j, \upsilon_{j+1}; \theta^j, \theta'_j) = {} & U(c_j(\theta^j), n_j(\theta^j)/\theta_j) + \beta \upsilon_{j+1}(\theta^j)(\theta_j) \\ & - U(c_j(\theta^{j-1}, \theta'_j), n_j(\theta^{j-1}, \theta'_j)/\theta_j) - \beta \upsilon_{j+1}(\theta^{j-1}, \theta'_j)(\theta_j) \end{aligned}$$

and

$$\begin{aligned} D_j^P(c_j, n_j, \upsilon_j, \upsilon_{j+1}; \theta^{j-1}, \theta'_{j-1}) = {} & \upsilon_j(\theta^{j-1})(\theta'_{j-1}) \\ & - \sum_{\theta_j} \left\{ U(c_j(\theta^j), n_j(\theta^j)/\theta_j) + \beta \upsilon_{j+1}(\theta^j)(\theta_j) \right\} \pi_j(\theta_j | \theta'_{j-1}) \end{aligned}$$

for all $(j, \theta^j, \theta'_j, \theta'_{j-1})$, $(c_j, n_j, \upsilon_{j+1}) : \Theta^j \to \mathbb{R}_+ \times [0, n_{\max}] \times \mathbb{R}^{\Theta_j}$, $\upsilon_j : \Theta^{j-1} \to \mathbb{R}^{\Theta_{j-1}}$. Finally, let $\upsilon_{J+1} \equiv 0$ and $\upsilon_{J+1,t} \equiv 0$ for all t in what follows. (Note that there is no need to characterize the subset of $\mathbb{R}^{\Theta_{j-1}}$ to which each $\upsilon_j(\theta^{j-1})$ must belong, given these terminal conditions and the fact that we will not be doing any backward induction in this proof.)

For a given initial condition $(\bar{K}_1, (\bar{\upsilon}_{j1})_{j=2}^J)$, where $\bar{K}_1 \in \mathbb{R}_+$ and $\bar{\upsilon}_{j1} : \Theta^{j-1} \to \mathbb{R}^{\Theta_{j-1}}$ for each j, define the auxiliary planning problem as follows: Choose $\xi = (x, ((\upsilon_{jt})_{j=1}^J)_{t=1}^\infty)$, where x is an allocation and $\upsilon_{jt} : \Theta^{j-1} \to \mathbb{R}^{\Theta_{j-1}}$ for each (t, j), to maximize $W(x)$ subject to the resource feasibility of x,

$$D_j^I(c_{jt}, n_{jt}, \upsilon_{j+1,t+1}; \theta^j, \theta'_j) \geq 0 \tag{17}$$

$$D_j^P(c_{jt}, n_{jt}, \upsilon_{jt}, \upsilon_{j+1,t+1}; \theta^{j-1}, \theta'_{j-1}) = 0 \tag{18}$$

for all $(t, j, \theta^j, \theta'_j, \theta'_{j-1})$, and the initial conditions $(K_1, (\upsilon_{j1})_{j=2}^J) = (\bar{K}_1, (\bar{\upsilon}_{j1})_{j=2}^J)$. Using (18), it is straightforward to see that $W(x) = \liminf_{H \to \infty} \frac{1}{H} \sum_{t=1}^H \upsilon_{1t}$ for any ξ satisfying the constraints. As well, because each n_{jt} is bounded and the resource constraint must hold at each t, we may without loss restrict each c_{jt}, υ_{jt}, and $K_t/(1+\eta)^{t-1}$ to be bounded from above and below by appropriate constants. Let $W^{APP*}(\bar{K}_1, (\bar{\upsilon}_{j1})_{j=2}^J)$ denote the maximum objective value of this problem as a function of its initial condition. ($W^{APP*}(\bar{K}_1, (\bar{\upsilon}_{j1})_{j=2}^J) \equiv -\infty$ if the constraint set given $(\bar{K}_1, (\bar{\upsilon}_{j1})_{j=2}^J)$ is empty.)

The following lemma clarifies the relationship between the auxiliary planning problem and the planning problem.

Lemma 3. *If, for a given $\bar{K}_1$,*

$$(\bar{\upsilon}_{j1})_{j=2}^J \in \arg \max_{(\upsilon_{j1})_{j=2}^J} W^{APP*}(\bar{K}_1, (\upsilon_{j1})_{j=2}^J), \tag{19}$$

the x-component of a solution to the auxiliary planning problem starting from $(\bar{K}_1, (\bar{\upsilon}_{j1})_{j=2}^J)$ solves the planning problem starting from $\bar{K}_1$.

Proof. If x^* satisfies the given description, it is resource feasible by definition, and is incentive compatible by (17), (18), and the one-shot deviation principle. To see that it is optimal, choose any feasible $x = ((c_{jt}, n_{jt})_{j=1}^J, K_t)_{t=1}^\infty$ and define $((\upsilon_{jt})_{j=1}^J)_{t=1}^\infty$ by $\upsilon_{j,t+j-1} = \Upsilon_j((c_{i,t+i-1}, n_{i,t+i-1})_{i=j}^J)$ for each j and t. Then $\xi = (x, ((\upsilon_{jt})_{j=1}^J)_{t=1}^\infty)$ satisfies the constraints of the auxiliary planning problem starting from $(\bar{K}_1, (\upsilon_{j1})_{j=2}^J)$, so

$$W(x) \leq W^{APP*}(\bar{K}_1, (\upsilon_{j1})_{j=2}^J) \leq W^{APP*}(\bar{K}_1, (\bar{\upsilon}_{j1})_{j=2}^J) = W(x^*)$$

as desired. □

Let us call $((c_j, n_j, \upsilon_j)_{j=1}^J, K)$ a stationary solution to the auxiliary planning problem if $\xi = (((c_{jt}, n_{jt}, \upsilon_{jt}) = (c_j, n_j, \upsilon_j))_{j=1}^J, K_t = (1+\eta)^{t-1}K)_{t=1}^\infty$ solves the auxiliary planning problem starting from $(\bar{K}_1 = K, (\bar{\upsilon}_{j1} = \upsilon_j)_{j=2}^J)$.

Lemma 4. *Let $((c_j^*, n_j^*)_{j=1}^J, \kappa^*)$ satisfy the conditions in Proposition 1, $\upsilon_j^* = \Upsilon_j((c_i^*, n_i^*)_{i=j}^J)$ for each j, and*

$$K^* = \kappa^* \sum_{j=1}^J \sum_{\theta^j} \left(\frac{1}{1+\eta}\right)^{j-1} n_j^*(\theta^j)\pi_j(\theta^j). \tag{20}$$

Then $((c_j^*, n_j^*, \upsilon_j^*)_{j=1}^J, K^*)$ is a stationary solution to the auxiliary planning problem.

Proof. Define $\xi^* = (((c_{jt}^*, n_{jt}^*, \upsilon_{jt}^*) = (c_j^*, n_j^*, \upsilon_j^*))_{j=1}^J, K_t^* = (1+\eta)^{t-1}K^*)_{t=1}^\infty$. This satisfies resource feasibility by (11), $\hat{r}(\kappa^*) = \eta$, (20), and Euler's theorem. It also satisfies (17) and (18) by (12) and the definition of $(\upsilon_j^*)_{j=1}^J$.

To verify its optimality, let us first follow Fernandes and Phelan (2000) and rewrite the dynamic mechanism design problem in the proposition as: Choose $(c_j, n_j, \upsilon_j)_{j=1}^J$, where

$\upsilon_j : \Theta^{j-1} \to \mathbb{R}^{\Theta_{j-1}}$ for each j, to maximize υ_1 subject to (11) and

$$D_j^I(c_j, n_j, \upsilon_{j+1}; \theta^j, \theta_j') \geq 0 \tag{21}$$

$$D_j^P(c_j, n_j, \upsilon_j, \upsilon_{j+1}; \theta^{j-1}, \theta_{j-1}') = 0 \tag{22}$$

for all $(j, \theta^j, \theta_j', \theta_{j-1}')$. Under the change of variables with $(u(c_j), v(n_j), \upsilon_j)_{j=1}^J$ instead of $(c_j, n_j, \upsilon_j)_{j=1}^J$ as the choice variable, this problem is smooth and concave. Moreover, once $(\upsilon_j)_{j=1}^J$ is substituted out as a linear function of $(u(c_j), v(n_j))_{j=1}^J$ using $(\Upsilon_j)_{j=1}^J$, the constraint (22) drops out and the constraint set has a non-empty interior. Hence there exist Lagrange multipliers $(\lambda^R, (\lambda_j^I, \lambda_j^P)_{j=1}^J)$ such that $(c_j^*, n_j^*, \upsilon_j^*)_{j=1}^J$ maximizes the Lagrangian:

$$L^{MDP}((c_j, n_j, \upsilon_j)_{j=1}^J) = \upsilon_1 - \lambda^R G + \sum_{j=1}^J \sum_{\theta^j} \left(\frac{\lambda^R}{(1+\eta)^{j-1}} \{\hat{w}(\kappa^*) n_j(\theta^j) - c_j(\theta^j)\} \right.$$
$$\left. + \sum_{\theta_j'} \lambda_j^I(\theta^j, \theta_j') D_j^I(c_j, n_j, \upsilon_{j+1}; \theta^j, \theta_j') + \sum_{\theta_{j-1}'} \lambda_j^P(\theta^{j-1}, \theta_{j-1}') D_j^P(c_j, n_j, \upsilon_j, \upsilon_{j+1}; \theta^{j-1}, \theta_{j-1}') \right) \pi_j(\theta^j)$$

and the complementary slackness conditions hold.

Consider the following Lagrangian for the auxiliary planning problem:

$$L^{APP}(\xi) = \liminf_{H\to\infty} \frac{1}{H} \sum_{t=1}^H \left\{ \upsilon_{1t} + \frac{\lambda^R}{(1+\eta)^{t-1}} \{F(K_t, N_t) - C_t - K_{t+1} + (1-\delta)K_t - G_t\} \right.$$
$$+ \sum_{j=1}^J \sum_{\theta^j} \left(\sum_{\theta_j'} \lambda_j^I(\theta^j, \theta_j') D_j^I(c_{jt}, n_{jt}, \upsilon_{j+1,t+1}; \theta^j, \theta_j') \right.$$
$$\left. \left. + \sum_{\theta_{j-1}'} \lambda_j^P(\theta^{j-1}, \theta_{j-1}') D_j^P(c_{jt}, n_{jt}, \upsilon_{jt}, \upsilon_{j+1,t+1}; \theta^{j-1}, \theta_{j-1}') \right) \pi_j(\theta^j) \right\}.$$

Using $F(K_t, N_t) \leq (\hat{r}(\kappa^*) + \delta)K_t + \hat{w}(\kappa^*)N_t$, $\hat{r}(\kappa^*) = \eta$, and the boundedness condition on ξ, we obtain

$$L^{APP}(\xi) \leq \liminf_{H\to\infty} \frac{1}{H} \sum_{t=1}^H L^{MDP}((c_{j,t+j-1}, n_{j,t+j-1}, \upsilon_{j,t+j-1})_{j=1}^J).$$

It then follows from the previous paragraph that L^{APP} is maximized at ξ^* and that the complementary slackness conditions hold.

Now suppose ξ^* did not solve the auxiliary planning problem, and let ξ^{**} denote a superior choice. Then using the constraints and the complementary slackness conditions, we have

$$L^{APP}(\xi^{**}) \geq W(x^{**}) > W(x^*) = L^{APP}(\xi^*),$$

where x^* and x^{**} are the x-components of ξ^* and ξ^{**}, respectively. This contradicts the above. $\square$

Lemma 5. *W^{APP*} is a constant function.*

Proof. Pick any two initial conditions $(\bar{K}_1, (\bar{\upsilon}_{j1})_{j=2}^J)$ and $(\bar{K}_1', (\bar{\upsilon}_{j1}')_{j=2}^J)$, and let ξ and ξ' solve the corresponding auxiliary planning problems. Then consider a deviation from ξ of the following form. For the first H periods set the consumption-labor profiles for all newborns to $(c_j = 0, n_j = n_{\max})_{j=1}^J$. From then on, set them to what they are under ξ'. For H sufficiently large, this together with a capital stock sequence which equals that under ξ' for $t \geq H+1$ defines a feasible allocation. Since this deviation equals ξ' after a finite number of periods, the no-discounting property of W implies that it gives welfare $W^{APP*}(\bar{K}_1', (\bar{\upsilon}_{j1}')_{j=2}^J)$. It follows that $W^{APP*}(\bar{K}_1, (\bar{\upsilon}_{j1})_{j=2}^J) \geq W^{APP*}(\bar{K}_1', (\bar{\upsilon}_{j1}')_{j=2}^J)$. Use symmetry. $\square$

Lemma 6. *If x^* satisfies the conditions in Proposition 1, it solves the planning problem.*

Proof. Let $((c_j^*, n_j^*)_{j=1}^J, \kappa^*)$ and x^* satisfy the conditions in Proposition 1. Define $(\upsilon_j^*)_{j=1}^J$ and K^* as in Lemma 4. Let $W^{PP*}(\bar{K}_1)$ denote the maximum value of the objective in the planning problem. We then have:

$$\begin{aligned}
W(x^*) &= \upsilon_1^* \quad (\text{by Lemma 2, since } (c_{jt}^*, n_{jt}^*) \to (c_j^*, n_j^*) \text{ and so } V_t(x^*) \to \upsilon_1^* \text{ as } t \to \infty) \\
&= W^{APP*}(K^*, (\upsilon_j^*)_{j=2}^J) \quad (\text{by Lemma 4}) \\
&= W^{APP*}(\bar{K}_1, (\bar{\upsilon}_{j1})_{j=2}^J) \quad (\text{by Lemma 5, where } (\bar{\upsilon}_{j1})_{j=2}^J \text{ satisfies (19)}) \\
&= W^{PP*}(\bar{K}_1) \quad (\text{by Lemma 3})
\end{aligned}$$

Hence x^* solves the planning problem. $\square$

The following lemma, which is a straightforward adaptation of Kocherlakota (2005), concludes the proof:

Lemma 7. *If x^* solves the planning problem, there exists a tax system T^* such that (T^*, x^*) solves (5) with $\mathcal{T} = \mathcal{T}^*$.*

Proof. We first construct a tax policy T^* and a candidate equilibrium as follows. Write $x^* = ((c_{jt}^*, n_{jt}^*)_{j=1}^J, K_t^*)_{t=1}^\infty$. For each t, define C_t^* and N_t^* by aggregating $(c_{jt}^*, n_{jt}^*)_{j=1}^J$ and set factor prices to $r_t^* = F_K(K_t^*, N_t^*) - \delta$ and $w_t^* = F_N(K_t^*, N_t^*)$. Let $M_{jt}^* = \Theta_j$ and $m_{jt}^*(\theta^j) = \theta_j$ for each (t, j, θ^j). Let each τ_{jt}^* take the form:

$$\tau_{jt}^*(h_{jt}) = \tau_{jt}^{n*}(\theta^j, w_t n_{jt}) + \tau_{jt}^{a*}(\theta^j, w_t n_{jt}) r_t (k_{jt} + b_{jt}),$$

and specify $(\tau_{jt}^{n*}, \tau_{jt}^{a*})$ as follows. First let $((\tau_{jt}^a)_{j=1}^J)_{t=1}^\infty$ satisfy:

$$u'(c_{j,t+j}^*(\theta^j)) = \beta u'(c_{j+1,t+j+1}^*(\theta^{j+1}))[1 + (1 - \tau_{j+1,t+j+1}^a(\theta^{j+1}))r_{t+j+1}^*] \tag{23}$$

for all (t, j, θ^{j+1}), and choose $((\tau_{jt}^n, k_{jt}^*, b_{jt}^*)_{j=1}^J, B_t^*)_{t=1}^\infty$ so as to satisfy the budget constraints

$$\begin{aligned} &c_{j,t+j}^*(\theta^j) + k_{j+1,t+j+1}^*(\theta^j) + b_{j+1,t+j+1}^*(\theta^j) \\ &\quad = w_{t+j}^* n_{j,t+j}^*(\theta^j) + [1 + (1 - \tau_{j,t+j}^a(\theta^j))r_{t+j}^*](k_{j,t+j}^*(\theta^{j-1}) + b_{j,t+j}^*(\theta^{j-1})) - \tau_{j,t+j}^n(\theta^j), \end{aligned} \tag{24}$$

for all (t, j, θ^j), the initial conditions on asset holdings, and the aggregation conditions

$$(K_{t+1}^*, B_{t+1}^*) = \sum_{j=1}^J \sum_{\theta^j} (k_{j+1,t+1}^*(\theta^j), b_{j+1,t+1}^*(\theta^j))\mu_{jt}(\theta^j)$$

for all t. Then, set

$$(\tau_{jt}^{n*}(\theta^j, w_t n_{jt}), \tau_{jt}^{a*}(\theta^j, w_t n_{jt})) = \begin{cases} (\tau_{jt}^n(\theta^j), \tau_{jt}^a(\theta^j)) & \text{if } w_t n_{jt} = w_t^* n_{jt}^*(\theta^j) \\ (w_t n_{jt} + 1, 1/r_t^* + 1) & \text{otherwise} \end{cases}$$

for each $(t, j, \theta^j, w_t n_{jt})$.

I claim that (T^*, x^*) solves the optimal tax problem (5) under $\mathcal{T}^*$. Since any equilibrium allocation is feasible, it is enough to show that $((c_{jt}^*, n_{jt}^*, k_{jt}^*, b_{jt}^*, m_{jt}^*)_{j=1}^J)_{t=1}^\infty$, $(C_t^*, N_t^*, K_t^*)_{t=1}^\infty$, and $(w_t^*, r_t^*)_{t=1}^\infty$ is an equilibrium given T^*. Markets clear and the marginal product conditions hold by construction, so it remains to check that households are optimizing. (The government's budget constraint is then implied by Walras' law). The argument for cohorts $t \geq 0$ is the following. If a household chooses $(m_{j,t+j})_{j=1}^J$, its labor choice must satisfy $n_{j,t+j}(\theta^j) = n_{j,t+j}^*((m_{i,t+i}(\theta^i))_{i=1}^j)$ for all (j, θ^j) so as to be budget feasible. Given this, it follows from (23) and (24) that choosing $c_{j,t+j}(\theta^j) = c_{j,t+j}^*((m_{i,t+i}(\theta^i))_{i=1}^j)$ and $k_{j+1,t+j+1}(\theta^j) = k_{j+1,t+j+1}^*((m_{i,t+i}(\theta^i))_{i=1}^j)$ for all (j, θ^j) is optimal. The conclusion then follows from the incentive compatibility of x^*. The argument for cohorts $t < 0$ is the same. $\square$

B Data Appendix

Data for aggregate and policy variables are from the Bureau of Economic Analysis's National Income and Product Accounts (NIPA) and Fixed Asset Tables (FA), the Federal Reserve Board's Flow of Funds Accounts (FOF), the Economic Report of the President (EROP), and the Social Security Administration's Annual Statistical Supplement to the Social Security Bulletin (SSA).

The mapping between model and data variables is straightforward for the following: the population growth rate is that of the civilian non-institutional population of ages 16 and above (EROP B-35); government debt is gross federal debt (EROP B-78); the social security tax rate is the sum of Old Age and Survivors Insurance (OASI) contribution rates for employers and employees (SSA 2.A3); and social security benefit expenses are those for the OASI (SSA 4.A1).

For the remaining variables, the mapping generally follows Cooley and Prescott (1995): capital is the total value of private fixed assets (FA 1.1), consumer durables (FA 1.1), inventories (NIPA 5.7.5.A/B), and land (FOF B.100, B.102, B.103); the components of gross domestic income (NIPA 1.10) are allocated to capital and labor income assuming that factor shares among the ambiguous components (components other than compensation of employees, net interest, rental income, and corporate profits) are the same as those among total income; service flows from consumer durables are imputed assuming that they yield the same rate of return as other components of capital; and gross domestic product/income and its components (NIPA 1.1.5 and 1.10) are adjusted by adding the imputed service flows from durables to consumption and capital income.

References

Aiyagari, S. R., and E. R. McGrattan (1998): "The Optimum Quantity of Debt," *Journal of Monetary Economics*, 42(3), 447–469.

Albanesi, S., and C. Sleet (2006): "Dynamic Optimal Taxation with Private Information," *Review of Economic Studies*, 73(1), 1–30.

Arias, E., L. R. Curtin, R. Wei, and R. N. Anderson (2008): "U.S. Decennial Life Tables for 1999-2001, United States Life Tables," *National Vital Statistics Reports*, 57(1), 1–40.

Atkeson, A., and R. E. Lucas, Jr. (1992): "On Efficient Distribution with Private Information," *Review of Economic Studies*, 59(3), 427–453.

Chang, Y., and S.-B. Kim (2006): “From Individual to Aggregate Labor Supply: A Quantitative Analysis Based on a Heterogeneous Agent Macroeconomy,” *International Economic Review*, 47(1), 1–27.

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, J. C., and D. Krueger (2006): “On the Optimal Progressivity of the Income Tax Code,” *Journal of Monetary Economics*, 53(7), 1425–1450.

Cooley, T. F., and E. C. Prescott (1995): “Economic Growth and Business Cycles,” in *Frontiers of Bysiness Cycle Research*, ed. by T. F. Cooley, pp. 1–38. Princeton University Press.

Diamond, P. A. (1965): “National Debt in a Neoclassical Growth Model,” *American Economic Review*, 55(5), 1126–1150.

Farhi, E., and I. Werning (2009): “Capital Taxation: Quantitative Explorations of the Inverse Euler Equation,” Discussion paper, Harvard and MIT.

Fernandes, A., and C. Phelan (2000): “A Recursive Formulation for Repeated Agency with History Dependence,” *Journal of Economic Theory*, 91(2), 223–247.

Fukushima, K., and Y. Waki (2009): “Computing Dynamic Optimal Mechanisms When Hidden Types Are Markov,” Working paper, University of Minnesota and FRB Minneapolis.

Golosov, M., and A. Tsyvinski (2006): “Designing Optimal Disability Insurance: A Case for Asset Testing,” *Journal of Political Economy*, 114(2), 257–279.

——— (2007): “Optimal Taxation With Endogenous Insurance Markets,” *Quarterly Journal of Economics*, 122(2), 487–534.

Gouveia, M., and R. P. Strauss (1994): “Effective Federal Individual Income Tax Functions: An Exploratory Empirical Analysis,” *National Tax Journal*, 47(2), 317–339.

Hall, R. E., and A. Rabushka (1995): *The Flat Tax*. Hoover Institution Press, second edn.

Hansen, G. D. (1993): “The Cyclical and Secular Behaviour of the Labour Input: Comparing Efficiency Units and Hours Worked,” *Journal of Applied Econometrics*, 8(1), 71–80.

Heathcote, J., K. Storesletten, and G. L. Violante (2005): “Two Views of Inequality over the Life Cycle,” *Journal of the European Economic Association*, 3(2/3), 765–775.

Huggett, M., and J. C. Parra (2009): "How Well Does the US Social Insurance System Provide Social Insurance?," Working paper, Georgetown University.

Kocherlakota, N. R. (2005): "Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation," *Econometrica*, 73(5), 1587–1621.

——— (2009): *The New Dynamic Public Finance*. Princeton University Press, forthcoming.

Mirrlees, J. A. (1971): "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 38(2), 175–208.

Rogerson, R., and J. Wallenius (2007): "Micro and Macro Elasticities in a Life Cycle Model with Taxes," NBER Working Paper 13017.

Rogerson, W. P. (1985): "Repeated Moral Hazard," *Econometrica*, 53(1), 69–76.

Storesletten, K., C. I. Telmer, and A. Yaron (2004): "Consumption and Risk Sharing Over the Life Cycle," *Journal of Monetary Economics*, 51(3), 609–633.

Tauchen, G. (1986): "Finite State Markov-Chain Approximations to Univariate and Vector Autoregressions," *Economics Letters*, 20(2), 177–181.

Weinzierl, M. (2008): "The Surprising Power of Age-Dependent Taxes," Discussion paper, Harvard University.