Optimal Progressive Taxation and Education Subsidies in a Model of Endogenous Human Capital Formation*

Dirk Krueger
University of Pennsylvania, CEPR and NBER

Alexander Ludwig
CMR at University of Cologne, MEA and Netspar

December 21, 2012

Abstract

In this paper we characterize quantitatively the optimal mix of progressive income taxes and education subsidies in a model with endogenous human capital formation, borrowing constraints, income risk. and incomplete financial markets. Progressive labor income taxes provide social insurance against idiosyncratic income risk and redistributes after tax income among ex-ante heterogeneous households. In addition to the standard distortions of labor supply progressive taxes also impede the incentives to acquire higher education, generating a non-trivial trade-off for the benevolent utilitarian government. The latter distortion can potentially be mitigated by an education subsidy. We find that the welfare-maximizing fiscal policy is indeed characterized by a substantially progressive labor income tax code and a positive subsidy for college education. Both the degree of tax progressivity and the education subsidy are larger than in the current U.S. status quo.

Keywords: Progressive Taxation, Capital Taxation, Optimal Taxation

J.E.L. classification codes: E62, H21, H24

*We thank Felicia Ionescu for helpful discussions and seminar participants at Boconi and the LSE for many useful comments. Krueger gratefully acknowledges financial support from the NSF under grant SES-0820494.

1 Introduction

In this paper we characterize quantitatively the optimal mix of progressive income taxes and education subsidies in a large-scale overlapping generations model with endogenous human capital formation, borrowing constraints, income risk, intergenerational transmission of wealth and ability and incomplete financial markets. Progressive labor income taxes provide social insurance against idiosyncratic income risk and redistributes after tax income among ex-ante heterogeneous households. In addition to the standard distortions of labor supply progressive taxes also impede the incentives to acquire higher education, generating a non-trivial trade-off for the benevolent utilitarian government. The latter distortion can potentially be mitigated by an education subsidy. We find that the welfare-maximizing fiscal policy is indeed characterized by a substantially progressive labor income tax code and a positive subsidy for college education. Both the degree of tax progressivity and the education subsidy are larger than in the current U.S. status quo.

This paper is situated at the intersection of two strands of the literature on optimal labor income taxation, discussed in more detail below.[1] Previous work (see Conesa and Krueger (2006) and Conesa et al. (2009) and the references therein) quantitatively characterized the optimal degree of labor income tax progressivity (within a parametric class of tax functions) in Auerbach and Kotlikoff (1987) style OLG models with idiosyncratic uninsurable wage risk, but took wages over the life cycle as exogenously given.[2] In this paper households partially choose, by deciding on whether to go college, the life cycle wage profile they will be subjected to during their working years. The government, taking as given the behavioral and general equilibrium responses to an (unexpected) tax reform along the transition path[3] induced by the reform, determines the policy that maximizes Utilitarian social welfare among those currently alive at the time of the reform.[4]

Second, a primarily theoretical literature has characterized the optimal combination of progressive labor income taxes and education subsidies in models that abstract from uninsurable income risk and precautionary asset accumulation (see e.g. Benabou (2002), Bovenberg and Jacobs (2005) and Jacobs and Bovenberg, 2010). The latter paper, in particular, highlights how an education subsidy can mitigate the distortions of progressive labor income taxes (motivated by redistributive societal concerns) on the household education decision. Our paper contributes to this strand of the literature by quantifying the optimal policy mix between education finance and progressive income taxation policies.

This paper views itself distinctly in the Ramsey tradition in that, in our attempt to characterize optimal taxation in large-scale OLG models with unin-

[1] See Piketty and Saez (2012) for a comprehensive overview of the theoretical literature.

[2] In addition, in contrast to these papers here we also fully take into account the transitional dynamics induced by the hypothetical tax reforms when computing the optimal tax system.

[3] Fehr and Kindermann (2012) extend Conesa et al.'s (2009) steady state analysis by computing optimal tax transitions in a model that abstracts from endogenous schooling decisions.

[4] The well-being of future generations enters the social welfare function through altruisitc preferences of those currently alive.

surable idiosyncratic risk and endogenous education choices, we restrict the choices of the government to simple (and thus easily implementable) tax policies. We fully acknowledge that the paper is therefore subject to much of the critique of this approach by the New Dynamic Public Finance literature[5] (see e.g. , Kocherlakota (2010), Farhi and Werning (2013), Golosov and Tsyvinski (2013) for representative papers, and Bohacek and Kapicka (2008) and Kapicka (2011) for the analysis of models with endogenous human capital accumulation and education subsidies).

The paper is organized as follows. After relating our contribution to the literature in the next section, in section 3 we construct a simple, analytically tractable model to argue why progressive income taxes and education subsidies might simultaneously be part of an optimal government fiscal policy in the presence of an endogenous education decision. In order to make that argument most clearly, in that section of the paper we abstract from general equilibrium effects of these policies as well as the dynamics induced by asset accumulation and the intergenerational transmission of talent and wealth. These elements are then introduced in the quantitative model in section 4 where we set up the model and define equilibrium for a given fiscal policy of the government. Section 5 describes the optimal tax problem of the government, including its objective and the instrument available to the government. After calibrating the economy to U.S. data (including current tax and education policies) in section 6 of the paper, part 7 displays the results and interpretation of the optimal taxation analysis. Section 8 concludes, and the appendices contain the proofs of the propositions from section 3 as well as details of the calibration and the computation of the quantitative version of the model from section 4 of the main paper.

2 Relation to the Literature

Methodologically, our paper builds on the large literature that uses quantitative OLG models in the spirit of Auerbach and Kotlikoff (1987), enriched by uninsurable idiosyncratic earnings risk as in Bewley (1986), Huggett (1993, 1997) and Aiyagari (1994), to study the optimal structure of the tax code in the Ramsey tradition, see Chamley (1986) and Judd (1985).

The optimal tax code in life cycle economies with a representative household in each generation was characterized in important papers by Alvarez et al. (1992), Erosa and Gervais (2002), Garriga (2003), Gervais (2009) and Bovenberg and Jacobs (2010), and in economies with private information in the Mirrleesian (1971) tradition, by Judd and Su (2006), Fukujima (2010), Bohacek and Kapicka (2008), Kapicka (2011), Findeisen and Sachs[6] and Weinzierl (2011).[7]

[5] Note however, that we do *not* rule out lump-sum taxes. Such taxes are *not optimal* since they contribute to an unfavorable *distribution* of lifetime utilities in society.

[6] The focus of the last three papers on optimal income taxation in the presence of human capital accumulation make them especially relevant for our work, although they abstract from explicit life cycle considerations.

[7] There is also large literature on the positive effects of various taxes on allocations and

Our paper aims at characterizing the optimal progressivity of the income tax code in an economy in which the public provision of redistribution and income insurance through taxation are desirable, but where progressive taxes not only distort consumption-savings and labor-leisure choices, but also household human capital accumulation choices. It is most closely related to the studies by Conesa and Krueger (2006), Conesa et al. (2009) and Karabarbounis (2011). Relative to their steady state analysis we provide a full quantitative *transition analysis* of the optimal tax code in a model with *endogenous education choices*.

In models in which progressive labor income taxes potentially distort education decisions a public policy that subsidizes these choices might be effective in mitigating the distortions from the tax code, as pointed out effectively by Bovenberg and Jacobs (2005). As in their theoretical analysis we therefore study such subsidies explicitly as part of the optimal policy mix in our quantitative investigation. Our focus of the impact of the tax code and education subsidies on human capital accumulation decisions also connects our work to the studies by Heckman et al. (1998, 1999), Benabou (2002), Caucutt et al. (2003), Bohacek and Kapicka (2010), Gallipoli et al. (2011), Guvenen et al. (2011), Holter (2011) and Kindermann (2012), although the characterization of the optimal tax code is not the main objective of these papers.

In our attempt to contribute to the literature on (optimal) taxation in life cycle economies with idiosyncratic risk and human capital accumulation we explicitly model household education decisions (and government subsidies thereof) in the presence of borrowing constraints and the intergenerational transmission of human capital as well as wealth. Consequently our works builds upon the huge theoretical and empirical literature investigating these issues, surveyed partially in, e.g. Cunha et al. (2006), Holmlund et al. (2011), Lochner and Monge (2011).[8]

3 A Simple Model

We now present a simple model[9] that allows us to make precise the intuition that with incomplete financial markets progressive labor income taxes might be part of optimal fiscal policy because it implements a more equitable consumption distribution than the laissez faire competitive equilibrium, but that it distorts both the labor supply *and* the education decision. The latter distortion can be partially offset by an education subsidy which then becomes part of an optimal policy mix as well. Relative to the quantitative model used in the next sections, the model analyzed here abstracts from general equilibrium feedbacks

prices in life cycle economies. See e.g. Hubbard and Judd (1986), Castañeda et al. (1999) for representative examples.

[8] A comprehensive survey of this literature is well beyond the scope of this introduction. We will reference the papers on which our modeling assumptions or calibration choices are based specifically in sections 4 and 6.

[9] The model shares important aspects with the analysis in Bovenberg and Jacobs (2005, 2009).

and the two key sources of dynamics, endogenous capital accumulation and the intergenerational transmission of talent and wealth.

3.1 The Environment

The economy lasts for one period and is populated by a continuum of measure 1 of households that differ by ability e. The population distribution of e is uniform on the unit interval, $e \sim U[0,1]$. Households value consumption and dislike labor according to the utility function

$$\log\left(c - \mu \frac{l^{1+\frac{1}{\psi}}}{1+\frac{1}{\psi}}\right).$$

These Greenwood, Hercowitz and Huffman (1988) preferences rule out wealth effects on labor supply (which greatly enhances the analytical tractability of the model), but at the same time make utility strictly concave in consumption, which induces a redistribution/insurance motive for a utilitarian social planner or government.

A household can either go to college or not. A household with ability e that has gone to college produces $(1+pe)w$ units of consumption per unit of labor, whereas a household without a college degree has labor productivity w. Here $w > 0$ and $p > 0$ (the college premium for the most able type) are fixed positive parameters. Going to college requires κw resources (but no time) where $\kappa > 0$ is a parameter.

3.2 Social Planner Problem

Prior to analyzing the competitive equilibrium without and with fiscal policy we establish, as a benchmark, how a social planner with utilitarian social welfare function would allocate consumption and labor across the population. The social planner problem chooses consumption and labor supply $c(e), l(e)$ for each type $e \in [0,1]$ as well as the set I of types that are being sent to college to solve

$$\begin{aligned} & \max_{c(e),l(e),I} \int \log\left(c(e) - \mu \frac{l(e)^{1+\frac{1}{\psi}}}{1+\frac{1}{\psi}}\right) de \\ & s.t \\ \int c(e)de + \kappa w \int_{e\in I} de \quad = \quad & \int_{e\notin I} wl(e)de + \int_{e\in I} (1+pe)wl(e)de \end{aligned}$$

The following proposition summarizes its solution, under the following

Assumption 1

$$\frac{\left(\frac{\mu^{\psi}(1+\psi)\kappa}{w^{\psi}} + 1\right)^{\frac{1}{1+\psi}} - 1}{p} < 1$$

Proposition 1 *Suppose assumption is satisfied. Then the solution to the social planner problem is characterized by an ability threshold e^{SP} such that all households with $e \geq e^{SP}$ are sent to college (and indexed with subscript c from now on) and the other households are not (and are indexed by n). Labor allocations are given by*

$$l_n = \left(\frac{w}{\mu}\right)^{\psi} \text{ for all } e < e^{SP} \tag{1}$$

$$l_c(e) = \left(\frac{(1+pe)w}{\mu}\right)^{\psi} \text{ for all } e \geq e^{SP} \tag{2}$$

Consumption allocations are characterized by

$$c(e) = c_n \text{ for all } e < e^{SP}$$

$$c(e) = c_c(e) = c_n + \mu \frac{l_c(e)^{1+\frac{1}{\psi}}}{1+\frac{1}{\psi}} - \mu \frac{l_n^{1+\frac{1}{\psi}}}{1+\frac{1}{\psi}} \text{ for all } e \geq e^{SP}$$

The optimal education threshold satisfies the first order condition

$$\kappa w = \frac{w^{1+\psi}}{\mu^{\psi}(1+\psi)} \left((1+pe^{SP})^{1+\psi} - 1\right) \tag{3}$$

and is given in closed form as

$$e^{SP} = \frac{\left(\frac{\mu^{\psi}(1+\psi)\kappa}{w^{\psi}} + 1\right)^{\frac{1}{1+\psi}} - 1}{p} = e^{SP}(\kappa, p, w, \mu) \tag{4}$$

Thus the larger is p, w and the smaller is κ, μ, the smaller is the education threshold and thus the more households are sent to college.

Proof. The threshold property of I follows from the fact that the cost of college is independent of e and the productivity benefits pe of being college-educated are strictly increasing in e. The other results are directly implied by the first order conditions (which in the case of the education threshold e^{SP} involves applying Leibnitz' rule to the resource constraint, after having substituted in the optimal labor allocations). Assumption 1 assures that $e^{SP} \in (0,1)$. Note that this assumption is purely in terms of the structural parameters of the model and requires that the college productivity premium p is sufficiently large, relative to the college cost κ. ■

Equation (3) has an intuitive interpretation. The social planner chooses the optimal education threshold such that the cost of education for the marginal type, κw, equals the net additional resources this marginal type generates with a college education, relative to producing without having obtained a college education. The term on the right hand side of (3) takes into account that college educated households work longer hours (this explains the exponent $1+\psi$) and the fact that college-educated households are compensated for their longer hours with extra consumption which explains the factor $\frac{1}{\mu^{\psi}(1+\psi)}$.

For future comparison with equilibrium consumption allocations we state:

Corollary 2 *The optimal consumption allocation satisfies* $c_n < c_c(e^{SP})$ *and* $c_c(e)$ *is strictly increasing in* e *for all* $e > e^{SP}$.

We depict the optimal consumption allocation in figure 1, together with two equilibrium allocations discussed in the next subsection.[10]

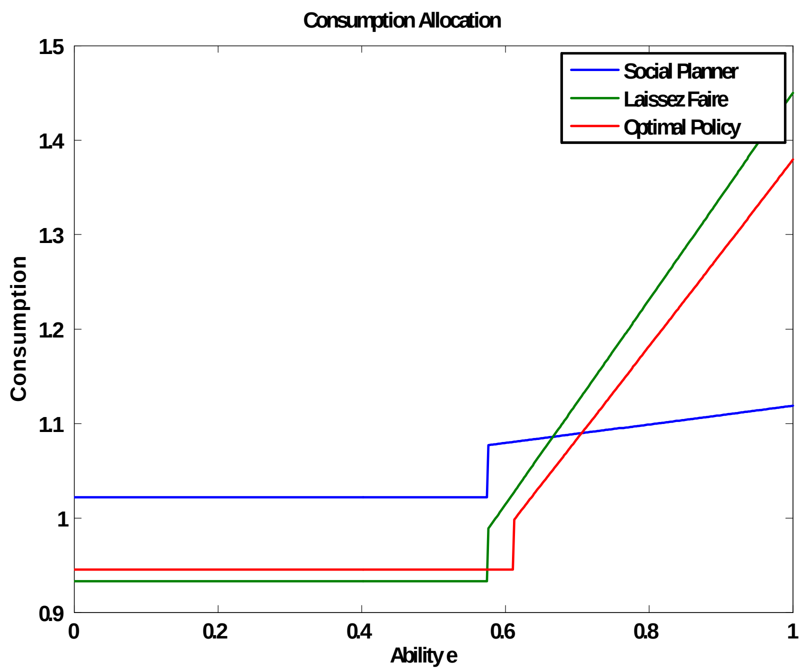


Figure 1: Optimal and Equilibrium Consumption Allocations

3.3 Competitive Equilibrium

Now we study the competitive equilibrium of this economy. To do so we first have to specify the market structure and the government policies. Households participate in two competitive markets, the goods market where they purchase consumption goods, and the labor market where they earn a wage per unit of labor supplied that equals their marginal product. In addition to choosing consumption and labor households decide whether to incur the cost κw of going to college. The benefit of doing so is a wage premium $epw > 0$. We denote by $\mathcal{I}(e) \in \{0, 1\}$ the college choice of household type e, with $\mathcal{I}(e) = 1$ if the household goes to college.

Financial markets, however, are assumed to be incomplete. Although there is no scope for intertemporal trade, in principle households would like to trade

[10] The specific parameter values that underly this figure are discussed below as well.

insurance contracts against the risk of being born as a low ability e type, prior to the realization of that risk. It is this insurance against this idiosyncratic wage risk that we rule out by assumption, and this fundamental market failure will induce a motive of insurance/redistribution for the benevolent, utilitarian government in this economy.[11]

We assume that the benevolent, utilitarian government (which from now on we will frequently refer to as the Ramsey government) has access to three fiscal policy instrument, a flat labor income tax with tax rate τ, a lump sum transfer/tax dw and a proportional education subsidy with rate θ. Note that by permitting $d < 0$ we allow the government to levy lump sum taxes; on the other hand it can also implement a progressive labor income tax schedule by setting $d, \tau > 0$. What we do not permit are policies that make taxes or subsidies type-specific by conditioning τ, d, θ on type e. Given these restrictions on the tax code we would not expect that the government can implement the solution to the social planner problem as a competitive equilibrium.

3.3.1 Definition and Characterization of Equilibrium

Now consider the problem of a generic household $e \in [0, 1]$. Given a fiscal policy the household's problem reads as

$$\max_{l(e), c(e) \geq 0, \mathcal{I}(e) \in \{0,1\}} \log \left(c(e) - \mu \frac{l(e)^{1+\frac{1}{\psi}}}{1+\frac{1}{\psi}} \right) \tag{5}$$

$$s.t.$$

$$c(e) = (1-\tau)(1+\mathcal{I}(e)pe)wl(e) + dw - \kappa w(1-\theta)\mathcal{I}(e) \tag{6}$$

Definition 3 *For a given fiscal policy (τ, d, θ) a competitive equilibrium are consumption, labor and education allocations $c(e), l(e), \mathcal{I}(e)$ such that*

1. For all $e \in [0, 1]$, the choices $c(e), l(e), \mathcal{I}(e)$ solve the household maximization problem (5).
2. The government budget constraint is satisfied:
$$dw + \kappa w\theta \int_{\{e:\mathcal{I}(e)=1\}} de = \tau \left(w \int_{\{e:\mathcal{I}(e)=0\}} l(e)de + w \int_{\{e:\mathcal{I}(e)=1\}} (1+pe)l(e)de \right)$$
3. The goods market clears
$$\int c(e)de + \kappa w \int_{\{e:\mathcal{I}(e)=1\}} de = w \int_{\{e:\mathcal{I}(e)=0\}} l(e)de + w \int_{\{e:\mathcal{I}(e)=1\}} (1+pe)l(e)de$$

We can completely characterize the competitive equilibrium, for a given fiscal policy. We summarize the results in the following

[11] What is insurance ex ante (prior to the realization) of the e draws, is redistribution among different e types ex post.

Proposition 4 *Given a policy (τ, d, θ), the optimal labor supply of households not going to college is given by*

$$l_n(\tau) = \left(\frac{(1-\tau)w}{\mu}\right)^{\psi} \tag{7}$$

whereas the optimal labor supply of households with a college education is given by

$$l_c(e;\tau) = \left(\frac{(1-\tau)(1+pe)w}{\mu}\right)^{\psi}. \tag{8}$$

The corresponding consumption allocations read as

$$c_n(\tau, d) = \frac{[(1-\tau)w]^{1+\psi}}{\mu^{\psi}} + dw \tag{9}$$

$$c_c(e;\tau, d, \theta) = \frac{[(1-\tau)w]^{1+\psi}(1+pe)^{1+\psi}}{\mu^{\psi}} + dw - \kappa w(1-\theta) \tag{10}$$

There is a unique education threshold e^{CE} such that all types with $e \geq e^{CE}$ go to college and the others don't. This threshold satisfies

$$c_n(\tau, d) - \frac{\mu l_n(\tau)^{1+\frac{1}{\psi}}}{1+\frac{1}{\psi}} = c_c(e^{CE};\tau, d, \theta) - \frac{\mu l_c(\tau, e^{CE})^{1+\frac{1}{\psi}}}{1+\frac{1}{\psi}}$$

and is explicitly given by[12]

$$e^{CE} = \frac{\left(\frac{(1-\theta)\mu^{\psi}(1+\psi)\kappa}{(1-\tau)^{1+\psi}w^{\psi}} + 1\right)^{\frac{1}{1+\psi}} - 1}{p} = e^{CE}(\tau, \theta; \kappa, p, w, \mu) \tag{11}$$

The threshold e^{CE} is strictly decreasing (the share of households going to college is strictly increasing) in θ, strictly increasing in τ and independent of d.

Proof. The equilibrium labor allocations follow directly from the first order conditions of the household problem. The equilibrium consumption allocations are then implied by plugging equilibrium labor supply into the household budget constraint (6). Thus lifetime utility conditional on not going to college is given by

$$\log\left(c_n(\tau, d) - \frac{\mu l_n(\tau)^{1+\frac{1}{\psi}}}{1+\frac{1}{\psi}}\right)$$

which is constant in ability e, and lifetime utility conditional on going to college reads as

$$\log\left(c_c(e;\tau, d, \theta) - \frac{\mu l_c(\tau, e)^{1+\frac{1}{\psi}}}{1+\frac{1}{\psi}}\right)$$

[12] This result assumes that $e^{CE} \leq 1$. An assumption similar to assumption 1 is required to assure this, and we assume such an assumption to hold for the range of policies considered below.

which is strictly increasing in e. The threshold result for the education decision thus follows, and the threshold itself is determined by the indifference between attending and not attending college at the threshold. ∎

3.3.2 Optimality of Equilibrium without Government Intervention?

In this subsection we show that the unregulated competitive equilibrium displays an optimal (in the sense of solving the social planner problem) labor and education allocation and thus optimal production. However, the consumption distribution is suboptimally highly dispersed in the competitive equilibrium without government policies. The first result (a simple corollary to proposition 4) states that the competitive equilibrium without government intervention has an optimal labor and education allocation. It follows directly from comparing equations (1),(2) and (4) in the social planner problem to equations (7)-(11) in the competitive equilibrium, evaluated at $\tau = d = \theta = 0$.

Corollary 5

$$\begin{aligned} l_n^{CE}(\tau &= 0) = l_n^{SP} \\ l_c^{CE}(e;\tau &= 0) = l_c^{SP}(e) \\ e^{CE}(\tau &= 0, \theta = 0) = e^{SP} \end{aligned}$$

This last result also implies that aggregate output, defined in the competitive equilibrium as

$$\begin{aligned} L^{CE}(\tau,\theta) &= e^{CE}(\tau,\theta) w l_n^{CE}(\tau) + w \int_{e^{CE}}^{1} (1+pe) l_c^{CE}(e;\tau) de \\ &= e^{CE}(\tau,\theta) w \left(\frac{(1-\tau)w}{\mu}\right)^{\psi} + w \int_{e^{CE}(\tau,\theta)}^{1} (1+pe) \left(\frac{(1-\tau)w(1+pe)}{\mu}\right)^{\psi} de \end{aligned}$$

is at the optimal level as well: $L^{CE}(\tau = 0, \theta = 0) = L^{SP}$. Note that aggregate output in the competitive equilibrium is strictly decreasing in the tax rate τ, strictly increasing in the college subsidy θ (since θ raises the share of households going to college and college-educated households are more productive and work longer hours), and independent of the lump-sum tax/subsidy d.

However, the next proposition shows that the consumption distribution in the competitive equilibrium is suboptimally dispersed since non-college households consume too little in the competitive equilibrium, so do non-productive college graduates, and the dependence of consumption on individual ability e is suboptimally high in the competitive equilibrium without government intervention:

Proposition 6 *In the competitive equilibrium for policy $\tau = d = \theta = 0$ (and thus $e^{CE} = e^{SP}$) we have*

$$\begin{aligned} c_n^{CE} &< c_n^{SP} \\ c_c^{CE}(e^{SP}) &< c_c^{SP}(e^{SP}) \\ \frac{\partial c_c^{CE}(e)}{\partial e} &> \frac{\partial c_c^{SP}(e)}{\partial e} \end{aligned}$$

Proof. See Appendix A ■

The equilibrium consumption allocation in the absence of government policy is depicted in figure 1 and shows the excess consumption inequality proved in proposition 6. The social planner, relative to the competitive equilibrium without policies, provides additional consumption insurance, both between education groups, and within the high education group. As discussed in the beginning of this section, the fundamental market failure that leads to the suboptimality of the competitive equilibrium is the absence of insurance markets against e-risk.[13]

Also note that the socially optimal allocation cannot be implemented as a competitive equilibrium with the policies (τ, d, θ) unless policies can be made e-type specific. This can be seen from recognizing that, under the restricted policies, insuring $\frac{\partial c_c^{CE}(e)}{\partial e} = \frac{\partial c_c^{SP}(e)}{\partial e}$ requires a positive labor income tax $\tau > 0$ that satisfies

$$\frac{1}{1+\frac{1}{\psi}} = (1-\tau)^{1+\psi} \tag{12}$$

but such a tax distorts the labor supply decisions of households, a distortion that cannot be corrected with the existing set of instruments (see equations (7) and (8)).

3.4 Towards Optimal Policy

3.4.1 Macroeconomic Effects of Progressive Taxation and Education Subsidies

The previous section has shown that in the competitive equilibrium without policy consumption of households without college is suboptimally low and consumption of college educated depends suboptimally strongly on their ability e. Investigating the household budget constraint (6) we observe that these two concerns can both be mitigated by implementing a lump-sum transfer $d > 0$ financed by a proportional labor income tax, $\tau > 0$. It thus might be part of the optimal policy mix. We now study the consequences of such a policy. In order to do so we note that the government budget constraint reads as (from now on suppressing the CE label whenever unnecessary):

$$d + \theta\kappa(1 - e^{CE}(\tau, \theta)) = \tau L^{CE}(\tau, \theta)/w \tag{13}$$

Recall that neither the education threshold nor aggregate output is a function of the lump-sum tax/subsidy d. This observation immediately results in the following:

Proposition 7 *An increase in lump-sum transfers d, financed by a raise in the income tax rate τ (that is, an increase in the progressivity of the tax code) leads to*

1. *A decline in the fraction e^{CE} of households attending college.*

[13] Of course there exist *nonutilitarian* welfare weights $\mu(e) \neq 1$ under which the socially optimal allocation arises as a competitive equilibrium without government intervention.

2. *A reduction in individual and aggregate labor supply and thus output L^{CE}.*

Proof. Follows directly from the fact that $l_n^{CE}(\tau), l_c^{CE}(e;\tau), e^{CE}(\tau,\theta)$ and $L^{CE}(\tau,\theta)$ are all strictly decreasing in τ and are independent of d. ■

Thus a τ-financed increase in lump-sum transfers $d > 0$ improves the consumption distribution by redistributing towards n-households (and low e college educated households), but it reduces aggregate output through reducing labor supply of all households and lowering the share $1 - e^{CE}$ of households that become more productive through a college education. The latter concern can be offset through education subsidies, as the next proposition shows.

Proposition 8 *An increase in college subsidies θ financed by a reduction in the transfers d leads to*

1. *An increase in the fraction of households attending college (e^{CE} decreases)*
2. *An increase in aggregate labor supply and thus output L^{CE}.*

Proof. Follows directly from the fact that e^{CE} is strictly decreasing in θ and $L^{CE}(\tau,\theta)$ is strictly decreasing in e^{CE} (output increases with more households going to college) and is independent of d. ■

Note that positive education subsidies, when financed by labor income taxes, not only increase aggregate output, but also redistribute from high e-types to low e-college types, but redistribute away from the very low e-types that do not go to college, hence do not enjoy the subsidy but still bear part of the income tax burden. To summarize, in light of an inefficiently dispersed consumption distribution in the unregulated equilibrium (relative to the one chosen by the utilitarian social planner) the implementation of a progressive tax system improves on the consumption distribution, but lowers average consumption by creating disincentives to work and go to college. The latter distortion can be offset by an appropriate education subsidy. It therefore is to be expected that the optimal fiscal policy in this model may feature progressive income taxes ($\tau, d > 0$) and a positive education subsidy, $\theta > 0$. The next subsection will demonstrate that this is indeed the case, at least for a non-empty subset of the parameter space.

3.4.2 The Optimal Policy Mix

Given the full characterization of a competitive equilibrium for a given fiscal policy (τ, d, θ) in proposition , we can now state the optimal fiscal policy problem of the Ramsey government as

$$\max_{\tau,d,\theta}\left\{e^{CE}(\tau,\theta)\log\left(\frac{[(1-\tau)w]^{1+\psi}}{(1+\psi)\mu^{\psi}}+dw\right)+\int_{e^{CE}(\tau,\theta)}^{1}\log\left(\frac{[(1-\tau)(1+pe)w]^{1+\psi}}{(1+\psi)\mu^{\psi}}+dw+\kappa(1-\theta)w\right)de\right\}$$

subject to

$$d+\theta\kappa(1-e^{CE}(\tau,\theta))=\tau L(\tau,\theta)/w$$

with

$$
\begin{aligned}
e^{CE}(\tau,\theta) &= \frac{\left(\frac{(1-\theta)\mu^{\psi}(1+\psi)\kappa}{(1-\tau)^{1+\psi}w^{\psi}}+1\right)^{\frac{1}{1+\psi}}-1}{p} \\
L(\tau,\theta)/w &= \left(\frac{(1-\tau)w}{\mu}\right)^{\psi}\left(e^{CE}(\tau,\theta)+\int_{e^{CE}(\tau,\theta)}^{1}(1+pe)^{1+\psi}de\right)
\end{aligned}
$$

In Appendix A we partially characterize the solution to the Ramsey problem analytically, here we present a simple quantitative example, employing the parameter values summarized in table 1.

Table 1: Parameter Values						
Parameter	σ	μ	ψ	w	p	κ
Value	1	2	0.1	1	1	0.55

Table 2 displays characteristics of the socially optimal and equilibrium allocations, the latter for three policy configurations. The second row displays equilibrium outcomes without government intervention, the third row for a restricted optimal policy where the education subsidy θ is constrained to equal zero. Finally, the last row summarizes the equilibrium under the optimal fiscal policy. Figure 2 below plots social welfare over the relevant range of fiscal policies, with τ and θ on the axes, and d adjusted to balance the government budget.

Table 2: Results for Example									
	τ	d	θ	e^*	L	l_n	$l_c(e^*)$	c_n	$c_c(e^*)$
SP	N/A	N/A	N/A	0.58	1.29	0.93	0.98	1.02	1.08
CE	0	0	0	0.58	1.29	0.93	0.98	0.93	0.98
CE	6.4%	8%	0	0.62	1.25	0.93	0.97	0.95	1.00
CE	9%	10.4%	4%	0.61	1.25	0.92	0.97	0.95	1.00

From table 2 we observe that the optimal policy indeed calls for progressive taxes and education subsidies. Comparing rows 1 and 2 shows, as proved above that the competitive equilibrium without government intervention has optimal labor supply, education and production allocations, but a consumption distribution in which low e types consume too little. Introducing a progressive income tax in row 3 leads to an improvement in that distribution, but at the expense of reduced output and a smaller fraction of households attending college. A positive (but quantitatively small) education subsidy raises this share, but requires extra government revenue and thus an even higher labor income tax rate τ. The resulting consumption distribution implied by the optimal fiscal policy is plotted in figure 1, alongside the socially optimal and the laissez faire allocation.

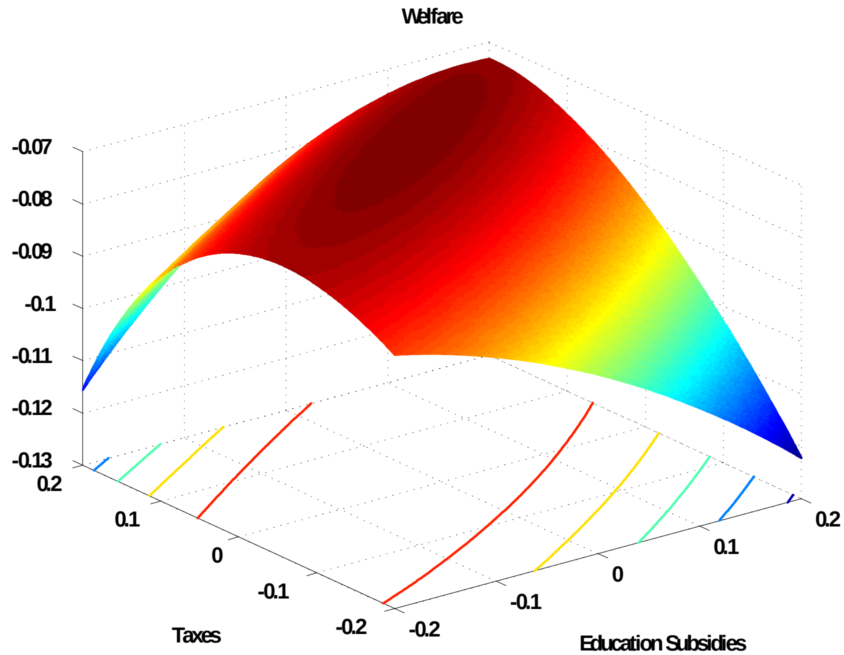


Figure 2: Welfare as a Function of τ and θ

To conclude this section, we have developed a simple model with risk-averse households that make endogenous labor supply and education decisions and used it to argue that progressive income taxes and positive education subsidies are part of an optimal second best policy mix, in the presence of incomplete insurance markets against income risk. We now turn to our quantitative analysis to investigate whether the same statement is true in a realistically calibrated dynamic general equilibrium model with overlapping dynasties, and to quantify the optimal degree of income tax progressivity and magnitude of optimal education subsidies.

4 The Quantitative Model

4.1 Demographics

Population grows at the exogenous rate χ. We assume that parents give birth to children at the age of j_f and denote the fertility rate of households by f, assumed to be the same across education groups.[14] Notice that f is also the number of children per household. Further, let φ_j be the age-specific survival

[14] Note that due to the endogeneity of the education decision in the model, if we were to allow differences in the age at which households with different education groups have children it is hard to assue that the model has a stationary joint distribution over age and skills.

rate. We assume that $\varphi_j = 1$ for all $j = 0, \ldots, j_r$ and $0 < \varphi_j \leq 1$ for all $j = j_r + 1, \ldots, J$, where j_r is the retirement age and J denotes the maximum age (hence $\varphi_J = 0$). The population dynamics are then given by

$$N_{t+1,0} = f \cdot N_{t,j_f} \tag{14}$$

$$N_{t+1,j+1} = \varphi_j \cdot N_{t,j}, \quad \text{for } j = 0, \ldots, J. \tag{15}$$

Observe that the population growth rate is then given by

$$\chi = f^{\frac{1}{j_f+1}} - 1. \tag{16}$$

4.2 Technology

We refer to workers that have completed college as skilled, the others as unskilled. Thus the skill level s of a worker falls into the set $s \in \{n, c\}$ where $s = c$ denotes college educated individuals. We assume that skilled and unskilled labor are imperfectly substitutable in production (see Katz and Murphy (1992) and Borjas, 2003) but that within skill groups labor is perfectly substitutable across different ages. Let $L_{t,s}$ denote aggregate labor of skill s, measured in efficiency units and let K_t denote the capital stock.

Total labor efficiency units at time t, aggregated across both education groups, is then given by

$$L_t = \left(L_{t,n}^{\rho} + L_{t,c}^{\rho}\right)^{\frac{1}{\rho}} \tag{17}$$

where $\frac{1}{1-\rho}$ is the elasticity of substitution between skilled and unskilled labor.[15] Note that as long as $\rho < 1$, skilled and unskilled labor are imperfect substitutes in production, and $\theta$the college wage premium is not constant, but will endogenously respond to changes in government policy.

Aggregate labor is combined with capital to produce output Y_t according to a standard Cobb-Douglas production function

$$Y_t = F(K_t, L_t) = K_t^{\alpha} L_t^{1-\alpha} = K_t^{\alpha} \left[\left(L_{t,n}^{\rho} + L_{t,c}^{\rho}\right)^{\frac{1}{\rho}}\right]^{1-\alpha} \tag{18}$$

where α measures the elasticity of output with respect to the input of capital services.

As always, perfect competition among firms and constant returns to scale in the production function implies zero profits for all firms at all t, and an indeterminate size distribution of firms. Thus there is no need to specify the ownership structure of firms in the household sector, and without loss of generality we can assume the existence of a single representative firm.

This representative firm rents capital and hires the two skill types of labor on competitive spot markets at prices $r_t + \delta$ and $w_{t,s}$, where r_t is the interest rate, δ the depreciation rate of capital and $w_{t,s}$ is the wage rate per unit of labor of skill s. Furthermore, denote by $k_t = \frac{K_t}{L_t}$ the "capital intensity"—defined as

[15] Katz and Murphy (1992) report an elasticity of substitution across education groups of $\sigma = 1.4$. This is also what Borjas (2003) finds, using a different methodology and dataset.

the ratio of capital to the CES aggregate of labor. Profit maximization of firms implies the standard conditions

$$r_t = \alpha k_t^{\alpha-1} - \delta \tag{19}$$

$$w_{t,n} = (1-\alpha)k_t^{\alpha}\left(\frac{L_t}{L_{t,n}}\right)^{1-\rho} = \omega_t\left(\frac{L_t}{L_{t,n}}\right)^{1-\rho} \tag{20}$$

$$w_{t,c} = (1-\alpha)k_t^{\alpha}\left(\frac{L_t}{L_{t,c}}\right)^{1-\rho} = \omega_t\left(\frac{L_t}{L_{t,c}}\right)^{1-\rho} \tag{21}$$

where $\omega_t = (1-\alpha)k_t^{\alpha}$ is the marginal product of total aggregate labor L_t. The college wage premium is then given by

$$\frac{w_{t,c}}{w_{t,n}} = \left(\frac{L_{t,n}}{L_{t,c}}\right)^{1-\rho} \tag{22}$$

which depends on the relative supplies of non-college to college labor and the elasticity of substitution between the two types of skills, and thus is endogenous in our model.

4.3 Household Preferences and Endowments

4.3.1 Preferences

Households are born at age $j = 0$ and form independent households at age j_a, standing in for age 18 in real time. Households give birth at the age j_f and children live with adult households until they form their own households. Hence for ages $j = j_f, \ldots, j_f + j_a - 1$ children are present in the parental household. Parents derive utility form per capita consumption of all households members and leisure that are representable by a standard time-separable expected lifetime utility function

$$E_{j_a}\sum_{j=j_a}^{J}\beta^{j-j_a}u\left(\frac{c_j}{1+\mathbf{1}_{\mathcal{J}_s}\zeta f}, \ell_j\right) \tag{23}$$

where c_j is total consumption, ℓ_j is leisure and $\mathbf{1}_{\mathcal{J}_s}$ is an indicator function taking the value one during the period when children are living in the respective household, that is, for $j \in \mathcal{J}_s = [j_f, j_f + j_a - 1]$, and zero otherwise. $0 \leq \zeta \leq 1$ is an adult equivalence parameter. Expectations in the above are taken with respect to the stochastic processes governing mortality and labor productivity risk.

We model an additional form of altruism of households towards their children. At parental age j_f, when children leave the house, the children's' expected lifetime utility enters the parental lifetime utility function with a weight $\upsilon\beta^{j_f}$, where the term β^{j_f} simply reflects the fact that children's' lifetime utility enters parental lifetime utility at age j_f, and the parameter υ measures the strength of parental altruism.[16]

[16]Evidently the exact timing when children lifetime utility enters that of their parents is

4.3.2 Human Capital Accumulation Technology

At age $j = 0$, before any decision is made, households draw their innate ability to go to college, $e \in \{e_1, e_2, \ldots, e_N\}$ according to a distribution $\pi_{s_p}(e)$ that depends on the education level of their parents $s_p \in \{n, c\}$.[17] A household with ability e incurs a per-period resource cost of going to college $w_{t,c}\kappa$ that is proportional to the aggregate wage of the high-skilled, $w_{t,c}$.[18] In case the government chooses to implement education subsidies, a fraction θ_t of the resource cost is borne by the government.

Going to college also requires a fraction $0 < \xi(e) < 1$ of time, for all j_c periods in which the household attends school. The dependence of the time cost function ξ on innate ability to go to college reflects the assumption that more able people require less time to learn and thus can enjoy more leisure time or work longer hours while attending college (the alternative uses of an individual's time.[19] The education decision is made at age j_a for all subsequent periods $j_a, \ldots, j_c$. A household that completed college has skill $s = c$, a household that did not has skill $s = n$.

4.3.3 Endowments

In each period of their lives households are endowed with one unit of productive time. A household of age j with skill $s \in \{n, c\}$ earns a wage

$$w_{t,s}\epsilon_{j,s}\gamma_s(e)\eta$$

per unit of time worked. Wages depend on a deterministic age profile $\epsilon_{j,s}$ that differs across education groups, on the skill-specific average wage $w_{t,s}$, a component $\gamma_s(e)$ that makes wages depend on innate ability and an idiosyncratic stochastic shock η. The shock η is mean-reverting and follows an education-specific Markov chain with states $\mathcal{E}_s = \{\eta_{s1}, \ldots, \eta_{sM}\}$ and transitions $\pi_s(\eta'|\eta) > 0$. Let Π_s denote the invariant distribution associated with π_s. Prior to making the education decision a household's idiosyncratic shock η are drawn from Π_n, respectively. We defer a detailed description of the exact forms for $\gamma_s(e)$ and $\pi_s(\eta'|\eta)$ to the calibration section.

---

inconsequential. We can simply rescale υ to offset changes in the time disount factor β^{j_f} and leave the effective degree of altruism $\upsilon\beta^{j_f}$ unchanged. Similarly, the parameter υ captures the utility parents receive from all of their f (identical) children. One could write $\upsilon = \tilde{\upsilon}f$, where $\tilde{\upsilon}$ is per-child altruism factor, but this of course leaves both the dynamic programming problem as well as the calibration of the model unchanged (since $\tilde{\upsilon}$ would turn out to equal υ/f in our calibration).

[17] Ability e in our model does not only capture innate ability in the real world since it also stands in for all characteristics of the individual at the age of the college decision, that is, everything learned in primary and secondary education. In our model one of the benefits of going to college is to be able to raise children that will (probabilistically) be more able to go to college.

[18] Gallipoli et al use a time cost instead of a monetary cost reflecting "psychic stress" based on Heckman, Lochner, Todd (2005). Our specification is closer to Caucutt et al. where the costs stand in for hiring a teacher to acquire education.

[19] With this time cost we also capture utility losses of poorer households who have to work part-time to finance their college education.

Households start their economic life at age j_a with an initial endowment of financial wealth $b \geq 0$ received as inter-vivos transfer from their parents.[20] Parents make these transfers, assumed to be noncontingent on the child's education decision[21], at their age j_f, after having observed their child's ability draw e. This transfer is restricted to be nonnegative. In addition to this one-time intentional intergenerational transfer b, all households receive transfers from accidental bequests. We assume the assets of households that die at age j are redistributed uniformly across all households of age $j - j_f$, that is, among the age cohort of their children. Let these age dependent transfers be denoted by $Tr_{t,j}$

4.4 Market Structure

We assume that financial markets are incomplete in that there is no insurance available against idiosyncratic mortality and labor productivity shocks. Households can self-insure against this risk by accumulating a risk-free one-period bond that pays a real interest rate of r_t. In equilibrium the total net supply of this bond equals the capital stock K_t in the economy, plus the stock of outstanding government debt B_t.

Furthermore we severely restrict the use of credit to self-insure against idiosyncratic labor productivity and thus income shocks by imposing a strict credit limit. The only borrowing we permit is to finance a college education through student loans. Households that borrow to pay for college tuition and consumption while in college face age-dependent borrowing limits of $\underline{A}_{j,t}$ (whose size depends on the degree to which the government subsidizes education) and also face the constraint that their balance of outstanding student loans cannot increase after they have completed school. This assumption rules out that student loans are used for general consumption smoothing over the life cycle.

The constraints $\underline{A}_{j,t}$ are set such that student loans need to be fully repaid by age j_r at which early mortality sets in. This insures that households can never die in debt and we do not need to consider the possibility and consequences of personal bankruptcy. Beyond student loans we rule out borrowing altogether.

[20] This is similar to Gallipoli et al. (2008). We model this as a one time payment only. The transfer payment captures the idea that parents finance part of the higher education of their children. Our simplifying assumptions of modelling these transfers are a compromise between incorporating directed inter-generational transfers of monetary wealth in the model and computational feasibility.

If we were to model flexible inver-vivo transfers at all ages $j = j_f, \ldots, j_f + j_c$, we would have to deal with two continuous state variables. Both their own as well as their parents' assets would be relevant for children's decisions at all ages $j = j_a, \ldots, j_f$. An additional continuous state variable is also required if we were to assume that parents commit to pay constant transfers b at all ages $j_f, \ldots, j_f + j_c$ which would perhaps have a more realistic flavor than assuming a one-time transfer. During those years b is a state variable for the childrens' problem. Note that if parental borrowing constraints are not binding one-time transfers are equivalent to a commitment to transfers for many periods (as long as the contingency of parental death is appropriately insured). Thus the issue whether our assumption is quantitatively important depends on the specification of the borrowing constraint, and, given this specification, whether the constraint often binds for households at age j_f.

[21] Note that parents of course understand whether, given b, children will go to college or not, and thus can affect this choice by giving a particular b.

This, among other things, implies that households without a college degree can never borrow

4.5 Government Policies

The government needs to finance an exogenous stream G_t of non-education expenditures and an endogenous stream E_t of education expenditures. It can do so by issuing government debt B_t, by levying linear consumption taxes $\tau_{c,t}$ and income taxes $T_t(y_t)$ which are not restricted to be linear. The initial stock of government debt B_0 is given. We restrict attention to a tax system that discriminates between the sources of income (capital versus labor income), taxes capital income $r_t a_t$ at the constant rate $\tau_{k,t}$, but permits labor income taxes to be progressive or regressive. Specifically, the total amount of labor income taxes paid takes the following simple linear form

$$T_t(y_t) = \max\{0, \tau_{l,t}\,(y_t - d_t \bar{y}_t)\} \tag{24}$$

$$= \max\{0, \tau_{l,t}\,(y_t - Z_t)\} \tag{25}$$

where y_t is household taxable income labor income (prior to a potential deduction) and $\bar{y}_t = \frac{Y_t}{N_t}$ is per-capita income in the economy. Note that the tax system is potentially progressive (if $d_t > 0$), and that lump-sum taxes are permitted, too (the case $\tau_{lt} = 0$ and $d_t < 0$) Therefore for every period there are three policy parameters on the tax side, $(\tau_{kt}, \tau_{lt}, d_t)$.

The government uses tax revenues to finance education subsidies θ_t and finance exogenous government spending

$$G_t = gy \cdot Y_t$$

where the share of output $gy = \frac{G_t}{Y_t}$ commanded by the government is a parameter to be calibrated from the data.[22]

In addition the government administers a pure pay-as-you-go social security system that collects payroll taxes $\tau_{ss,t}$ and pays benefits $p_{t,j}(e, s)$, which will depend on the wages a household has earned during her working years, and thus on her characteristics (e, s) as well as on the time period in which the household retired (which, given today's date t can be inferred from the current age j of the household). In the calibration section we describe how we approximate the current U.S. system with its progressive benefit schedule through the function $p_{t,j}(e, s)$. Since we are interested in the optimal progressivity of the income tax schedule *given* the current social security system it is important to get the progressivity of the latter right, in order to not bias our conclusion about the desired progressivity of income taxes. In addition, the introduction of social security is helpful to obtain more realistic life cycle saving profiles and an empirically more plausible wealth distribution.

[22] Once we turn to the determination of optimal tax and subsidy policies we will treat G rather than gy as constant. A change in policy changes output Y_t and by holding G fixed we assume that the government does not respond to the change in tax revenues by adjusting government spending (if we held gy constant it would).

Since the part of labor income that is paid by the employer as social security contribution is not subject to income taxes, taxable labor income equals $(1 - 0.5\tau_{ss,t})$ per dollar of labor income earned, that is

$$y_t = (1 - 0.5\tau_{ss,t})w_{t,s}\epsilon_{j,s}\gamma_s(e)\eta l$$

4.6 Competitive Equilibrium

We deal with time sequentially, both in our specification of the model as well as in its computation. For a given time path of prices and policies it is easiest to formulate the household problem recursively, however. In order to do so for the different stages of life we first collect the key decisions and state variables in a time line.

4.6.1 Time Line

1. Newborns individuals are economically inactive but affect parental utility until they form a new household at age j_a.

2. At age j_a a new adult household forms. Initial state variables are age $j = j_a$, parental education s^p, own education $s = n$ (the household does not have a college degree before having gone to college). Then an ability level $e \sim \pi_{s_p}(e)$ is drawn. Then parents decide on the inter-vivos transfer b, which constitute the initial endowment of assets (generically denoted by a). Then initial idiosyncratic labor productivity η is drawn according to Π_n. Thus the state of a household is $z = (j_a, e, s = n, \eta, a = b)$.

3. Given state z, at age j_a the educational decision is made. If a household decides to go to college, she immediately does so at age j_a, and continues until she graduates at age $j = j_c$. Her education state switches to $s = c$ at that age.

4. Ages $j = j_a, \ldots, j_c - 1$. The education decision has been made. The household problem differs between non-college and college households. College households at ages $j_a, \ldots, j_c$ work for non-college wages. A household that goes to college but works part time does so for non-college wages, so that her wage equals
$$w_{t,n}\epsilon_{j,n}\gamma_n(e)\eta$$
where η evolves according to $\pi_n(\eta'|\eta)$. Since after age j_c college-educated households work for college wages, the continuation value changes for college households at age $j = j_c$.

5. Age $j = j_c$: College households still work at non-college wages but have continuation utility involving college income states. The idiosyncratic shock η is re-drawn from Π_c and now evolves according to $\pi_c(\eta'|\eta)$ for college-educated households.

6. Ages $j_c+1,\ldots,j_f-1$: Between age of j_f-1 and j_f the decision problem changes because children now enter the utility function and households maximize over per capita consumption $c_j/(1+\zeta f)$.

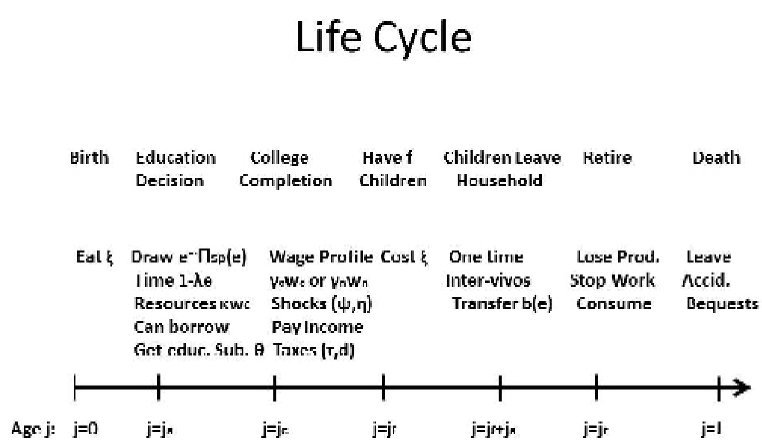
7. Ages $j_a+j_f,\ldots,j_a+j_f-1$: Between age of j_a+j_f-1 and j_a+j_f the decision problem changes again since at j_a+j_f children leave the household and the decision about the inter-vivos transfer b is made and lifetime utilities of children enter the continuation utility of parents.

8. Age j_f: Households make transfers b to their children conditional on observing the skills e of their children.

9. Age $j_a+j_f+1,\ldots,j_r-1$: Only utility from own consumption and leisure enters the lifetime utility at these ages. Labor productivity falls to zero at retirement which is at age j_r.

10. Ages $j=j_r,\ldots,J$: Households are now in retirement and only earn income from capital and from social security benefits $p_{t,j}(e,s)$.

The key features of this time line are summarized in figure



4.6.2 Recursive Problems of Households

We now spell out the dynamic household problems at the different stages in the life cycle recursively.

Child at $j=0,\ldots,j_a-1$ Children live with their parents and command resources, but do not make own economic decisions.

Education decision at j_a Before households make the education decision households draw ability e, their initial labor productivity η and receive inter-vivos transfers b. We specify an indicator function for the education decision as $\mathbf{1}_s = \mathbf{1}_s(e, \eta, b)$, where a value of 1 indicates the household goes to college. Recall that households, as initial condition, are not educated in the first period, $s = n$ and that age is $j = 0$. The education decision solves

$$\mathbf{1}_s(e, \eta, b) = \arg\max\{V_t(j = 0, e, s = n, \eta, a = b), V_t(j = 0, e, s = c, \eta, a = b)\}$$

where $V_t(j = 0, e, s, \eta, a = b)$ is the lifetime utility at age $j = 0$, conditional on having chosen (but not necessarily completed) education $s \in \{n, c\}$.

Problem at $j = j_a, \ldots, j_c - 1$ After having made the education decision, from ages $j = j_a$ to $j = j_c - 1$ households choose how much to work, how much to consume and how much to save.

$$V_t(j, e, s, \eta, a) = \max_{\substack{c, l \in [0, 1 - \mathbf{1}_s \xi(e)] \\ a' \geq -\mathbf{1}_s \underline{A}_{j,t}}} \left\{ u(c, 1 - \mathbf{1}_s \xi(e) - l) + \beta \varphi_j \sum_{\eta'} \pi_n(\eta' | \eta) V_{t+1}(j + 1, e, s, \eta', a') \right\} \tag{26}$$

subject to the budget constraint[23]

$$(1 + \tau_{c,t})c + a' + \mathbf{1}_s(1 - \theta_t)\kappa w_{t,c} + T_t(y_t) = (1 + (1 - \tau_{k,t})r_t)(a + Tr_{t,j}) + (1 - \tau_{ss,t})w_{t,n}\epsilon_{j,n}\gamma_n(e)\eta l \tag{27}$$

$$\text{where } y_t = (1 - 0.5\tau_{ss,t})w_{t,n}\epsilon_{j,n}\gamma_n(e)\eta l \tag{28}$$

Problem at j_c For households with $s = n$, the problem is identical to that at ages $j < j_c$. Households with $s = c$ face exactly the same constraints as before, but their Bellman equation now reads as

$$V_t(j_c, e, c, \eta, a) = \max_{\substack{c, l \in [0, 1 - \xi(e)] \\ a' \geq -\mathbf{1}_s \underline{A}_{j,t}}} \left\{ u(c, 1 - \xi(e) - l) + \beta \varphi_j \sum_{\eta'} \Pi_c(\eta') V_{t+1}(j_c + 1, e, c, \eta', a') \right\}$$

The expectation of the continuation utility is now taken with respect to the stochastic process governing college-educated idiosyncratic productivity (the mean-reverting part η').

[23] At age j_a assets a equal to the transfers from parents. Since these enter the budget constraint of children in the period they are given, for j_a the first term on the right hand side of the budget constraint reads as

$$a + (1 + (1 - \tau_{k,t})r_t)Tr_{t,j}.$$

Problem at $j_c+1,\dots,j_f-1$ At these ages education is completed, thus no time and resource cost for education is being incurred. The problem reads as

$$V_t(j,e,s,\eta,a) = \max_{\substack{c,l\in[0,1]\\ a'\geq -\mathbf{1}_s\underline{A}_{j,t}}} \left\{ u(c,1-l) + \beta\varphi_j \sum_{\eta'} \pi_s(\eta'|\eta) V_{t+1}(j+1,e,s,\eta',a') \right\} \tag{29}$$

subject to the budget constraint

$$\begin{aligned} (1+\tau_{c,t})c + a' + T_t(y_t) &= (1+(1-\tau_{k,t})r_t)(a+Tr_{t,j}) + (1-\tau_{ss,t})w_{t,s}\epsilon_{j,s}\gamma_s(e)\eta l \quad (30)\\ \text{where } y_t &= (1-0.5\tau_{ss,t})w_{t,s}\epsilon_{j,s}\gamma_s(e)\eta l \quad (31) \end{aligned}$$

Problem at ages $j_f,\dots,j_f+j_a-1$ At these ages children live with the household and thus resource costs of children are being incurred. The problem reads as

$$V_t(j,e,s,\eta,a) = \max_{\substack{c,l\in[0,1]\\ a'\geq -\mathbf{1}_s\underline{A}_{j,t}}} \left\{ u\left(\frac{c}{1+\zeta f},1-l\right) + \beta\varphi_j \sum_{\eta'} \pi_s(\eta'|\eta) V_{t+1}(j+1,e,s,\eta',a') \right\} \tag{32}$$

subject to the budget constraint

$$\begin{aligned} (1+\tau_{c,t})c + a' + T_t(y_t) &= (1+(1-\tau_{k,t})r_t)(a+Tr_{t,j}) + (1-\tau_{ss,t})w_{t,s}\epsilon_{j,s}\gamma_s(e)\eta l \quad (33)\\ \text{where } y_t &= (1-0.5\tau_{ss,t})w_{t,s}\epsilon_{j,s}\gamma_s(e)\eta l \quad (34) \end{aligned}$$

Problem at j_f+j_a This is the age of the household where children leave the home, parents give them an inter-vivos transfer b and the children's' lifetime utility enters that of their parents. The dynamic problem becomes

$$\begin{aligned} V_t(j,e,s,\eta,a) = \max_{\substack{c(e'),l(e')\in[0,1],b(e')\geq 0\\ a'(e')\geq -\mathbf{1}_s\underline{A}_{j,t}}} \sum_{e'} \pi_s(e') \Big\{ u(c(e'),1-l(e')) \\ + \beta\varphi_j \sum_{\eta'} \pi_s(\eta'|\eta) V_{t+1}(j+1,e,s,\eta',a'(e')) \\ + \upsilon \sum_{\eta'} \Pi_n(\eta') \max\left[V_t(j_a,e',n,\eta',b(e')), V_t(j_a,e',c,\eta',b(e'))\right] \Big\} \end{aligned} \tag{35}$$

subject to

$$\begin{aligned} (1+\tau_{c,t})c(e') + a'(e') + b(e')f + T_t(y_t) &= (1+(1-\tau_{k,t})r_t)(a+Tr_{t,j}) + (1-\tau_{ss,t})w_{t,s}\epsilon_{j,s}\gamma_s(e)\eta l(e')\\ \text{where } y_t &= (1-0.5\tau_{ss,t})w_{t,s}\epsilon_{j,s}\gamma_s(e)\eta l(e') \end{aligned}$$

Note that since parents can observe the ability of their children e' before giving the transfer, the transfer b (and thus all other choices in that period) are contingent on e'. Also notice that all children in the household are identical. Since parents do not observe the initial labor productivity of their children, parental choices cannot be made contingent on it, and expectations over η' have to be taken in the Bellman equation of the parents over the lifetime utility of their children.[24]

Problem at $j_f + 1, \ldots, j_r - 1$ Now children have left the household, and the decision problem exactly mimics that in ages $j \in \{j_c + 1, \ldots, j_f - 1\}$. Observe that there is a discontinuity in the value function along the age dimension from age j_f to age $j_f + 1$ because the lifetime utility of the child does no longer enter parental utility after age j_f.

Problem at $j_r, \ldots, J$ Finally, in retirement households have no labor income (and consequently no labor income risk). Thus the maximization problem is given by

$$V_t(j, e, s, a) = \max_{c, a' \geq 0} \left\{ u(c, 1) + \beta \varphi_j V_{t+1}(j + 1, e, s, a') \right\} \tag{36}$$

subject to the budget constraint

$$(1 + \tau_{c,t})c + a' = (1 + (1 - \tau_{k,t})r_t)(a + Tr_{t,j}) + p_{t,j}(e, s) \tag{37}$$

4.7 Definition of Equilibrium

Let $\Phi_{t,j}(e, s, \eta, a)$ denote the share of agents, at time t of age j with characteristics (e, s, η, a). For each t and j we have $\int d\Phi_{t,j} = 1$

Definition 9 *Given an initial capital stock K_0, initial government debt level B_0 and initial measures $\{\Phi_{0,j}\}_{j=0}^{J}$ of households, and given a stream of government spending $\{G_t\}$, a competitive equilibrium is sequences of household value and policy functions $\{V_t, a'_t, c_t, l_t, \mathbf{1}_{s,t}, b_t\}_{t=0}^{\infty}$, production plans $\{Y_t, K_t, L_{t,n}, L_{t,c}\}_{t=0}^{\infty}$, sequences of tax policies, education policies, social security policies and government debt levels $\{T_t, \tau_{l,t}, \tau_{c,t}, \theta_t, \tau_{ss,t}, p_{t,j}, (.), B_t\}_{t=0}^{\infty}$, sequences of prices $\{w_{t,n}, w_{t,c}, r_t\}_{t=0}^{\infty}$, sequences of transfers $\{Tr_{t,j}\}_{t=0,j}^{\infty}$ and sequences of measures $\{\Phi_{t,j}\}_{t=1}^{\infty}$ such that*

1. *Given prices, transfers and policies, $\{V_t\}$ solve the Bellman equations described in subsection 4.6.2 and $\{V_t, a'_t, c_t, l_t, \mathbf{1}_{s,t}, b_t\}$ are the associated policy functions.*

[24] Note that we make parents choose their transfers uncontingent on the schooling choice of their children. Mechanically it is no harder to let this choice be contingent on the schooling choice (it then simply would be two numbers). Note that permitting such contingency affects choices, since making transfers contingent permits parents to implicitly provide better insurance against (η, ψ)-risk. If the transfers also could be conditioned on η and ψ, then I conjecture that it does not matter whether they in addition are made conitingent on the education decision of the children or not. Note that in any case, parents can fully think through what transfer induced what education decision.

2. *Interest rates and wages satisfy (4.2).*

3. *Transfers are given by*

$$Tr_{t+1,j-j_f+1} = \frac{N_{t,j} \int (1-\varphi_j) a'_t (j,e,s,\eta,a)\, d\Phi_{t,j}}{N_{t+1,j-j_f+1}} \text{ for all } j \geq j_f \quad (38)$$

4. *Government policies satisfy the government budget constraints*

$$\begin{aligned} \tau_{ss,t} \sum_s w_{t,s} L_{t,s} &= \sum_{j=j_r}^{J} N_{t,j} \int p_{t,j}(e,s) d\Phi_{t,j} \\ G_t + E_t + (1+r_t) B_t &= B_{t+1} + \sum_j N_{t,j} \int T_t(y_t) d\Phi_{t,j} + \tau_{k,t} r_t \left(K_t + B_t\right) + \tau_{c,t} C_t, \end{aligned}$$

where, for each household, taxable income y_t was defined in the recursive problems in subsection 4.6.2 and aggregate consumption and education expenditures are given by

$$E_t = \theta_t \kappa w_{t,c} \sum_{j=j_a}^{j_c} N_{t,j} \int_{\{(e,s,\eta,a): s=c\}} d\Phi_{t,j} \quad (39)$$

$$C_t = \sum_j N_{t,j} \int c_t\left(j,e,s,\eta,a\right) d\Phi_{t,j} \quad (40)$$

5. *Markets clear in all periods t*

$$L_{t,s} = \sum_j N_{t,j} \int \epsilon_{j,s} \gamma_s(e) \eta l_t \left(j,e,s,\eta,a\right) d\Phi_{t,j} \text{ for } s \in n,c \quad (41)$$

$$K_{t+1} + B_{t+1} = \sum_j N_{t,j} \int a'_t \left(j,e,s,\eta,a\right) d\Phi_{t,j} + N_{t,j_f} \sum_{e'} \int \pi_s(e') b'(e'; j,e,s,\eta,a) d\Phi_{t,j_f} \quad (42)$$

$$K_{t+1} = Y_t + (1-\delta) K_t - C_t - CE_t - G_t - E_t. \quad (43)$$

where Y_t is given by (18) and it is understood that the integration in (41) *is only over individuals with skill s. Also*

$$CE_t = (1-\theta_t) \kappa w_{t,c} \sum_{j=j_a}^{j_c} N_{t,j} \int_{\{(e,s,\eta,a): s=c\}} d\Phi_{t,j} \quad (44)$$

is aggregate private spending on education.

6. $\Phi_{t+1,j+1} = H_{t,j}\left(\Phi_{t,j}\right)$ *where $H_{t,j}$ is the law of motion induced by the exogenous population dynamics, the exogenous Markov processes for labor productivity and the endogenous asset accumulation, education and transfer decisions $a'_t, \mathbf{1}_{s,t}, b_t$.*

The law of motion for the measures is explicitly states as follows. Define the Markov transition function at time t for age j as

$$Q_{t,j}\left((e,s,\eta,a),(E\times\mathcal{S}\times\mathcal{E}\times\mathcal{A})\right)=\begin{cases}\sum_{\eta'\in\mathcal{E}}\pi_s(\eta'|\eta) & \text{if } e\in E,\ s\in\mathcal{S},\ \text{and } a'_t\left(j,e,s,\eta,a\right)\in\mathcal{A}\\ 0 & \text{else}\end{cases}$$

That is, the probability of going from state (e,s,η,a) into a set of states $(E\times\mathcal{S}\times\mathcal{E}\times\mathcal{A})$ tomorrow is zero if that set does not include the current education level and education type, and $\mathcal{A}$ does not include the optimal asset choice.[25] If it does, then the transition probability is purely governed by the stochastic shock process for η.

The age-dependent measures are then given, for all $j\geq 1$, by

$$\Phi_{t+1,j+1}((E\times\mathcal{S}\times\mathcal{E}\times\mathcal{A}))=\int Q_{t,j}\left(.,(E\times\mathcal{S}\times\mathcal{E}\times\mathcal{A})\right)d\Phi_{t,j}$$

The initial measure over types at age $j=0$ is more complicated. Households start with assets equal to bequests from their parents determined by the bequest function b_t, draw initial mean reverting productivity according to $\Pi_n(\eta')$, determine education according to the index function $\mathbf{1}_{s,t}$ evaluated at their draw e',η' and the optimal bequests of the parents:

$$\begin{aligned}
&\Phi_{t+1,j=j_a}(\{e'\}\times\{n\}\times\{\eta'\}\times\mathcal{A})\\
&=\Pi_n(\eta')\pi_n(e')\int(1-\mathbf{1}_{s,t}(e',\eta',b_t(e,n,\eta,a;e')))\mathbf{1}_{\{b_t(e,n,\eta,a;e')\in\mathcal{A}\}}\Phi_{t,j_f(n)+j_a}(\{e\}\times\{n\}\times\{\eta\}\times da)\\
&+\Pi_n(\eta')\pi_c(e')\int(1-\mathbf{1}_{s,t}(e',\eta',b_t(e,c,\eta,a;e')))\mathbf{1}_{\{b_t(e,c,\eta,a;e')\in\mathcal{A}\}}\Phi_{t,j_f(c)+j_a}(\{e\}\times\{c\}\times\{\eta\}\times da)\\
&\Phi_{t+1,j=j_a}(\{e'\}\times\{c\}\times\{\eta'\}\times\mathcal{A})\\
&=\Pi_n(\eta')\pi_n(e')\int\mathbf{1}_{s,t}(e',\eta',b_t(e,n,\eta,a;e'))\mathbf{1}_{\{b_t(e,n,\eta,a;e')\in\mathcal{A}\}}\Phi_{t,j_f(n)+j_a}(\{e\}\times\{n\}\times\{\eta\}\times da)\\
&+\Pi_n(\eta')\pi_c(e')\int\mathbf{1}_{s,t}(e',\eta',b_t(e,c,\eta,a;e'))\mathbf{1}_{\{b_t(e,c,\eta,a;e')\in\mathcal{A}\}}\Phi_{t,j_f(c)+j_a}(\{e\}\times\{c\}\times\{\eta\}\times da)
\end{aligned}$$

Definition 10 *A stationary equilibrium is a competitive equilibrium in which all individual functions and all aggregate variables are constant over time.*

[25] There is one exception: at age $j=j_c$ college-educated households redraw their fixed effect. For this group therefore the transtion function at that age reads as

$$Q_{t,j}\left((e,s,\eta,a),(E\times\mathcal{S}\times\mathcal{E}\times\mathcal{A})\right)=\begin{cases}\sum_{\eta'\in\mathcal{E}}\pi_s(\eta'|\eta) & \text{if } e\in E,\ s\in\mathcal{S} \text{ and } a'_t\left(j,e,s,\eta,a\right)\in\mathcal{A}\\ 0 & \text{else}\end{cases}$$

5 Thought Experiment

5.1 Social Welfare Function

Utilitarian for people initially alive

$$SWF(T) = \sum_j N_{t,j} \int V_1(j, e, s, \eta, a; T) d\Phi_{1,j}$$

where $V_1(.; T, \tau_k)$ is the value function in the first period of the transition induced by new tax system (T, τ_k) and $\Phi_1 = \Phi_0$ is the initial distribution of households in the old stationary equilibrium.[26]

5.2 Optimal Tax System

Given initial conditions (K_0, B_0) and a cross-section of households Φ_0 determined by a stationary (to be calibrated policy $\tau_{k,0}, \tau_{l,0}, \theta_0, d_0, b_0 = B_0/Y_0$, the optimal tax reform is defined as the sequence $T^* = \{\tau_{k,t}, \tau_{l,t}, \theta_t, d_t, B_t\}_{t=1}^{\infty}$ that maximizes the social welfare function, i.e. that solves

$$(T^*, \tau_k^*) \in \arg\max_{T \in \Gamma} SWF(T, \tau_k)$$

Here Γ is the set of policies for which an associated competitive equilibrium exists.

Unfortunately the set Γ is too large a policy space to optimize over. The hope is that we can characterize the optimal one-time policy reform, by restricting the sequences that are being optimized over to

$$\begin{aligned} \tau_{k,t} &= \tau_{k,0} \\ \tau_{l,t} &= \tau_{l,1} \\ \theta_t &= \theta_1 \\ d_t &= d_1 \end{aligned}$$

for all $t \geq 1$. Note that the associated debt to GDP ratio will of course not be constant over time. Since all admissible policies defined by $(\tau_{k,2}, \tau_{l,2}, \theta_2, d_2)$ have to lie in Γ, from the definition of equilibrium there must be an associated sequence of $\{B_t\}$ such that the government budget constraint is satisfied in every period. This imposes further restrictions on the set of possible triples $(\tau_{l,1}, \theta_1, d_1)$ over which the optimization of the social welfare function is carried out.

Note that in this version of the paper we restrict the capital income tax rate to remain at its initial (calibrated) value by imposing $\tau_{k,t} = \tau_{k,0}$, that is, in

[26] Note that future generations' lifetime utilities are implicitly valued through the value functions of their parents. Of course there is nothing wrong in principle to additionally include future generations' lifetime utility in the social welfare function with some weight, but this adds additional free parameters (the social welfare weights).

this version of the paper we only determine the optimal mix of (progressive) labor income taxes and education subsidies. Future versions will include the determination of the optimal capital income tax reform as well.[27]

6 Calibration

6.1 Demographics

We take survival probabilities from the Social Security Adminstration life tables. The total fertility rate f population growth rate in the economy is assumed to be $f = 1.14$, reflecting the fact that a mother on average has about 2.3 children. This number also determines the population growth rate in the economy. Households form at age 18 and require 4 years to complete a college education. They have children at age 30 that leave the household 18 years later. Retirement occurs at age 65 and the maximum life span is 100. We describe the remaining model calibration at a yearly frequency, but in our computations we consider a period length of four years.

6.2 Labor Productivity Process

Recall that a household of age j with ability e. skill $s \in \{n, c\}$ and idiosyncratic shock η earns a wage of

$$w_s \epsilon_{j,s} \gamma_s(e)$$

where w_s is the skill-specific wage per labor efficiency unit.

We estimate the deterministic, age- and education-specific component of labor productivity $\{\varepsilon_{j,s}\}$ from PSID data and for both education groups we normalize the mean productivity at labor market entry such that $\epsilon_{j_a,n} = 1$. An appropriate college wage premium will be delivered in the model through the calibration of the $\gamma_s(e)$ term.

We choose the Markov chain driving the stochastic mean reverting component of wages η as a two state Markov chain with education-specific states for log-wages $\{-\sigma_s, \sigma_s\}$ and transition matrix

$$\Pi = \begin{pmatrix} \pi_s & 1-\pi_s \\ 1-\pi_s & \pi_s \end{pmatrix}$$

In order to parameterize this Markov chain we first estimate the following process on the education-specific PSID samples selected by Karahan and Ozkan (2012):

[27]For the case of optimal capital income taxation it would potentially be important to consider tax reforms that are pre-announced one period prior to being implemented. Of course households will respond and adjust their behavior at the time $t = 1$ of the announcement of the reform already. In addition to believing that this timing assumption is a realistic description of the policy lags for a model at annual frequency, it would also avoid the well-known issue in the Ramsey optimal taxation literature that capital is supplied inelastically in period 1 (since the reform was unforeseen in period 0) and thus there would be a very strong incentive to tax this fixed factor in period of the reform.

$$\begin{aligned}\log w_t &= \alpha + z_t \\ z_t &= \rho z_{t-1} + \eta_t\end{aligned}$$

where α is a individual-specific fixed effect that is assumed to be normally distributed (with cross-sectional variance σ_α^2). The estimation results are summarized in the following table:[28]

Table 3: Estimates for Earnings Process			
Group	ρ	σ_η^2	σ_α^2
College	0.969	0.0100	0.0474
Non-College	0.928	0.0192	0.0644

For each education group we choose the two numbers (π_s, σ_s) such that the two-state Markov chain for wages we use has exactly the same persistence and conditional variance as the AR(1) process estimated above.[29] This yields parameter choices given in the next table:

Table 4: Markov Chain for Wages			
Group	π_s	σ_s	$\mathcal{E}_s$
College	0.9408	0.191	$\{0.8113, 1.1887\}$
No College	0.8713	0.250	$\{0.7555, 1.2445\}$

After de-logging, the wage states were normalized so that the mean of the stochastic component of wages equals to 1. We observe that college educated agents face somewhat smaller wage shocks, but that these shocks are slightly more persistent than for non-college educated households.

This leaves us with the ability-dependent fixed component of wages $\gamma_s(e)$. We will calibrate the parameters governing this two wage component so that our model under the status quo policy matches selected wage or earnings observations from the data. We assume that the dependence of wages on ability takes the following log-linear form, for $s \in \{n, c\}$

$$\ln \gamma_s(e) = \vartheta_{0s} + \vartheta_{1s} e.$$

[28] For the details of the sample selection we refer the reader to Karahan and Ozkan (2012) and we thank the authors for providing us with the estimates for the process specified in the main text. In their paper they estimate a richer stochastic process (which, if implemented in our framework, would lead to at least one additional state variable).

[29] The (unconditional) presistence of the AR(1) process is given by ρ and the conditional variance by σ_η^2 whereas the corresponding statistics for the Markov chain read as $2\pi_s - 1$ and σ_s^2, respectively.

For a model where a period lasts 4 years and the AR(1) process is estimated on yearly data, the corresponding statistics are ρ^4 and $(1 + \rho^2 + \rho^4 + \rho^6)\sigma_\eta^2$.

and thus is determined by the four parameters $(\vartheta_{0s}, \vartheta_{1s})$. The distribution of e is discussed in subsection 6.7. We normalize $\vartheta_{0c} = 0$. The remaining three parameters $\vartheta_{0n}, \vartheta_{1n}, \vartheta_{1n}$ are chosen jointly such that the stationary equilibrium of the status quo economy the attains the following targets:

1. A college wage premium of 80% as in U.S. data for the later part of the 2000's (see e.g. Heathcote et al. 2010). This, roughly speaking, pins down ϑ_{0n} which we expect to be less than zero.

2. The variances of the fixed effect for both education groups displayed in the last column of table 3. Note that the variances in the model are given by
$$\sigma_{\alpha s}^2 = Var(\ln \gamma_s(e)) = (\vartheta_{1s})^2 \, Var(e|s)$$
and thus are a function of the parameter ϑ_{1s} and the model-endogenous sorting (by ability) of households into the two different education classes.

Remark 11 *The average college wage premium is*
$$\overline{wp} = \frac{E(\gamma_c(e))}{E(\gamma_n(e))} = \frac{E(\exp(\vartheta_{1c}e)\,|s=c)}{\exp(\vartheta_{0n})E(\exp(\vartheta_{1n}e)\,|s=n)}.$$
For the marginal household with ability $\bar{e}$ not going to college, the expected college premium is
$$\begin{aligned}\frac{w_c(\bar{e})}{w_n(\bar{e})} &= \frac{\exp(\vartheta_{1c}\bar{e})}{\exp(\vartheta_{0n})\exp(\vartheta_{1n}\bar{e})} = \overline{wp} * \frac{\frac{\exp(\vartheta_{1c}\bar{e})}{E(\exp(\vartheta_{1c}e)|s=c)}}{\frac{\exp(\vartheta_{1n}\bar{e})}{E(\exp(\vartheta_{1n}e)|s=n)}} \\ &= \overline{wp} * \frac{\frac{\exp(\vartheta_{1c})\exp(\bar{e})}{\exp(\vartheta_{1c})E(\exp(e)|s=c)}}{\frac{\exp(\vartheta_{1n})\exp(\bar{e})}{\exp(\vartheta_{1n})E(\exp(e)|s=n)}} = \overline{wp} * \frac{\frac{\exp(\bar{e})}{E(\exp(e)|s=c)}}{\frac{\exp(\bar{e})}{E(\exp(e)|s=n)}} = \overline{wp} * \frac{E(\exp(e)\,|s=n)}{E(\exp(e)\,|s=c)}\end{aligned}$$
Thus as long as the education decision has the alleged threshold property such that low e households don't do got college whereas high e households do, the wage premium for the marginal type $\bar{e}$ of going to college is smaller than the average college premium. This is an important observation for the interpretation of the quantitative results.

6.3 Technology

The parameters to be calibrated are (α, δ, ρ). As a benchmark we choose $\rho = 1$, that is, skilled and unskilled labor are perfect substitutes in production. We will investigate the quantitative importance of this crucial assumption in later versions of this paper. The capital share is set to $\alpha = 1/3$. Furthermore we target an investment to output ratio of 20% and a capital-output ratio of 2.65. Accounting for population growth this implies a yearly depreciation rate of 8.4% and thus a yearly interest rate of about 4.2%. The capital-output ratio (equivalently, the real interest rate) will be attained by appropriate calibration of the preference parameters (especially the time discount factor β), as discussed below.

6.4 Government Policy

In the benchmark economy the six policy parameters to be determined are $(\tau_k, \tau_l, \tau_c, \tau_p, d, b, gy)$. We choose $b = 0.6$ and $gy = 0.17$ to match a government debt to GDP ratio of 60% and government consumption (net of tertiary education expenditure) to GDP ratio of 17%. Consumption taxes can estimated from NIPA data as in Mendoza, Razin and Tesar (1994) who find $\tau_c \approx 0.05$. For the capital income tax rate, we adopt Chari and Kehoe's (2006) estimate of $\tau_k = 28.3\%$ for the early 2000's.

The payroll tax $\tau_{ss} = 12.4\%$ is chosen to match the current social security payroll tax (excluding Medicare). We model social security benefits $p_{t,j}(e, s)$ as concave function of average wages earned during a household's working life, in order to obtain a reasonably accurate approximation to the current progressive US benefit formula, but without the need to add a continuous state variable to the model. The details of the calibration of social security benefits are contained in appendix B.1.

We calibrate the labor income tax deduction to match the sum of standard deductions and exemptions from the US income tax code, for a married household with two children. In 2009 a married couple with 2 children had a standard deduction of \$11400 plus 4 times the standard exemption of 3650, totaling \$26,000. Per capita GDP in 2009 was \$46,400, which amounts to \$185,640 for a family of 4.Thus we calibrate the deduction in the benchmark economy to $d = 0.14$, that is 14% of GDP per capita. Finally the marginal tax rate on labor income τ_l is chosen to balance the government budget.

6.5 Preferences

The bequest parameter υ is chosen so that in equilibrium a fraction of 0.32% of total wealth is given as inter-vivos transfers, which Nishiyama reports as the number from the 1986 SCF (summarized by Gale and Scholz, 1994). The same source states that total bequests given in year account for 1% of total wealth, and we evaluate, as an independent test of the model, whether the accidental bequests in our economy amount to approximately the same amount. We specify the period utility function as

$$u(c, l) = \frac{\left[c^{\mu}\left(1 - \mathbf{1}_s \xi(e) - l\right)\right]^{1-\sigma}}{1 - \sigma}$$

We a prioi choose $\sigma = 4$ and then determine the time discount factor β and the weight on leisure μ in the utility function such that in the benchmark model the capital-output ratio is 3 and households on average work 1/3 of their time.[30]

[30] These preferences imply a Frisch elasticity of labor supply of $\left(\frac{1-\mu(1-\sigma)}{\sigma}\right)\left(\frac{1-l}{l}\right)$, and with an average labor supply of $l = 1/3$ one could be worried that the Frisch labor supply elasticity, which, given the parameter estimates will be around 1 for most households, is implausibly high. But note that this elasticity of labor supply of entire households, not that of white prime age males on which many lower empirical estimates are based.

The coefficient of relative risk aversion with this formulation equals $\sigma\mu + 1 - \mu \approx 2$.

6.6 Education Costs and Subsidies

We choose the resource cost for college education κ and the share of expenses borne by the government θ in the benchmark model to match the total average yearly cost of going to college, as a fraction of GDP per capita, $\frac{\kappa w_c}{\bar{y}}$, and the cost net of government subsidies, $\frac{(1-\theta)\kappa w_c}{\bar{y}}$.

To calculate the corresponding numbers from the data we turn to Ionescu and Simpson (2010) who report an average net price (tuition, fees, room and board net of grants and education subsidies) for a four year college (from 2003-04 to 2007-08) to be $\$58,654$ and for a two year college of $\$20,535$. They also report that 67% of all students that finish college completed a 4 year college and 33% a two year college. Thus the average net cost of tuition and fees for one year of college is

$$0.67 * 58,654/4 + 0.33 * 20,535/2 = \$13,213$$

Average GDP per capita during this time span was, in constant 2005 dollars, \$42,684. Thus

$$\frac{(1-\theta)\kappa w_c}{\bar{y}} = 13,213/42,684 = 0.31$$

Furthermore education at a glance (OECD 2008, Table B1.1a) reports that per student expenditures for 2006 on tertiary education equals $\$21,588$.[31] As a fraction of 2005 GDP per capita this equals

$$\frac{\kappa w_c}{\bar{y}} = \frac{21,588}{42,664} = 0.506$$

Consequently we find $1-\theta = 0.31/0.506 = 61.2\%$ and thus a subsidy rate of $\theta = 38.8\%$. The cost parameter κ is calibrated so that the equilibrium of the benchmark model has to be calibrated within the model so that in the model $\frac{\kappa w_c}{\bar{y}} = 0.506$.

6.7 Ability Transitions and College Time Costs

Newly formed households draw their ability from a distribution $\pi_{s^p}(e)$ that depends on the education level of their parents. We interpret $e \in [0,1]$ as basic ability to succeed in college. Based on their ability e the time requirement for attending class and studying in college is given by the linear function

$$\xi(e) = 1 - e$$

so that the children with lowest ability face prohibitively large time costs of going to college, $\xi(e=0) = 1$.

We assume that the distribution $\pi_{s_p}(e)$ follows a normal distribution with parameters μ_{s^p}, σ which is truncated to $[0,1]$ and then discretized to 10 values, $e \in \{e_1 = 0, \ldots, e_{10} = 1\}$. We choose the education (of the parents) specific

[31] These figures exclude expenditures for R&D activities.

means μ_{s^p} to match college completion rates of students by parental education levels, and choose the variance σ such that the probability mass of the original normal distributions located in the unit interval $[0,1]$ is 90% on average over the two groups. The implied coefficient of variation of time spent studying is 0.54 which is well in the range of estimates reported by Babcock (2009).

To obtain college completion rates of students by parental education we turn to the National Education Longitudinal Study (NELS:88).[32] We compute the percent of individuals from this nationally representative sample who were first surveyed as eighth-graders in the spring of 1988, that by 2000 had obtained at least a Bachelors degree, conditional on the highest education level of their parents. We identify $s_p = c$ in our model with the highest education of a parent being at least a Bachelors degree (obtained by 1992). We find that for students with parents in the $s_p = c$ category 63.3% have completed a Bachelors degree. The corresponding number for parents with $s_p = n$ is 28.8%. Although in the model these shares are endogenously determined, they are mainly driven by the values for the education specific means μ_{s^p}.

6.8 Borrowing Constraints

The borrowing constraints faced by agents pursuing a college degree allow such an agent to finance a fraction $\phi \in [0,1]$ of all tuition bills with credit and specify a constant (minimum) payment rp such that at the age of retirement all college loans are repaid. Formally

$$\underline{A}_{j,t} = (1+r_t)\underline{A}_{j-1,t-1} + \phi(1-\theta_t)\kappa w_{t,c}$$

for all $j = 0, \ldots, j_c$ where $\underline{A}_{-1,t} = 0$ is understood. For $j = j_c + 1, \ldots, j_r - 1$ we specify

$$\underline{A}_{j,t} = (1+r_t)\underline{A}_{j-1,t-1} - rp$$

and rp is chosen such that the terminal condition $\underline{A}_{j_r-1,t} = 0$ is met.

The parameter ϕ to be calibrated determines how tight the borrowing constraint for college is. Note that in contrast rp is not a calibration parameter but an endogenously determined repayment amount that insures that households don't retire with outstanding student loans.

The maximum amount of publicly provided student loans for four years is given by \$27,000 for dependent undergraduate students and \$45,000 for independent undergraduate students (the more relevant number given that our students are independent households).[33] Relative to GDP per capita in 2008 of \$48,000, this given maximum debt constitutes 14% and 23.4% of GDP per capita. Compare that to the 31% of total costs computed above, this indicates that shows independent undergraduate students can borrow at most approximately 75% of the cost of college, and thus we set $\phi = 0.75$. The following table summarizes the parameters used in our optimal tax computations.

[32] http://nces.ed.gov/surveys/nels88/

[33] Note that about 66% of students finishing four year colleges have debt, and *conditional* on having debt the average amount is \$23,186 and the median amount is \$20,000.

Table 5: Calibration		
Parameter	Interpretation	Value
Exogenously Calibrated Parameters		
Population		
j_a	Age at HH form. (age 18)	0
j_c	Age, coll. compl. (age 21)	3
j_f	Fertility Age (age 30)	11
j_r	Retirement Age (age 65)	46
J	Max. Lifetime (age 100)	81
f	Fertility Rate	1.14
$\{\varphi_j\}$	Survival Probabilities	Life Tables SSA
Labor Productivity		
$\{\varepsilon_{j,s}\}$	Age Profile	Estimates (PSID)
$\mathcal{E}_s$ and $\pi_s(\eta'\|\eta)$	Stochastic Part of Wages	Estimates (PSID)
σ_e	Std. Dev of e (Coverage of trunc. Normal=90%)	0.278
Preferences		
σ	Coef. of Rel. Risk Aversion = 2	4
ζ	Equivalence Scale	0.3
Technology		
α	Capital Share	0.333
δ	Depreciation	0.084
Ability and Education		
ϕ	Tightness of Borrowing Constraint	0.75
Government Policy		
θ	Education Subsidy	0.388
d	Labor Income Tax Deduction	0.14
τ_c	Consumption Tax Rate	0.05
τ_k	Capital Income Tax Rate	0.283
b	Debt-GDP Ratio	0.6
gy	Gov. Cons to GDP Ratio	0.17
τ^p	Social Security Payroll Tax	0.124
Parameters Calibrated in Equilibrium (Targets in Brackets)		
Preferences		
β	Time Discount Rate (K/Y)	0.989
υ	Altruism Parameter (Avg. Transfers)	0.167
μ	Leisure Share (Fraction of h worked)	0.347
Ability and Education		
μ_{s_p}	Mean Ability (Coll. Compl. Rate by s_p)	$[0.605, 0.639]$
κ	Resource Cost of Coll. (Spend. on Tert. Educ.)	0.434
ϑ_{0s}	$\gamma_s(e) = \vartheta_{0s} + \vartheta_{1s}e$ ($\vartheta_{0s} = 1$, ϑ_{0n} avg. skill prem.)	$[-0.328, 0]$
ϑ_{1s}	(s-specific variance of fixed effect)	$[1.587, 1.479]$
Government Policy		
τ_l	Labor Income Tax Rate (Budget Bal.)	0.175

7 Results

7.1 How the Model Works: The Education Decision

Prior to presenting the optimal tax results it is instructive to discuss how households make their key economic decisions *for a given policy.* Ours is a fairly standard life cycle model with idiosyncratic wage risk, and thus the life cycle profiles of consumption, asset and labor supply are consistent with those reported in the literature (see e.g. Conesa et al. (2009), figure 1). Instead, here we explore how the optimal education decision is made, as a function of the initial characteristics of the household. This focus is further warranted by the observation that the optimal policy will have a strong impact on this decision and will result in a significant change in the share of households obtaining an education in the aggregate, which is in turn important for understanding the optimality of the policy in the first place.

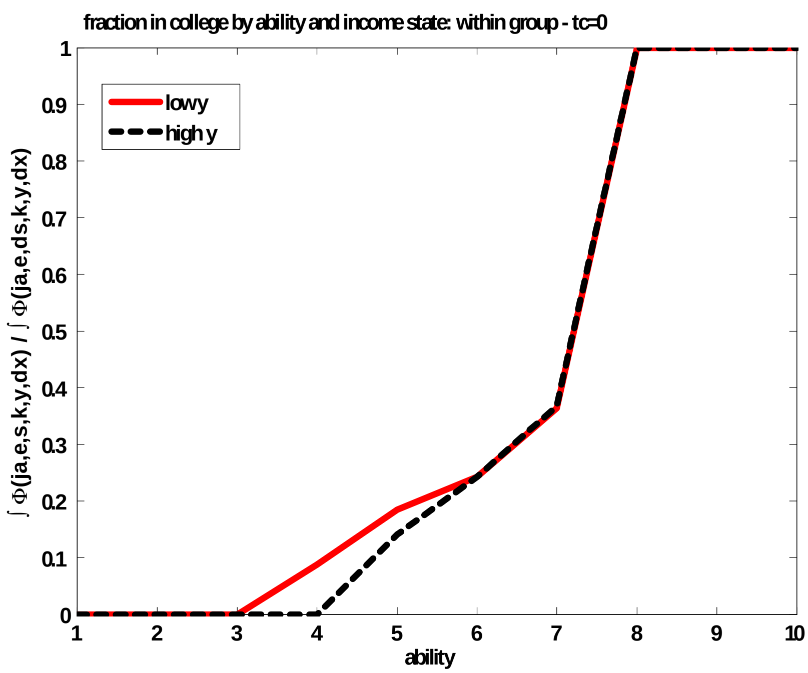


Figure 3: Fraction of Households Deciding to Go to College

Recall that households, at the time of the college decision (that is, at age j_a) differ according to (e, η, b), that is, their ability to go to college e, their wages outside college (as determined by the idiosyncratic shock η), and their initial asset levels resulting from parental transfers b. In figure 3 we display the share of households deciding to go to college, under the status quo policy, as a function of e, both for households with low and with high η (and thus high income y)

realizations. All households with high abilities ($e \geq e_8$) go to college, and non of the households with very low ability ($e \leq e_3$) do. For households in the middle of the ability distribution, their decision depends on the attractiveness of the outside option of working in the labor market: a larger share of households with lower opportunity costs (low η and thus y) attends to college. Finally, a share strictly between zero and one, conditional on η, indicates that wealth heterogeneity among the youngest cohort (which in turn stems from wealth and thus transfer heterogeneity of their parents) is an important determinant of the college decision for those in the middle of the ability distribution ($e \in [e_4, e_7]$).

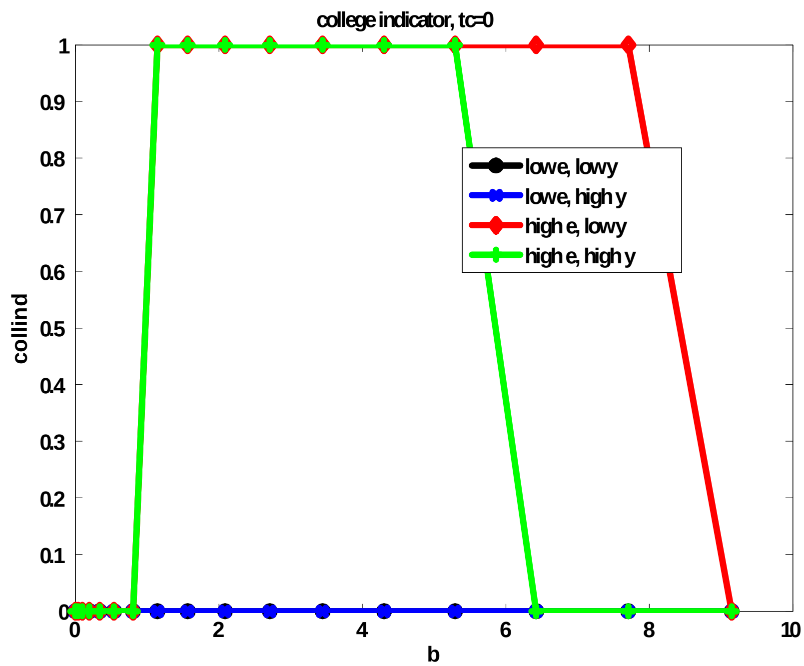


Figure 4: College Decision by Initial Assets

This point is further reinforced by figure 4 which displays the college decision indicator function in dependence of initial assets b, and conditional on the non-college wage realization. A value of 0 on the y-axis stands for not attending college, a value of 1 represents the decision to go to college. Assets on the x-axis are normalized such that a value of $b = 1$ stands for assets equal to one time average asset holdings of the parental generation at the age intergenerational transfers are given. We display the policy function for those with low ability ($e = e_3$) and those with high ability ($e = e_6$). We make several observations. First, low-ability households never go to college, independent of their other characteristics (the low e policy function is identically equal to zero). For households with sufficiently high ability ($e > e_3$) other characteristics matter. As discussed above, a higher non-college wage (high y) reduces the incidence of attending college. Finally and perhaps most interestingly, the effects of initial wealth on the college decision are non-monotone. For households at the low end of the wealth distribution the borrowing constraint is important. Although the government subsidies college (in the status quo it covers a 38.8% share of the costs) and although households can borrow 75% of the remaining resource costs, at zero or close to zero wealth household might still not be able to afford college. That is, either it is impossible for these households to maintain positive consumption even by working full time while attending college, or the resulting low level of consumption and/or leisure make such a choice suboptimal. As parental transfers increase the borrowing constraint is relaxed and able households decide to go to college. Finally, sufficiently wealthy households that expect to derive a significant share of their lifetime income through capital income find it suboptimal to invest in college and bear the time and resource cost in exchange for larger labor earnings after college. Note, however, that although this last result follows from the logic of our model, it is not important quantitatively since the stationary asset transfer distribution puts essentially no mass on initial assets $b \geq 5$.

7.2 The Optimal Policy

Starting from the status quo, the optimal policy as defined above is characterized by a significantly more progressive tax system with a marginal tax rate of $\tau_l =$ 24.1% and a deduction of $d = 32\%$ of average household income. The associated optimal education policy subsidies the resource cost of going to college at a rate of $\theta = 70\%$, close to doubling the subsidy, relative to the status quo policy. The resulting welfare gain from the policy reform and its implied economic transition is equivalent to a uniform increase in consumption (over time, states of the world and households) of approximately 1.2%. In the next two subsections we, in turn, characterize the long run and then transitional consequences of this fundamental tax reform, before turning to an interpretation of the welfare gains implied by it in section 7.2.3.

7.2.1 Comparison of Initial and Final Steady States

In table 6 we summarize the impact of the policy reform on macroeconomic aggregates in the long run, by comparing stationary equilibria under the status quo and the dynamically optimal policy. All variables are denoted in per capita terms.

Table 6: Steady State Comparison			
Var.	Status Quo	Opt. Pol.	Change
τ_l	17.5%	24.1%	6.6%
d	14.0%	32.0%	18%
θ	38.8%	70.0%	31.2%
Y	0.612	0.620	1.3%
D/Y	60.0%	76.8%	16.8%
K	0.406	0.402	-1.0%
L	1.139	1.166	2.4%
K/L	0.542	0.524	-3.3%
w	0.547	0.541	-1.1%
r	4.7%	4.9%	0.2%
$hours$	33%	31.7%	-1.3%
C	0.398	0.405	1.9%
Trans/Assets	0.33%	0.30%	-0.024%
college share	43.9%	57.8%	14%
$Gini(c)$	0.309	0.286	-0.023
$Gini(h)$	0.117	0.112	-0.006
$Gini(a)$	0.607	0.581	-0.026

From the table we observe that the increase in the progressivity of the tax code and simultaneous rise in education subsidy triggers a significant decline in hours worked, by 1.3%. Furthermore, the expansion of government debt along the transition (see next subsection) crowds out physical capital accumulation, so that the steady state capital stock is now 1% lower than in the status quo. The capital-labor ratio falls by 3.3%, and wages per efficiency units decline by 1.1%, whereas the return on assets (and thus the interest rate on government debt) rises by 20 basis points.

However, the policy does not lead to a substantial decline in per capita output, as one might suspect from the decline in capital and hours worked; in fact GDP per capita increases by 1.3%. Key to this finding is the increase in the share of households attending college and thus the shares of workers with a college degree, which is up by 14 percentage points. The doubling of the college subsidy rate more than offsets the reduced incentives to acquire human capital due to a more progressive tax system. The improved skill distribution in the population in turn results in a larger effective labor supply in the new steady state, despite the fact that average hours are significantly down. Aggregate consumption in turn rises by 1.9% in the long run. On the distributional side, consumption, leisure and wealth inequality fall on account of a more redistributive labor income tax schedule, most significantly so for consumption. Overall, the significant reduction in hours worked and increase in aggregate consumption as well as a more equal distribution of resources indicates substantial welfare gains from this policy reform.

7.2.2 Transitional Dynamics

However, this discussion ignores the fact that it takes time (and resources) to build up a more skilled workforce, suggesting that an explicit consideration of the transition path is important. At any point in time, the youngest cohort constitutes just a small share of the overall workforce, so even if the education decision of this cohort is changed drastically on impact in favor of more college education, it takes years, if no decades, until the skill composition of the entire workforce changes significantly (as older, less skilled cohorts retire and younger, more skilled cohorts take over). In figure 5 we plot the evolution of the key macroeconomic variables along the policy-induced transition path. The lower right panel which displays both the share of the youngest cohort going to college as well as the overall fraction of the population highlights the observation described above. Whereas the share of the youngest cohort going to college moves strongly on policy impact, it takes approximately 60 years until the overall skill distribution have reached a level close to its new steady state value. It is this dynamics that a restriction to a steady state policy analysis would miss completely.

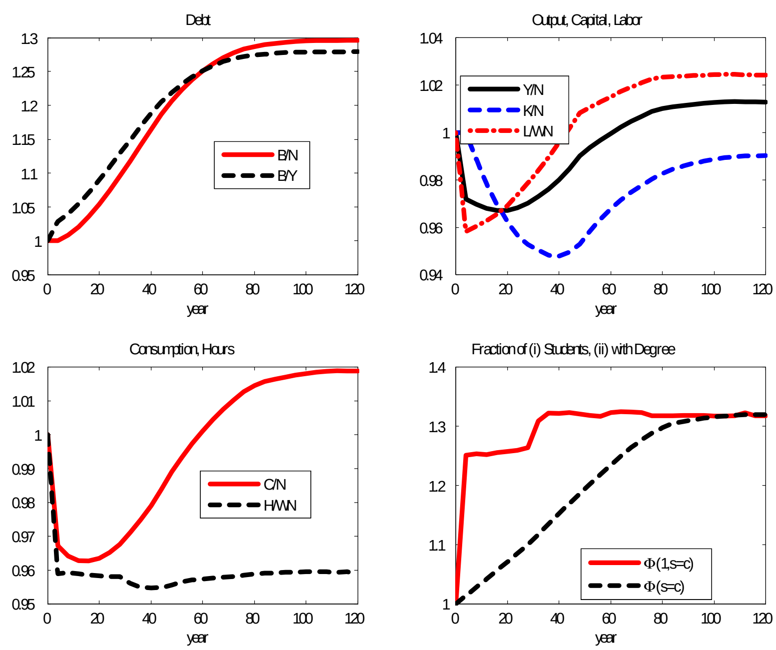


Figure 5: Evolution of Macroeconomic Aggregates

That this omission has potentially profound consequences can be seen from the upper right and the lower left panel of figure 5 which show the evolution of GDP per capita (together with that of capital and effective labor supply), and that of consumption (together with average hours worked). The graphs show that while hours worked respond significantly on impact and the remain fairly flat, effective labor supply falls early on and then recovers as the skill composition of the population changes. Consequently the drop in GDP and consumption per capita is very substantial early in the transition, prior to education-driven transitional growth setting in.

The dynamics of government debt (which is mechanically determined, through the sequence of government budget constraints, given its initial level and the tax and education policies) mirrors that of GDP per capita, as the upper left panel of figure 5 displays. During the transitional years of low economic activity (due to a falling capital stock and lower hours worked) the government accumulates debt and the debt to GDP ratio rises from its 60% level in the initial steady state. As the economy recovers the debt-to-GDP ratio then stabilizes at its higher steady state level of about 77%.

Finally, both the public debt-induced capital crowding-out effect and the early collapse and subsequent recovery of effective labor supply induces substantial swings of the capital-labor ratio and associated movements of wages and interest rates. As figure 6 shows, after the initial collapse of hours and implied increase of this ratio the recovery of effective labor supply and continued decline in the capital stock leads to a lower capital-labor ratio and wages as well as higher interest rates in the long run, whereas households living through the transition enjoy higher wages and lower returns for most of their lifetime.

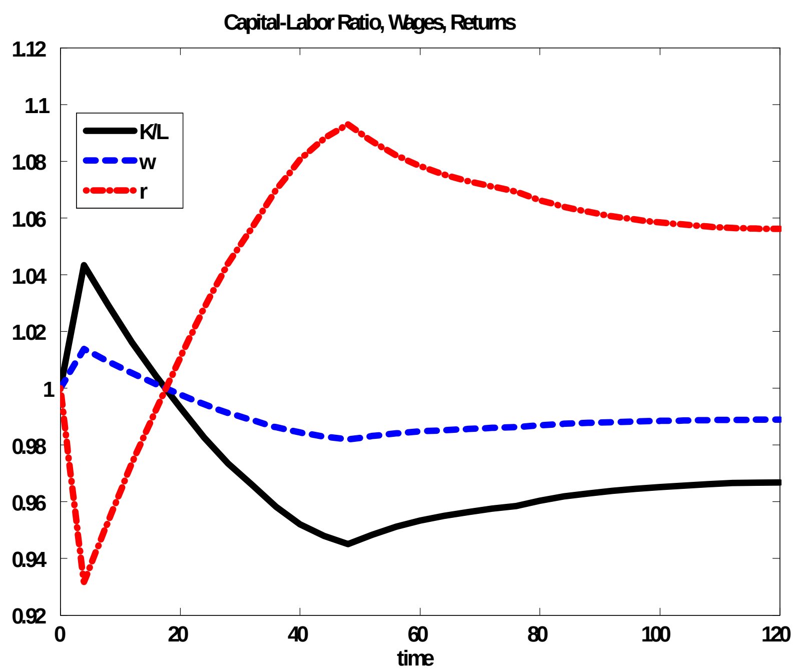


Figure 6: Evolution of Capital-Labor Ratio along the Transition

To conclude this section, the analysis of the transition path induced by the optimal policy indicates that a steady state analysis of welfare might potentially be problematic since it ignores the transitional costs of temporarily lower output and consumption induced by a progressive tax reform that slows down labor supply and capital accumulation.

7.2.3 Interpreting the Optimal Policy and Welfare Results

Despite the previous discussion, the tax reform *does* increase social welfare significantly (in the order of 1.2% of consumption) even when the transitional cost of the policy is fully taken into account. An important element of these gains stems from a more equal distribution of consumption (and also leisure). The substantial increase in labor income tax progressivity induces a gradual reduction, over time, in earnings, consumption and wealth inequality, in the order of about 2-3 points for the Gini coefficient, depending on the variable. The cross-sectional dispersion of leisure, on the other hand, changes relatively little, with a Gini coefficient that falls by less than one percentage point. Thus the aggregate welfare gains documented above stem primarily from two sources, a decline in aggregate hours worked and resulting increase in leisure for the typical household, and from a more equal consumption and leisure distribution. They are significantly mitigated by a substantial decline in aggregate output and thus consumption that the policy brings about *in the short run,* due to lower incentives to work and the crowding out of physical capital by government debt.

Finally note that, relative to the status quo, the optimal policy mix induces a substantial increase in college attendance (and thus, over time, a rising share of high-skilled households) despite the fact that the incentives from the labor income tax side for acquiring a college degree have substantially weakened. The optimal choice of $\theta = 0.7$ is crucial for this finding. More generally, this result points to the important interaction of progressive taxes and education subsidies that Bovenberg and Jacobs (2005) stressed theoretically. In fact, had θ remained constant at its status quo value of 0.388, a change in the progressivity of the labor income tax alone (to $d = 32\%$) would have led to a decline in the share of the young cohort going to college by 3% in the short run and 6% in the long run, an optimal of only $d = 30\%$ and welfare gains of only 0.3% of consumption. In this sense an important complementarity exists between progressive taxation and education subsidy.

8 Conclusions

In this paper we characterized the optimal mix of progressive income taxes and education subsidies and argued that a substantially progressive labor income tax and a positive education subsidy constitute part of the optimal fiscal constitution once household college attendance decisions are endogenous. In our thought experiment we took the tax on capital income as exogenously given. Ongoing and future work will determine whether these conclusions remain robust once the government chooses not only the progressivity of the labor income tax, but also the optimal mix between capital and labor income taxes.

Furthermore, we made several important auxiliary assumptions that deserve further scrutiny. On the calibration side, the assumption of a perfect substitutability between skilled and unskilled labor implies that an expansion of the stock of college-educated workers has no impact on their relative productivity and thus wages. As such, this parametric assumption gives education subsidies a potentially (too) important role in raising aggregate labor productivity and thus societal welfare.

Finally we determined the optimal tax policy as one which maximizes utilitarian social welfare among households currently alive.[34] We also documented that the optimal tax reform is not uniformly preferred to the status quo, implying that other social welfare functions imply alternative tax policies as optimal. We leave for future work a detailed analysis which elements of our optimal fiscal constitution remains intact if these alternative societal rankings of individual household preferences are considered.

[34] Although Utilitarian social welfare is commonly used in the literature, it is of course but one choice for the social welfare function. For alternative criteria and their merits, see e.g. Saez and Stantcheva (2012) or Weinzierl (2012).

References

[1] Aiyagari, R. (1994), "Uninsured Idiosyncratic Risk and Aggregate Saving," *Quarterly Journal of Economics*, 109(3): 659-684.

[2] Alvarez, Y., J. Burbidge, T. Farrell, and L. Palmer (1992), "Optimal Taxation in a Life-Cycle Model," *Canadian Journal of Economics*, 25, 111-122.

[3] Atkeson, A., V. V. Chari, and P. Kehoe (1999), "Taxing Capital Income: A Bad Idea," *Federal Reserve Bank of Minneapolis Quarterly Review*, 23, 3-17.

[4] Auerbach, A. and L. Kotlikoff (1987), *Dynamic Fiscal Policy*, Cambridge University Press.

[5] Babcock, P. (2009), "Real Costs of Nominal Grade Inflation? New Evidence from Student Course Evaluations," *Working paper*, UC Santa Barbara.

[6] Bakis, O., B. Kaymak and M. Poschke (2011), "On the Optimality of Progressive Income Redistribution," Working paper.

[7] Benabou, R. (2002), "Tax and Education Policy in a Heterogeneous-Agent Economy: What Levels of Redistribution Maximize Growth and Efficiency?" *Econometrica*, 70, 481-517.

[8] Bewley, T. (1986), "Stationary Monetary Equilibrium with a Continuum of Independently Fluctuating Consumers," in Werner Hildenbrand and Andreu Mas-Colell (eds.) *Contributions to Mathematical Economics in Honor of Gerard Debreu*, North-Holland, 79-102.

[9] Bohacek, R. and M. Kapicka (2008), "Optimal Human Capital Policies," *Journal of Monetary Economics*, 55, 1-16.

[10] Bohacek, R. and M. Kapicka (2010), "A Quantitative Analysis of Educational Reforms in a Dynastic Framework," Working Paper.

[11] Borjas, G. (2003), ""The Labor Demand Curve is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market," Quarterly Journal of Economics, 118, 1335-1374

[12] Bovenberg, L., and B. Jacobs (2005), "Redistribution and Education Subsidies are Siamese Twins," *Journal of Public Economics*, 89, 2005–2035.

[13] Brown, M., M. Mazzocco, J. Scholz and A. Seshadri (2006), "Tied Transfers," Working Paper.

[14] Castañeda, A., J. Díaz-Giménez and V. Ríos-Rull (1999), "Earnings and Wealth Inequality and Income Taxation: Quantifying the Trade-offs of Switching to a Proportional Income Tax in the U.S." Working Paper.

[15] Chamley, C. (1986), "Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives," *Econometrica*, 54, 607-622.

[16] Chari, V. V., and P. Kehoe (1999), "Optimal Fiscal and Monetary Policy," In *Handbook of Macroeconomics*, ed. John B. Taylor and Michael Woodford, 1671-1745. Amsterdam: North Holland.

[17] Chari, V. V., and P. Kehoe (2006), "Modern Macroeconomics in Practice: How Theory Is Shaping Policy," *Journal of Economic Perspectives,* 20, 3-28.

[18] Conesa, J. and D. Krueger (2006), "On the Optimal Progressivity of the Income Tax Code," *Journal of Monetary Economics*, 53, 1425-1450.

[19] Conesa, J., S. Kitao and D. Krueger (2009), "Taxing Capital: Not a Bad Idea after All!," *American Economic Review*, Vol 99, 25-48.

[20] Cunha, F., J. Heckman, L. Lochner and D. Masterov (2006), Interpreting the Evidence on Life Cycle Skill Formation, Handbook of the Economics of Education, Vol. 1, 697-812.

[21] Caucutt, E., K. Kumar and S. Imrohoroglu (2006), "Does the Progressivity of Income Taxes Matter for Human Capital and Growth?," *Journal of Public Economic Theory*, 8, 95-118.

[22] Erosa, A. and M. Gervais (2002), "Optimal Taxation in Life Cycle Economies," *Journal of Economic Theory*, 105, 338-369.

[23] Fehr, H. and F. Kindermann (2012), "Optimal Taxation with Current and Future Cohorts," *CESifo Working Paper* 3973.

[24] S. Findeisen and D. Sachs (2011), "Education and Optimal Dynamic Taxation," Working Paper.

[25] Fukushima, K. (2010), "Quantifying the Welfare Gains from Flexible Dynamic Income Tax Systems," *Working paper*, University of Wisconsin.

[26] Gallipoli, G., C. Meghir and G. Violante (2011), "Equilibrium Effects of Education Policies: A Quantitative Evaluation," Working Paper.

[27] Garriga, C. (2003), "Optimal Fiscal Policy in Overlapping Generations Models," Working Paper.

[28] Gale, W. and J. Scholz (1994), "Intergenerational Transfers and the Accumulation of Wealth," *Journal of Economic Perspectives*, *American Economic Association*, 8, 145-60.

[29] Gervais, M. (2009), "On the Optimality of Age-Dependent Taxes and the Progressive U.S. Tax System," Working Paper.

[30] Greenwood, J, Z. Hercowitz, and G. Hoffman (1988), "Investment, Capacity Utilization, and the Real Business Cycle," *American Economic Review*, 78, 402-417.

[31] Guvenen, F., B. Kuruscu and S. Ozkan (2011), "Taxation of Human Capital and Wage Inequality: A Cross-Country Analysis," Working Paper.

[32] Heckman, J., L. Lochner and P. Todd (2005), "Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond," NBER Working Paper 11544.

[33] Heckman, J., L. Lochner and C. Taber (1998), "Tax Policy and Human-Capital Formation," *American Economic Review*, 88, 293-297.

[34] Heckman, J., L. Lochner and C. Taber (1999), "Human Capital Formation and General Equilibrium Treatment Effects: A Study of Tax and Tuition Policy," *Fiscal Studies*, 20, 25-40.

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

[36] Holmlund, H., M. Lindahl and E. Plug (2011), The Causal Effect of Parents' Schooling on Children's Schooling: A Comparison of Estimation Methods," *Journal of Economic Literature*, 49, 615-651.

[37] Holter, H. (2011), "Accounting for Cross-Country Differences in Intergenerational Earnings Persistence: The Impact of Taxation and Public Education Expenditure," Working Paper.

[38] Hubbard, G. and K. Judd (1986), "Liquidity Constraints, Fiscal Policy, and Consumption," *Brookings Papers on Economic Activity*, 1986, 1-50.

[39] Huggett, M. (1993), "The Risk-Free Rate in Heterogeneous-Agent Incomplete-Insurance Economies," *Journal of Economic Dynamics and Control*, 17, 953-969.

[40] Huggett, M. (1997), "The One-Sector Growth Model with Idiosyncratic Shocks: Steady States and Dynamics," *Journal of Monetary Economics*, 39, 385-403.

[41] Ionescu, F. and N. Simpson (2011), "Credit Scores and College Investment: Implications for Student Loan Policies," *Working paper*, Colgate University.

[42] Jacobs, B. and L. Bovenberg (2010), "Human Capital and Optimal Positive Taxation of Capital Income," *International Tax and Public Finance*, 17, 451-478.

[43] Judd, K. (1985), "Redistributive Taxation in a Simple Perfect Foresight Model," *Journal of Public Economics*, 28, 59-83.

[44] Judd, K. and C. Su (2006), "Optimal Income Taxation with Multidimensional Taxpayer Types," Working Paper.

[45] Kapicka, M. (2011), "The Dynamics of Optimal Taxation when Human Capital is Endogenous," Working Paper.

[46] Karabarbounis, M. (2011), "Heterogeneity in Labor Supply Elasticity and Optimal Taxation," Working Paper.

[47] Katz, L. and K. Murphy (1992), "Changes in Relative Wages, 1963-1987: Supply and Demand Factors," *Quarterly Journal of Economics*, 107, 35-78.

[48] Kindermann, F. (2012), "Welfare Effects of Privatizing Public Education When Human Capital Investments Are Risky," *Journal of Human Capital*, 6, 87-123.

[49] Lochner, L. and A. Monge (2011), "Credit Constraints in Education," forthcoming, *Annual Review of Economics*.

[50] Mendoza, E., A. Razin and L. Tesar (1994), "Effective Tax Rates in Macroeconomics: Cross-Country Estimates on Factor Income and Consumption," *Journal of Monetary Economics,* 34, 297-323.

[51] Mirrless, J. (1971), "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 38, 175-208.

[52] Piketty, T. and E. Saez (2012), "Optimal Labor Income Taxation," *NBER Working Paper* 18521.

[53] Rupert, P. and G. Zanella (2010), "Revisiting Wage, Earnings and Hours Profiles," *Working paper*, UC Santa Barbara.

[54] Saez, E. and S. Stantcheva (2012), "Optimal Tax Theory with Endogenous Social Marginal Welfare Weights," Working paper, UC Berkeley.

[55] Stinebrickner, T. and R. Stinebrickner (2008), "The Causal Effect of Studying on Academic Performance," *Frontiers in Economic Analysis & Policy, Berkeley Electronic Press*, 8(1), 1868-1868.

[56] Weinzierl, M. (2011) "The Surprising Power of Age-Dependent Taxes," *Review of Economic Studies*, 78, 1-29.

[57] Weinzierl, M. (2012) "The Promise of Positive Optimal Taxation: A Generalized Theory Calibrated to Survey Evidence on Normative Preferences Explains Puzzling Features of Policy," *NBER Working Paper* 18599.

A Theoretical Appendix

A.1 Proof of Proposition 6

Proof. Recall that $e^{CE}(\tau = \theta = 0) = e^{SP}$ and thus aggregate output net of education costs coincide in the unregulated equilibrium and the solution to the social planner problem. Consequently aggregate consumption is identical as well:

$$C^{CE}(\tau = 0, \theta = 0) = L(\tau = 0, \theta = 0) - \kappa w(1 - e^{CE}(\tau = \theta = 0)) = L^{SP} - \kappa w(1 - e^{SP}) = C^{SP}.$$

In the social optimum the utilitarian social planner equalizes lifetime utility across all household types (this follows directly from the first order conditions of the planning problem) and therefore

$$c_c^{SP}(e) = c_n^{SP} + \mu \frac{\left(l_c^{SP}(e)\right)^{1+\frac{1}{\psi}}}{1+\frac{1}{\psi}} - \mu \frac{\left(l_n^{SP}\right)^{1+\frac{1}{\psi}}}{1+\frac{1}{\psi}} \tag{45}$$

for all e, and thus especially for $e = e^{SP}$. In the competitive equilibrium, *at the education threshold (and only there)* we have

$$c_c^{CE}(e^{CE}) = c_n^{CE} + \mu \frac{\left(l_c^{CE}(e^{CE})\right)^{1+\frac{1}{\psi}}}{1+\frac{1}{\psi}} - \mu \frac{\left(l_n^{CE}\right)^{1+\frac{1}{\psi}}}{1+\frac{1}{\psi}}$$

But for $\tau = d = \theta = 0$ we have $e^{CE} = e^{SP}$ and thus

$$c_c^{CE}(e^{CE}) - c_n^{CE} = c_c^{SP}(e^{SP}) - c_n^{SP} \tag{46}$$

that is, the consumption premium of the marginal type going to college is the same in the unregulated equilibrium and in the social planner problem. Now we show that for all types $e \geq e^{SP}$ we have $\frac{\partial c_c^{CE}(e)}{\partial e} > \frac{\partial c_c^{SP}(e)}{\partial e}$. For this we note that from (45) and (10), evaluated at $\tau = 0$,

$$\begin{aligned}
\frac{\partial c_c^{SP}(e)}{\partial e} &= \frac{(1+\psi)pw^{1+\psi}\left[(1+pe)\right]^{\psi}}{\left(1+\frac{1}{\psi}\right)\mu^{\psi}} \\
\frac{\partial c_c^{CE}(e)}{\partial e} &= \frac{(1+\psi)pw^{1+\psi}(1+pe)^{\psi}}{\mu^{\psi}}
\end{aligned}$$

Thus as long as $\psi < \infty$ we have $\frac{\partial c_c^{CE}(e)}{\partial e} > \frac{\partial c_c^{SP}(e)}{\partial e}$ for all $e \geq e^{CE}(\tau = \theta = 0) = e^{SP}$. But since $C^{CE}(\tau = 0, \theta = 0) = C^{SP}$ it then follows from (46) that

$$\begin{aligned}
c_n^{CE} &< c_n^{SP} \\
c_c^{CE}(e^{SP}) &< c_c^{SP}(e^{SP}).
\end{aligned}$$

Otherwise we would have $c^{CE}(e) > c^{SP}(e)$ for all e, which violates the resource constraint. ∎

A.2 Analytical Characterization of Optimal Ramsey Policy

Before turning to the first order conditions with respect to the policy variables we note that

$$\begin{aligned}
\frac{\partial e^{CE}(\tau,\theta)}{\partial \tau} &= \frac{(1-\theta)(1+\psi)\mu^{\psi}\kappa}{p(1-\tau)^{2+\psi}w^{\psi}}\left(\frac{(1-\theta)\mu^{\psi}(1+\psi)\kappa}{(1-\tau)^{1+\psi}w^{\psi}}+1\right)^{\frac{-\psi}{1+\psi}} \\
\frac{\partial e^{CE}(\tau,\theta)}{\partial \theta} &= -\frac{\mu^{\psi}\kappa}{p(1-\tau)^{1+\psi}w^{\psi}}\left(\frac{(1-\theta)\mu^{\psi}(1+\psi)\kappa}{(1-\tau)^{1+\psi}w^{\psi}}+1\right)^{\frac{-\psi}{1+\psi}} = -\frac{\partial e^{CE}(\tau,\theta)}{\partial \tau}\frac{(1-\tau)}{(1-\theta)(1+\psi)} \\
\frac{\partial L(\tau,\theta)}{\partial \tau} &= -\left(\frac{1-\tau}{\mu}\right)^{\psi} w^{1+\psi}\frac{\partial e^{CE}(\tau,\theta)}{\partial \tau}\left((1+pe^{CE}(\tau,\theta))^{1+\psi}-1\right) - \frac{\psi L(\tau,\theta)}{1-\tau} \\
\frac{\partial L(\tau,\theta)}{\partial \theta} &= -\left(\frac{1-\tau}{\mu}\right)^{\psi} w^{1+\psi}\frac{\partial e^{CE}(\tau,\theta)}{\partial \theta}\left((1+pe^{CE}(\tau,\theta))^{1+\psi}-1\right)
\end{aligned}$$

Attaching Lagrange multiplier λ to the government budget constraint we obtain as first order conditions with respect to d, θ and τ, respectively

$$\begin{aligned}
\lambda &= \frac{e^{CE}(\tau,\theta)}{nc_n} + \int_{e^{CE}(\tau,\theta)}^{1}\frac{de}{nc_c(e)} = \int_0^1 \frac{de}{nc(e)} \\
&\lambda\left[(1-e^{CE}(\tau,\theta)) - \theta\frac{\partial e^{CE}(\tau,\theta)}{\partial \theta} + \tau\frac{\partial e^{CE}(\tau,\theta)}{\partial \theta}\left(\left(\frac{1-\tau}{\mu}\right)^{\psi} w^{\psi}/\kappa\right)\left[(1+pe^{CE}(\tau,\theta))^{1+\psi}-1\right]\right] \\
&= \int_{e^{CE}(\tau,\theta)}^{1}\frac{de}{nc_c(e)} \\
&\lambda\left[\left(\frac{1-\tau(1+\psi)}{1-\tau}\right)L(\tau,\theta) - \frac{\partial e^{CE}(\tau,\theta)}{\partial \tau}\left(\tau\left(\frac{1-\tau}{\mu}\right)^{\psi} w^{1+\psi}\left((1+pe^{CE}(\tau,\theta))^{1+\psi}-1\right) - \theta\kappa w\right)\right] \\
&= \frac{w^{1+\psi}(1-\tau)^{\psi}}{\mu^{\psi}}\left(\frac{e^{CE}(\tau,\theta)}{nc_n} + \int_{e^{CE}(\tau,\theta)}^{1}\frac{[(1+pe)]^{1+\psi}\,de}{nc_c(e)}\right)
\end{aligned}$$

Here

$$\begin{aligned}
nc_n &= \frac{[(1-\tau)w]^{1+\psi}}{(1+\psi)\mu^{\psi}} + dw \\
nc_c(e) &= \frac{[(1-\tau)(1+pe)w]^{1+\psi}}{(1+\psi)\mu^{\psi}} + dw + \kappa(1-\theta)w
\end{aligned}$$

A given dollar of tax revenue can either be spent on lump-sum income subsidies or education subsidies. Combing the first two equations yields

$$\frac{e^{CE}(\tau,\theta)}{nc_n} = \lambda\left[\begin{array}{c} e^{CE}(\tau,\theta)) + \theta\frac{\partial e^{CE}(\tau,\theta)}{\partial\theta} - \tau\frac{\partial e^{CE}(\tau,\theta)}{\partial\theta}\left(\left(\frac{1-\tau}{\mu}\right)^{\psi} w^{\psi}/\kappa\right) * \\ \left[(1+pe^{CE}(\tau,\theta))^{1+\psi}-1\right] \end{array}\right]$$

A sharper theoretical characterization of the optimal policy mix is **TBC**

B Calibration Appendix

B.1 Details of the Calibration of Social Security Benefits

The U.S. system is characterized by an indexation to "average indexed monthly earnings" (AIME). This sum the 35 years of working life with the highest individual earnings relative to average earnings. Social security benefits are then calculated as a concave function of AIME.

We approximate this system as follows. First, we define AIME of a type $(\hat{e},\hat{s})$ household that retires in year t_r as

$$\bar{y}_{t_r}(\hat{e},\hat{s}) = \frac{\sum_{j=j_s}^{j_r-1} w_{t-(j_r-1-j),\hat{s}}\epsilon_{j,\hat{s}}\gamma_{\hat{s}}(\hat{e})}{\sum_{e,s}\sum_{j=j_s}^{j_r-1} w_{t-(j_r-1-j),s}\epsilon_{j,s}\gamma_s(e)} \tag{47}$$

as the sum of yearly wages, averaged across all η, for the cohort entering into retirement in year t_r, normalized such that $\sum \bar{y}_{t_r}(e,s,k) = 1$. For simplicity, we start the sum in (47) after college completion and thereby do not account for the lower wages of college attendees while in college.

The primary insurance amount (PIA) of the cohort entering into retirement in period t_r, $pia_{t_r}(e,s)$, is then computed as

$$pia_{t_r}(e,s) = \begin{cases} s_1\bar{y}_{t_r}(e,s,k) & \text{for } \bar{y}_{t_r}(e,s,k) < b_1 \\ s_1 b_1 + s_2\left(\bar{y}_{t_r}(e,s,k) - b_1\right) & \text{for } b_1 \leq \bar{y}_{t_r}(e,s,k) < b_2 \\ s_1 b_1 + s_2\left(b_2 - b_1\right) + s_3\left(\bar{y}_{t_r}(e,s,k) - b_2\right)) & \text{for } b_2 \leq \bar{y}_{t_r}(e,s,k) < b_3 \\ s_1 b_1 + s_2\left(b_2 - pb_1\right) + s_3\left(b_3 - b_2\right) & \text{for } \bar{y}_{t_r}(e,s,k) \geq b_3 \end{cases}$$

for slopes $s_1 = 0.9$, $s_2 = 0.32$, $s_3 = 0.15$ and bend points $b_1 = 0.24$, $b_2 = 1.35$ and $b_3 = 1.99$.

Pensions for all pensioners of age $j \geq j_r$ in period t are then given by

$$p_{t,j}(e,s) = \rho_t w_t(1-\tau_{ss,t}) \cdot pia_{t_r}(e,s)$$

where ρ, the net pension benefit level, governs average pensions.

Budget balance requires that

$$\tau_{ss,t}\sum_s w_{t,s}L_{t,s} = \sum_{j=j_r}^{J} N_{t,j}\int p_{t,j}(e,s)d\Phi_{t,j}$$

and thus

$$\tau_{ss,t}\sum_s w_{t,s}L_{t,s} = \rho_t w_t(1-\tau_{ss,t})\sum_{j=j_r}^{J} N_{t,j}\int pia_{t_r}(e,s)d\Phi_{t,j}$$

C Computational Appendix

[To be completed and re-written]