Redistributive Taxation in a Partial Insurance Economy

Jonathan Heathcote
Federal Reserve Bank of Minneapolis and CEPR

Kjetil Storesletten
University of Oslo and CEPR

Giovanni L. Violante
New York University, CEPR, and NBER

THIS VERSION: OCTOBER 2013

Abstract

We study optimal taxation in a partial insurance economy

*We thank ... The opinions expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.
1 Introduction

The rest of the paper is organized as follows. Section 2 presents our tax function and discusses its properties. Section 3 describes the economic environment. Section 4 contains a characterization of the equilibrium allocations in closed form. Section 5 solves analytically for social welfare as a function of the fiscal parameters chosen by the government (progressivity and size of the publicly provided good) and as a function of all other structural parameters of the model. Section 6 calibrates the model and explores the quantitative implications of the theory for the optimal degree of progressivity. Section 7 contains three extensions: transitional dynamics, the analysis of a progressive consumption tax, and a political-economic analysis. Section 8 concludes. All proofs are in the Appendix.

2 The tax function

We study the optimal degree of progressivity within a particular class of tax/transfers policies that has a tradition in public finance, starting from Feldstein (1969). More recently, Benabou (2000, 2002) has introduced this class of policies into dynamic macroeconomic models with heterogeneous agents. Under this scheme, tax revenue as a function of individual income $y_i$ (the only variable upon which taxes can be conditioned) is given by

$$T(y_i) = y_i - \lambda y_i^{1-\tau}. \quad (1)$$

The retention function in equation (1) implies the following mapping between disposable (post-government) earnings $\bar{y}_i$ and pre-government earnings $y_i$:

$$\bar{y}_i = \lambda y_i^{1-\tau}. \quad (2)$$

This tax scheme is indexed by two parameters only, $\tau$ and $\lambda$. Equation (2) reveals that $(1 - \tau)$ measures the elasticity of post-tax to pre-tax income. Thus, $\tau$ determines the de-

---

\(^1\)Musgrave (1979) refers to $1 - \tau$ as the coefficient of residual income progression. As discussed in Benabou (2000), it has been proved that the post-tax income distribution induced by a fiscal scheme Lorenz-dominates (i.e., displays less inequality than) the one induced by another fiscal scheme (for all pre-tax income distributions), if and only if the first scheme's progression coefficient $(1 - \tau)$ is everywhere smaller. See, e.g., Kakwani (1977). In this sense, $\tau$ is a very natural and parsimonious measure of progressivity.
gree of progressivity of the tax system and is the key object of interest in our analysis: our characterization of the optimal degree of progressivity will be developed in terms of $\tau$.

This tax scheme encompasses both progressive and regressive schemes. A common definition of progressivity ( regressivity) of a tax scheme requires that the ratio of marginal to average tax rate be greater ( smaller) than one for every level of income $y_i$. Within our class, we have:

$$\frac{T'(y_i)}{T(y_i)/y_i} = \frac{1 - \lambda (1 - \tau) y_i^{-\tau}}{1 - \lambda y_i^{-\tau}}. \quad (3)$$

The case $\tau = 0$ implies a ratio exactly equal to one and corresponds to a flat tax of size $1 - \lambda$. When $\tau > 0$, the ratio in equation (3) is larger than one and the tax system is progressive. A limiting case is the “confiscatory rate” $\tau = 1$. This is the largest degree of progressivity admissible which implies post-government earnings equal to $\lambda$ for every household and, hence, full redistribution. Conversely, when $\tau < 0$, the ratio in (3) is less than one and the tax system is regressive.

Given $\tau$, the second parameter, $\lambda$, determines the “break-even income level”. At income level $y^0 = \lambda^{1/\tau} > 0$, the average tax rate is zero and the marginal tax rate is exactly $\tau$. Therefore, given $\tau$, $\lambda$ shifts up and down the tax function and determines, loosely, the average level of taxation in the economy. The existence of this strictly positive break-even income level implies that at every income level below (above) $y^0$, if the system is progressive (regressive), the average tax rate is negative and households obtain a net transfer from the government. Therefore, this schedule is better seen as a tax and transfer schedule, a property that has implications for the empirical measurement of $\tau$.

From the perspective of optimal taxation, two restrictions implicit in $T(y_i)$ are that (i) it does not allow for lump-sum transfers in cash, since $T(0) = 0$; and (ii) it is either globally convex in income, if $\tau > 0$, or globally concave, if $\tau < 0$. As a result, marginal tax rates are monotonic in income. The same restriction applies to the average tax rate. [GV: we need more here. Say that Mirlees schemes use heavily transfer? say that we allow for a universal transfer in goods? say that in the US we don’t see lump-sum transfers but we do see transfers in goods and services? say that Mirlees optimal tax rates can be non-monotonic and use heavily transfer?].
Figure 1: Representation of the actual US tax/transfer system through our tax/transfer function. The estimated value of $\tau_{US}$ is 0.151.

While its functional form is inevitably restrictive, it has two important virtues. First and foremost, it offers a remarkably good representation of the actual tax/transfer scheme in the US. Second, it implies a log-linear relation between pre- and post-government log income that is crucial to preserving tractability.

To demonstrate the empirical fit of this functional form, we use data from the Panel Study of Income Dynamics (PSID) for survey years 2000, 2002, 2004, and 2006. We restrict attention to all households with age 25-60 where either the head or the spouse work at least 260 hours per year (one quarter part-time).\footnote{GV: note that we miss the very top of the income distribution, the emphasis of much of the Mirrlees literature} Pre-government household income includes labor earnings, private transfers (transfers include alimony, child support, help from relatives, miscellaneous transfers, private retirement income, annuities and other retirement income), and income from interests, dividends and rents. Post-government income equals pre-government income minus federal and state income taxes computed based on the NBER’s TAXSIM program, plus public transfers (AFDC/TANF, SSI and other welfare receipts, social security benefits, unemployment benefits, worker’s compensation and veterans’ pensions).

We estimate $\tau_{US}$ by least squares using equation (2) in log form. The point estimate is $\tau_{US} = 0.151$ ($S.E. = 0.003$). The simple model fits the empirical relationship between pre-
and post-government income distributions remarkably well \((R^2 = 0.96)\) as demonstrated in Figure 1(a), where we collapse our 13,721 observations into 50 quantiles (each containing 2\% of total observations).\(^3\) Figure 1(b) plots the average and marginal tax rates, as a function of income (with mean income normalized to 1) implied by our tax/transfer scheme evaluated at \(r^{US}\).

## 3 Economic environment

[JH maybe a brief introductory model description, if not in the text] We describe the economy in steady-state and omit time subscripts from all individual and aggregate variables.

**Demographics:** We adopt the Yaari “perpetual youth” structure. At every age \(a\), an agent survives into the next period with constant probability \(\delta < 1\), and a cohort of newborn agents of size \((1 - \delta)\) enters the economy. Thus, agents of age \(a\) have measure \((1 - \delta) \delta^a\) and the population is always stationary and has measure one. We index agents with \(i \in [0, 1]\).

**Lifecycle:** The life of every individual \(i\) in the model starts with an initial investment in human capital. After choosing her skill level \(s_i\) at age \(a = 0\), the individual enters the labor market and starts facing random fluctuations in her labor productivity \(z_i\). Every period she supplies hours of work \(h_i \geq 0\) to the market, and she consumes a private consumption good \(c_i > 0\) and a publicly provided good \(G > 0\).\(^4\)

**Technology:** Output \(Y\) is a CES aggregator over the continuum of skill types \(s \in [0, \infty)\)

\[
Y = \left\{ \int_0^\infty [N(s) \cdot m(s)]^{\frac{\theta - 1}{\sigma}} ds \right\}^{\frac{\sigma}{\theta - 1}}
\]

where \(\theta \geq 1\) is the elasticity of substitution across skill types, \(N(s)\) denotes average effective hours worked by skill type \(s\), and \(m(s)\) is the density of individuals with skill level \(s\).

---

\(^3\)The coordinates of each circle in the figure are the mean of the corresponding quantile of the pre-government income distribution (X axis), and the mean post-government income across the observations in that same quantile (Y axis).

\(^4\)There are two possible interpretations for \(G\). The first is that it is a pure public good, like national defense. The second is that it an excludable and rival good produced by the government and distributed uniformly across households, such as public health-care. Under the second interpretation, \(G\) is analogous to a universal lump-sum transfer in kind.
The aggregate resource constraint of the economy is

\[ Y = \int_0^1 c_i di + G \]  

and hence, the rate of transformation between private and public consumption is one.

**Preferences:** Preferences over private consumption \((c)\), hours worked \((h)\), publicly-provided goods \((G)\), and skill-investment \((s)\) effort for individual \(i\) are given by

\[ U_i = v_i(s_i) + (1 - \beta \delta) \mathbb{E}_0 \sum_{a=0}^{\infty} (\beta \delta)^a u_i(c_{ia}, h_{ia}, G), \]  

where \(\beta < 1\) is the discount factor, common to all individuals, and the expectation is taken over future histories of idiosyncratic productivity shocks, whose process is described below.

The disutility of the initial skill-investment \(s_i \geq 0\) takes the quadratic form

\[ v_i(s_i) = -\frac{1}{\kappa_i} \frac{s_i^2}{2\mu} \]  

where \(\kappa_i \geq 0\) is a parameter, heterogeneous across individuals, which determines the utility cost of acquiring skills: the larger \(\kappa_i\) the smaller the cost, so one can think of \(\kappa_i\) as innate learning ability. We assume that \(\kappa_i \sim \text{Exp}(\eta)\), an exponential distribution with parameter \(\eta\). The parameter \(\mu\) is a scaling constant. Without loss of generality, we set \(\mu = \eta\). this simplifies the expressions but does not affect the welfare analysis (\(\mu\) simply rescales the level of \(s\)). As we demonstrate below, the combination of quadratic skill investment cost and exponential distribution yields a Pareto-tail in the income distribution, a desirable feature of the model.

The period utility function \(u_i\) is specified as

\[ u_i(c_{ia}, h_{ia}, G) = \log c_{ia} - \exp((1 + \sigma) \varphi_i) h_{ia}^{1+\sigma} \frac{1}{1+\sigma} + \chi \log G \]  

where \(\exp((1 + \sigma) \varphi_i)\) measures the disutility of work effort, also heterogeneous across agents.\(^5\) The individual-specific parameter \(\varphi_i\) is normally distributed as \(\varphi_i \sim N\left(\frac{v_\varphi}{2}, v_\varphi\right)\),

where \(v_\varphi\) denotes its cross-sectional variance. We assume that \(\kappa_i\) and \(\varphi_i\) are uncorrelated.

\(^5\)The term \((1 + \sigma)\) multiplying \(\varphi_i\) is an innocuous and convenient normalization that simplifies the hours allocation.
The parameter $\chi \geq 0$, common across all agents, measures the taste for the publicly provided good $G$ relative to private consumption. Finally, $\sigma > 0$ measures aversion to hours fluctuations, and $1/\sigma$ is the Frisch elasticity of labor supply. It is also useful to define the tax-modified Frisch elasticity

$$\frac{1}{\tilde{\sigma}} = \frac{1 - \tau}{\sigma + \tau}$$

which measures the after-tax elasticity of hours worked to a transitory wage shock. The model can be generalized to allow for a fraction of public expenditure to be wasted, so that only a fraction $\gamma \in (0, 1]$ of output consumed by the government is delivered to consumers. In that case, utility from public expenditure would be given by $\chi \log(\gamma G)$. It is immediate that this amounts to adding an irrelevant constant to preferences.

**Labor productivity and earnings:** Individual efficiency units of labor $z_{ia}$ obey the statistical model

$$\log z_{ia} = \alpha_{ia} + \varepsilon_{ia},$$

a sum of two (cross-sectionally and longitudinally) orthogonal components, $\alpha_{ia}$ and $\varepsilon_{ia}$. The first component $\alpha_{ia}$ follows the unit root process $\alpha_{ia} = \alpha_{i,a-1} + \omega_{it}$, with innovation $\omega_{ia} \sim i.i.d. N \left(-\frac{v_\omega}{2}, v_\omega\right)$ and with initial condition $\alpha_{i0} = 0, \forall i$. The second one, $\varepsilon_{ia}$, is an i.i.d. shock distributed as $N \left(-\frac{v_\varepsilon}{2}, v_\varepsilon\right)$. This permanent-transitory error-component model for individual labor productivity has a long tradition in labor economics (for a survey, see Meghir and Pistaferri, 2011). A law of large numbers (e.g., Uhlig, 1996) can be applied so that individual-level shocks induce no aggregate uncertainty in the economy as a whole.

Individual earnings $y_{ia}$ are, therefore, the product of three components:

$$y_{ia} = p(s_i) \times \exp(\alpha_{ia} + \varepsilon_{ia}) \times h_{ia}.$$  

The first component $p(s_i)$ is the equilibrium price for the efficiency units supplied by an individual of skills $s_i$; the second component is the individual stochastic labor efficiency; the

---

6Setting the dispersion in initial conditions $\alpha_{i0}$ to zero is not restrictive because newborn agents enter the economy with heterogeneous skill levels $s_i$ (also a fixed individual effect) which reflect the dispersion in innate learning ability $\kappa_i$.

7The empirical autocovariance function for individual wages displays a sharp decline at the first lag, indicating the presence of a transitory component in wages. At the same time, within-cohort wage dispersion increases approximately linearly with age, suggesting the presence of permanent shocks.
third component is the number of hours worked by the individual. Equation (11) shows the key determinants of individual earnings: (i) skills accumulated before labor market entry, in turn a consequence of innate learning ability \( \kappa \); (ii) fortune in labor market outcomes determined by the realization of shocks; and (iii) work effort, determined by the degree of diligence \( \varphi \), a fixed individual trait. Since taxes are only a function of earnings \( y_i \), the optimal degree of progressivity must necessarily trade-off these three separate forces.

**Financial assets:** We adopt the “partial-insurance” structure developed in Heathcote, Storesletten, and Violante (2013) and postulate that there are only two types of financial assets in the economy. First, a non state-contingent bond \( b \) with price \( q \). Second, a full set of insurance claims against the \( \varepsilon \) shock. Let denoted by \( B(E) \) and \( Q(E) \) the quantity and the price, respectively, of the claim that pays if \( \varepsilon \in E \subseteq E \). These assumptions on spanning imply that the \( \varepsilon \) shocks are fully insurable, whereas the \( \alpha \) shocks can be, potentially, smoothed only by borrowing and saving through the risk-free bond. Our model encompasses several market structures. First, when \( v_\omega = 0 \), the economy displays full insurance. When \( v_\varepsilon = 0 \), it is a bond-economy, as in Huggett (1993). In general, when \( v_\omega > 0 \) and \( v_\varepsilon > 0 \), ours is a partial insurance economy, i.e., an economy that offers insurance opportunities in between a bond economy and complete markets.\(^8\)

Finally, for convenience, we assume that there exist actuarially fair annuities against survival risk which pay \( \delta^{-1} \) units of consumption if the agent survives to the next period, and zero otherwise. All assets in the economy are in zero net supply, and newborn agents start without any initial asset endowment.

**Markets:** The final consumption good, all types of labor services, and all financial assets are traded in competitive markets. The publicly-provided good \( G \) cannot be purchased privately. The price of the final good is the numeraire of the economy.

**Government:** The government runs the tax/transfer scheme described in Section 2 and provides each household with an amount of public goods/services equal to \( G \). Without loss of

---

\(^8\)The complete-markets assumption with respect to \( \varepsilon \) implies that one could allow for a much more general statistical process for \( \varepsilon \) than i.i.d., as done in Heathcote, Storesletten and Violante (2013). As we show below, all that matters in our analysis of optimal taxation is the cross-sectional variance of insurable wage risk which can be estimated independently of the time-series process for \( \varepsilon \). Therefore, to keep the exposition simpler, we maintain the i.i.d. assumption.
generality, we assume that government expenditures are set as a fraction \( g \) of aggregate output, i.e. \( G = gY \). Since we abstract from public debt, the government budget constraint holds period by period and reads as

\[
G = \int_{0}^{1} [y_i - \lambda y_i^{1-\tau}] \, di. \tag{12}
\]

We let the government choose the pair \((g, \tau)\), with \( \lambda \) being residually determined by equation (12).

### 3.1 Agent’s problem

At age \( a = 0 \), the agent begins by choosing her skill level. Combining equations (6) and (7), it is immediate that the first-order necessary and sufficient condition for the skill choice \( s \) is

\[
\frac{1}{\kappa_i \eta} \frac{s}{\partial s} = (1 - \beta \delta) \mathbb{E}_0 \sum_{a=0}^{\infty} (\beta \delta)^a \frac{\partial u_i (c_{ia}, h_{ia}, G)}{\partial s} \tag{13}
\]

stating that the marginal disutility of skill investment for an individual with learning ability \( \kappa_i \) must equal the discounted present value of expected benefits from the skill investment throughout her working lifetime.

The timing of the agent’s problem during her subsequent working life is as follows. At the beginning of every period \( a \), the innovation \( \omega_{ia} \) to the random walk shock \( \alpha_{ia} \) is realized. Then, the insurance markets against the \( \varepsilon \) shocks open and the individual buys insurance claims. Finally, \( \varepsilon_{ia} \) is realized and the individual supplies hours worked \( h_{ia} \) to the labor market, receives her wage payment, and chooses consumption \( c_{ia} \), and bond holdings \( b_{i,a+1} \) for next period.

Consider an individual who enters the period with bond holdings \( b_{ia} \). Her budget constraint in the middle of the period, when the insurance purchases are made, is:

\[
b_{ia} = \int_{E} Q (\varepsilon) B (\varepsilon) \, d\varepsilon \tag{14}
\]

whereas her budget constraint at the end of the period, after the realization of \( \varepsilon_{ia} \), reads as

\[
c_{ia} + \deltaqb_{i,a+1} = \lambda \left[ p (s_i) \exp (\alpha_{ia} + \varepsilon_{ia}) h_{ia} \right]^{1-\tau} + B (\varepsilon_{ia}), \tag{15}
\]

where the \( \delta \) pre-multiplying the bond price reflects the return on the annuity for survivors.
After her skill choice, at age $a = 0$ the problem for an agent is to choose a sequence of consumption and hours worked in order to maximize (6) subject to a sequence of budget constraints of the form (14) – (15), and the wage process in equation (10). In addition, agents face limits on borrowing that rule out Ponzi schemes, and non-negativity constraints on consumption and hours worked.

3.1.1 A special case: the representative agent problem

It is useful to solve for a special case of the agent’s problem. When $v_\varphi = v_\omega = v_\varepsilon = 0$ and $\theta = \infty$, there is no ex-ante heterogeneity in the taste for leisure nor uncertainty in efficiency units. There is no skill investment either, since skill levels are perfect substitutes in production. Since there is no ex-post heterogeneity in productivity, the economy collapses to a representative-agent model. In steady-state, the representative agent problem becomes

$$\max_{C,H} \quad \log C - \frac{H^{1+\sigma}}{1+\sigma} + \chi \log G$$

subject to

$$Y = H$$
$$C = \lambda Y^{1-\tau}$$

where the first constraint is the production technology (since $N = H$), and the second is the budget constraint. The consumption and hours allocations of the representative agent, who takes the fiscal variables $(\lambda, g, \tau)$ as given, are

$$\log C^{RA}(g, \tau) = \log \lambda^*(g, \tau) + \frac{1}{(1+\sigma)} \log(1-\tau)$$

$$\log H^{RA}(\tau) = \frac{1}{(1-\tau)(1+\sigma)} \log(1-\tau)$$

where $\lambda^*(g, \tau)$ is determined residually from the government budget constraint. Equation (18) shows that labor supply is decreasing in the degree of progressivity of the tax system $\tau$ because the marginal income tax faced by representative agent rises with $\tau$. Note that, with log utility, the average level of taxation ($\lambda$) has no impact on labor supply, which is only affected by the progressivity parameter $\tau$. This also explains why hours worked $H^{RA}$ are independent of the level of expenditures $g$. 9
4 Equilibrium

We now adopt a recursive formulation to define a stationary competitive equilibrium for our economy. The individual state vector for the end-of-period consumption/saving and labor supply decisions is given by \((\varphi, \alpha, \varepsilon, s, b)\). Since the choice over insurance claims is made before the realization of \(\varepsilon\), the state vector for this middle-of-the-period decision is \((\varphi, \alpha, s, b)\). Finally, since initial wealth \(b_0 = 0\), and \(\varepsilon\) is drawn only after the skill choice, the state vector for the skill accumulation decision at age \(a = 0\) reduces to the pair of fixed individual effects \((\kappa, \varphi)\). Note that, because of the perpetual youth structure, age is not a state variable.\(^9\)

Given \((g, \tau)\), a stationary recursive competitive equilibrium for our economy is a value \(\lambda\), asset prices \(Q(\cdot)\) and \(q\), skill prices \(p(s)\), decision rules \(s(\kappa, \varphi)\), \(c(\varphi, \alpha, \varepsilon, s, b)\), \(h(\varphi, \alpha, \varepsilon, s, b)\), \(b'(\varphi, \alpha, \varepsilon, s, b)\), and \(B(\varepsilon; \varphi, \alpha, s, b)\), and aggregate quantities \(N(s)\) such that:

1. Households solve the problem described in Section 3.1, and \(s(\kappa, \varphi)\), \(c(\varphi, \alpha, \varepsilon, s, b)\), \(h(\varphi, \alpha, \varepsilon, s, b)\), \(b'(\varphi, \alpha, \varepsilon, s, b)\), and \(B(\varepsilon; \varphi, \alpha, s, b)\) are the associated decision rules.

2. Labor markets for each skill type clear and \(p(s)\) is the value of the marginal product from an additional unit of effective hours of skill type \(s\):

\[
p(s) = \left( \frac{Y}{N(s) \cdot m(s)} \right)^{1/\theta}
\]

3. Asset markets clear: \(q\) is such that the net demand for the bond is zero, and the prices \(Q(\cdot)\) of insurance claims are actuarially fair.

4. The government budget is balanced, and \(\lambda\) is determined residually by equation (12).

Proposition 1 [competitive equilibrium]. There exists a competitive equilibrium characterized by no bond trading across individuals, i.e., \(b'(\varphi, \alpha, \varepsilon, s, b) = 0\) for all \((\varphi, \alpha, \varepsilon, s, b)\).

The interest rate \(r^* = -\log q\) which supports this equilibrium satisfies

\[
\rho - r^* = (1 - \tau) \left[ (1 - \tau) + 1 \right] \frac{\nu_{\omega}}{2}, \tag{19}
\]

\(^9\)This outcome also depends on the fact that \(\varepsilon\) is i.i.d. over time. In Heathcote, Storesletten and Violante (2013), we show that if \(\varepsilon\) has some persistence, then age remains a state.
where \( \rho = -\log \beta \) is the agents’ discount rate.

The proof for Proposition 1 is based on a “guess and verify” strategy. We first guess that the bond is not traded, and solve for the equilibrium consumption allocation. Next, we use the consumption allocation to construct the expected marginal rate of substitution and show that it is independent of any individual state. As a consequence, every agent assigns the same value to the bond. This result verifies that, if a bond were available it would not be traded in equilibrium.

Expression (19) reveals that, in equilibrium, the intertemporal dissaving motive (the left-hand side of this equation) determined by the gap between \( \rho \) and \( r \) exactly equals the precautionary saving motive (the right-hand side), which is increasing in the size of the uninsurable wage risk \( v_\omega \), and decreasing in the progressivity parameter \( \tau \) because, as \( \tau \) grows, the government provides more social insurance and the private precautionary demand for savings falls. At the equilibrium interest rate \( r^* \), these two saving motives exactly offset each other, and no agent wants to either borrow or lend. As discussed in Heathcote, Storesletten and Violante (2013), this result is a generalization of the insight in Constantineides and Duffie (1996).\(^{10}\)

Proposition 1 has two implications which are instrumental for the analytical tractability of our model. First, individual wealth is a redundant state variable: individuals start their life with zero wealth and stay with zero wealth forever.\(^{11}\) The power of this result lies in the fact that all remaining individual states are exogenous variables.\(^{12}\) Second, there is no self-insurance via non-contingent borrowing and lending, against \( \alpha \) shocks. In contrast, there is perfect insurance, by assumption, against \( \varepsilon \) shocks. Thus, in this equilibrium, there is a stark dichotomy between one type of risk which is uninsured, and another that is fully insured. In the aggregate, there is only partial insurance. In what follows, we will use the

\(^{10}\)Here, we further generalize the environment by endogenizing the wage through the skill investment decision and a technology featuring imperfect substitution across skill types. See Section 2.3.2 in Heathcote, Storesletten and Violante (2013) for a thorough discussion of the relationship with Constantineides and Duffie (1994).

\(^{11}\)Because bond holdings are zero, even the net position of insurance claims for every individual must equal to zero from the middle-of-the-period budget constraint (14).

\(^{12}\)The skill level is endogenously determined, but fixed after age zero, hence pre-determined with respect to consumption and labor supply decisions.
label “uninsurable” to denote the $\alpha$ shock (and its innovations $\omega$) and the label “insurable” to denote the $\varepsilon$ shock.

The payoff from analytical tractability is illustrated by the next two propositions which describe the equilibrium allocations and skill prices in closed form, and it culminates with Proposition 4 where we derive an analytical solution for the social welfare function. In what follows, we make the dependence of the equilibrium allocations and prices on $(g, \tau)$ explicit in preparation for our analysis of the optimal taxation problem.

**Proposition 2 [hours and consumption].** In equilibrium, the hours-worked allocation is given by

$$
\log h (\varphi, \varepsilon; \tau) = \log H^{RA} (\tau) - \frac{1}{\tilde{\sigma}(1 - \tau)} M (v_{\varepsilon}; \tau) - \varphi + \frac{1}{\tilde{\sigma}} \varepsilon
$$

(20)

where $H^{RA}$ are hours worked of the “representative agent” in equation (20) and $M (v_{\varepsilon}; \tau) = \frac{(1 - \tau)(1 - \tau(1 + \tilde{\sigma})) v_{\varepsilon}}{\tilde{\sigma}}$.

The consumption allocation is given by

$$
\log c (\varphi, \alpha, s; g, \tau) = \log C^{RA} (g, \tau) + M (v_{\varepsilon}; \tau) - (1 - \tau) \varphi + (1 - \tau) \log p (s; \tau) + (1 - \tau) \alpha
$$

(21)

where $C^{RA}$ is consumption of the “representative agent” in equation (21).

With log utility and zero wealth holdings, the income effect of uninsurable shocks $\alpha$ and skill level $s$ exactly cancels out with their substitution effect on labor supply, and hours worked are independent of $(s, \alpha)$. The hours worked allocation is composed of four terms. The first term is hours of the representative agent which, as explained, fall with progressivity. The second term captures the welfare-improving effect of insurable wage variation. As illustrated in Heathcote, Storesletten and Violante (2008), larger dispersion of the insurable shock $\varepsilon$ allows agents to work when they are more productive and take leisure when they are less productive, with the net effect of raising average leisure, average productivity, and welfare. Progressivity weakens this channel and leads to an increase in labor supply (and a decline in welfare through a productivity loss from the distortion, as we will see). The third term captures the fact that a higher disutility of work effort motivates agents to choose a lower labor supply. The fourth term shows that the response of hours worked to an insurable shock
which does not trigger any income effect precisely because of its insurability) is mediated by the tax-modified Frisch elasticity $1/\hat{\sigma}$. Progressivity lowers this elasticity towards zero.

The consumption allocation is additive in five separate components. The first component is consumption of the representative agent, described in Section 3.1.1. The second component shows that insurable variation in productivity increases average productivity, and has a positive level effect on average consumption. From the expression for $\mathcal{M}(v_\varepsilon; \tau)$, it is easy to see that higher progressivity reduces this productivity gain. Since hours worked are decreasing in the disutility of work $\varphi$, so are earnings and consumption. Consumption is increasing in the skill level $s$, through its price, and in the uninsurable component of wages $\alpha$. The redistributive role of progressive taxation is evident from the fact that a larger $\tau$ shrinks the pass-through from heterogeneity in initial conditions $\varphi, \kappa$ (and hence $s$) and ex-post realizations of the uninsurable wage shocks $\alpha$ onto consumption. Finally, because of the separability between consumption and leisure and the complete-market structure, $c$ is independent of $\varepsilon$.

**Proposition 3 [skill price and skill choice].** In equilibrium, the skill price is given by

$$\log p(s; \tau) = \pi_0(\tau) + \pi_1(\tau) \cdot s(\kappa; \tau)$$

where $\pi_0(\tau) = \frac{1}{2(\theta-1)} \left[ \log (1 - \tau) - \ln (\theta) \right] + \frac{1}{(\theta-1)} \log \left( \frac{\theta}{\theta-1} \right)$, and $\pi_1(\tau) = \sqrt{\frac{1}{\theta(1-\tau)}}$.

The skill investment allocation is given by

$$s(\kappa; \tau) = \eta \sqrt{\frac{1 - \tau}{\theta}} \cdot \kappa$$

and, thus, the equilibrium skill density $m(s)$ is exponential with parameter $\sqrt{\frac{\theta}{1-\tau}}$.

The first important result is that the log of the equilibrium skill price has a “Mincerian” shape, i.e., it is an affine function of $s$. The constant $\pi_0(\tau)$ is the base log-price of the lowest skill level ($s = 0$) and $\pi_1(\tau)$ is the marginal return to skill. As is evident from (23), a higher value for $\tau$ (more progressivity of the tax system) depresses skill investment and compresses the skill distribution towards zero. Because of imperfect substitution in produc-

---

13This result stems for three key assumptions we made: (i) log utility, quadratic skill investment cost, and (iii) exponential distribution of $\kappa$. See the Appendix for more details.
tion, the implied relative scarcity of high skill types increases the marginal return $\pi_1(\tau)$ and reduces the base price $\pi_0(\tau)$. [GV: cite Stiglitz 1986]

With the solution for the skill price and the consumption allocation in hand, the expression for the skill choice $s$ is easy to understand. Using the period-utility specification in (8) and the consumption allocation in (21), together with the skill price in (22), into the first-order condition (13), expression (23) follows immediately. Note that the skill investment decision is independent of $\varphi$ (and it would also be independent of $\alpha_0$ if there was heterogeneity in initial labor productivity within skill types). The logic is that, with log utility, the welfare gain from skill investment is proportional to the log change in wages, which is independent of the level of wages and hours.

**Corollary 3.1 [distribution of skill prices].** The variance of log skill prices is

$$\text{var} \left( \log p(s; \tau) \right) = \frac{1}{\theta^2},$$

and the distribution of skill prices (in levels) is Pareto. It follows that the upper tail of the equilibrium wage and income distributions is Pareto with parameter $\theta$.

Equation (23) reveals that because (i) the skill distribution retains the exponential shape of the distribution of learning ability $\kappa$, and (ii) the log of the skill price is affine in $s$, the variance of the log skill price is $\text{var} \left( \log p(s) \right) = 1/\theta^2$ and is, therefore, independent of $\tau$. It may appear surprising that inequality in pre-tax wages is independent of $\tau$. The reason is that progressivity sets in motion two offsetting forces. On the one hand, as discussed earlier, higher progressivity increases the equilibrium skill premium $\pi_1(\tau)$, which tends to raise inequality (the price effect). On the other hand, higher progressivity compresses the distribution of skills (the quantity effect). These two forces exactly cancel out and the variance of log wages is independent of $\tau$.

Since the exponential of an exponentially distributed random variable is Pareto, the distribution of skill prices in levels is Pareto with parameter $\theta$. The other stochastic components of wages (and hours worked) are lognormal. Therefore, the equilibrium distributions of wages and earnings (in level) have a Pareto tail, a robust feature of their empirical counterparts (see, e.g., Atkinson, Piketty and Saez, 2011, for a survey). Moreover, the distributions of
log-wage and log-earnings can be solved for in closed form, since they are Exponentially-Modified Gaussian (EMG) distributions given by the linear combinations of an exponential random variable $p(s; \tau)$ and a Normal random variable $(\alpha + \varepsilon)$. This is a useful result for our political-economic analysis of Section 7.4.

We now briefly discuss how taxation affects aggregate quantities in our model.

**Corollary 3.2 [aggregate quantities].** *Average hours worked and average effective hours are independent of skill type $s$ and given by*

$$
H(s; \tau) = H(\tau) = (1 - \tau)^{\frac{1}{1+\sigma}} \cdot \exp \left( \frac{\tau (1 + \sigma) \nu_{\varepsilon}}{\sigma^2} \right),
$$

**(24)**

$$
N(s; \tau) = N(\tau) = H(\tau) \cdot \exp \left( \frac{1}{\sigma} \nu_{\varepsilon} \right).
$$

**(25)**

*Output is given by*

$$
Y(\tau) = \mathbb{E}[p(s; \tau)] \cdot N(\tau)
$$

*where $\mathbb{E}[p(s; \tau)] = \frac{1}{\theta} \cdot \exp(\pi_0(\tau))$. Aggregate labor productivity is*

$$
\frac{Y}{H} = \frac{Y}{N} \cdot \frac{N}{H} = \mathbb{E}[p(s; \tau)] \cdot \exp \left( \frac{1}{\sigma} \nu_{\varepsilon} \right)
$$

Note that progressivity affects aggregate output through two channels: $\tau$ affects the average skill price (wage) via its impact on skill investment choices, and $\tau$ affects average hours worked. From equation (25) the elasticity of aggregate hours $N(s; \tau)$ with respect to $\tau$ at $\tau = 0$ is $\frac{-1}{1+\sigma}$. From equation (22) the elasticity of the average skill price $\pi_0(\tau)$ is $\frac{-1}{2(\theta-1)}$. We will return to these two elasticities in Section 5.2 when characterizing the conditions under which the optimal tax system is progressive.

[JH: Jon’s notes report elasticities of aggregate hours and earnings with respect to $(1-\tau)$. Can we find any empirical papers with comparable estimates? Another point we could note is that progressivity impacts the levels of hours and earnings, but does not impact dispersion in hours or earnings. Is there any empirical evidence on that point? The RED paper doesn’t show any action around 1986]

Before turning to the characterization of the optimal degree of progressivity, we state a useful result on social efficiency of the competitive equilibrium.
Corollary 3.3 [efficiency of equilibrium]. When \( v_\omega = 0 \) and \( \chi = 0 \), the competitive equilibrium with \( \tau = 0 \) is Pareto optimal.

There are two reasons why the competitive equilibrium with \( \tau = 0 \) is not generally efficient in our environment. The first is that there are no private markets for insuring the \( \omega \) shock. The second is that there is an externality in the individual labor supply decision when \( \chi > 0 \) so that the publicly provided good is valued: agents do not internalize that, by working more hours, the quantity of public good will increase.

Corollary 3.3 states that when the economy features complete markets with respect to wage shocks \( (v_\omega = 0) \) and does not feature any public-good externality \( (\chi = 0) \), one can find planner weights (a function of initial conditions \( \varphi \) and \( \kappa \)) such that the equilibrium allocations with \( \tau = 0 \) are efficient.

5 The social welfare function

We begin by analyzing the optimal choice of progressivity \( \tau \) and public good provision \( g \) for a utilitarian government that chooses \( (g, \tau) \) once and for all in order to maximize social welfare. In defining an objective function for the government, we face two decisions. First, the overlapping-generations structure of the economy requires us to take a stand on how the government weights currently alive and future generations.

We assume that the planner discounts future generations with the same discount factor \( \beta \) as the individual use. This implies that the planner will, at any point in time, weight all individuals alive the same, and the social welfare function is time consistent (cf. Obstfeld, 1986).

Second, the presence of human capital in the model introduces two related issues. As is well understood from the capital taxation literature, when human capital investment of the current cohorts is sunk the government has a temptation to tax it because such taxation is not distortionary. Moreover, when the stock of human capital adjusts slowly a change in the tax system induces transitional dynamics. In our benchmark analysis, we sidestep these two issues by making the assumption that the choice of skills is fully reversible at any point. In Section 7.2 we generalize our characterization of optimal progressivity by making the
polar opposite assumption that skills are fully irreversible. In this version of the model there are transitional dynamics between initial and final steady-state, and we can analyze how the motive to tax the existing stock of skills affects the optimal choice of progressivity.

The next lemma shows that, under these two assumptions, the social welfare function takes a familiar form.

**Lemma 1 [social welfare function].** If the government (i) puts equal weight (normalized to 1) on all currently alive agents and discounts at rate \( \beta \) utility of all agents in future cohorts; and (ii) skill investment is reversible, then the social welfare function \( W(g, \tau) \) equals average period utility in the cross-section, or

\[
W(g, \tau) = (1 - \delta) \sum_{j=0}^{\infty} \delta^j \mathbb{E} [u(c(\varphi, \alpha_j, s, g, \tau), h(\varphi, \varepsilon; \tau), G(g, \tau))] - \mathbb{E} [v(s, \kappa)]
\]

where the first expectation is taken with respect to the equilibrium cross-sectional distribution of \((\varphi, \alpha_j, s, \varepsilon)\), and the second expectation with respect to the cross-sectional distribution of \((s, \kappa)\).\(^{14}\)

The next proposition shows that we can solve for social welfare explicitly as a function of \((g, \tau)\) and of the other structural parameters of the model.

**Proposition 4 [closed-form solution].** The economy’s social welfare function \( W(g, \tau) \) is

\(^{14}\)The uninsurable shock \( \alpha_j \) is indexed by age \( j \) because the conditional variance of the unit root process \( \alpha \) depends on age \( j \).
given by:

\[
W(g, \tau) = \log(1 - g) + \chi \log g + (1 + \chi) \frac{\log(1 - \tau)}{(1 + \sigma)(1 - \tau)} - \frac{1}{(1 + \sigma)} 
\]

\[
+ (1 + \chi) \left[ -\frac{1}{\sqrt{\theta - 1}} \log \left( \sqrt{\frac{\theta}{1 - \tau}} \right) + \frac{\theta}{\theta - 1} \log \left( \frac{\theta}{\theta - 1} \right) \right] 
\]

\[
- \frac{1}{2\theta} (1 - \tau) 
\]

\[
- \left[ -\log \left( 1 - \left( \frac{1 - \tau}{\theta} \right) \right) - \left( \frac{1 - \tau}{\theta} \right) \right] 
\]

\[
- (1 - \tau)^2 \frac{\nu_\nu}{2} 
\]

\[
- \left[ (1 - \tau) \frac{\delta}{1 - \delta} \frac{\nu_\omega}{2} - \log \left( 1 - \delta \exp \left( -\frac{(1 - \tau)\nu_\omega}{2} \right) \right) \right] 
\]

\[
+ (1 + \chi) \left[ \frac{1}{\delta} v_\epsilon - \sigma \frac{1}{\sigma^2} \frac{v_\epsilon}{2} \right]. 
\]

The function \( W(g, \tau) \) is globally concave in \( g \) and, if \( \sigma \geq 2 \), it is globally concave in \( \tau \).

The proof of Proposition 4 in the Appendix shows that, in order to obtain the expression in equation (27), one has first to solve for the value \( \lambda(g, \tau) \) that balances the government budget. This value is needed to fully solve for the consumption allocation (21). Next, plugging the consumption, hours and skill allocations into (26), and appropriately rearranging, one obtains this analytical formulation of the social welfare function. As we show in the Appendix, aside from the term multiplying \( v_\epsilon \) in the last row of (27), the social welfare function is globally concave in \( \tau \). The term involving \( v_\epsilon \) is also globally concave in \( \tau \) if \( \sigma \geq 2 \), a condition that is satisfied in the calibration.\(^{15}\)

### 5.1 Decomposition of the social welfare function

We now demonstrate that every term in (27) has an economic interpretation and captures one of the many forces that determine the optimal level of progressivity (and the optimal level of public good provision) in the economy. All these forces are additively separable in the social welfare function, and do not interact with each other, making their interpretation very transparent.

\(^{15}\)See the Appendix for a more thorough interpretation of sufficient conditions.
5.1.1 Welfare of the representative agent

By substituting allocations (17) and (18) into the objective function of the representative agent problem of Section 3.1.1 (after solving for $\lambda$), one obtains the indirect welfare function of the representative agent

$$W^{RA}(g, \tau) = \log(1 - g) + \chi \log g + (1 + \chi) \frac{\log(1 - \tau)}{(1 + \hat{\sigma})(1 - \tau)} - \frac{1}{(1 + \hat{\sigma})}$$

(28)

which is precisely the first line of the social welfare function in (27). Note that this is the only place where $g$ appears and, as a result, by differentiating (28) we can determine the optimal level of public expenditures in the economy. It is easy to verify that

$$g^* = \frac{\chi}{1 + \chi}$$

(29)

which states that the government equates the marginal rates of substitution between private and public consumption, $\chi g / (1 - g)$ from the representative agent problem in (16) to the marginal rate of transformation between the two goods, equal to one. This choice of $g$ is independent of $\tau$. [GV: shall we say it’s Samuleson’s condition? KS: DEFINITELY!]

What is the optimal degree of progressivity in the representative agent economy? Differentiation of (28) with respect to $\tau$ yields

$$\tau^*_{RA} = -\chi$$

(30)

or, the benevolent government in the representative agent economy would set regressive taxes, proportionately to the relative taste for the public good. In this economy, there is an externality. The individual agent does not take into account that, by working more and producing more, output increases and the government can distribute more of the public good. A regressive tax increases labor supply, as is clear from the hours allocation (18), and entirely corrects this externality as stated in the following corollary.\(^{16}\)

[GV: Is the Feldstein result related?]

**Corollary 4.1 [efficiency in the representative agent model].** If $v_\varphi = v_\omega = v_\varepsilon = 0$ and $\theta = \infty$ (i.e., in the representative agent economy), $g^* = \frac{\chi}{1 + \chi}$ and $\tau^* = -\chi$ implement the first best.

\(^{16}\)If there was a private market for $G$, and $G$ was a non-rival non-excludable good, then the welfare theorems would apply and the first best could be implemented with $\lambda = 1$ and $\tau = 0$. 

19
To gain some intuition, note that one could alternatively implement the first best with lump sum taxes that do not distort labor supply and finance the desired amount of $G$. At the equilibrium level of income $Y$, the marginal tax rate with $\tau$ set to $-\chi$ is zero, but the average tax rate is positive to finance expenditures. Therefore, the system replicates a lump sum tax. [KS: perhaps say that the regressive tax is the only way for the Benabou tax system to implement the FB]

### 5.1.2 Welfare from skill investment

The second, third and fourth line in (27) are all related to the skill investment choice. To begin with, we have that

$$\log \left( \frac{Y}{N} \right) = \frac{-1}{\theta - 1} \log \left( \sqrt{\frac{\theta}{\theta - 1}} \right) + \frac{\theta}{\theta - 1} \log \left( \frac{\theta}{\theta - 1} \right),$$

i.e., this term is the log of aggregate productivity (output per efficiency unit of labor) in the economy, also equal to $\log \mathbb{E} [p (s; \tau)]$. It captures the fact that the pattern of skill investments determines productivity through the CES technology: output is larger the more evenly distributed are skills. Equation (31) shows that higher progressivity $\tau$ reduces skill investment and productivity. Note that, in welfare, productivity is multiplied by $(1 + \chi)$ because the higher the desire for public good, the more valuable is an additional unit of output (as more can be allocated to $G$ without foregoing private consumption).

Skill investment is not costless. The average skill investment costs in the population is

$$\mathbb{E} [v (s, \kappa)] = \frac{1}{2\theta} (1 - \tau),$$

or the term in the third line of (27) which enters with a negative sign and reduces welfare. This term is decreasing in $\tau$ because larger level of progressivity $\tau$ reduces skill acquisition and its costs as well. When combining (31) and (32), it is easy to show that the productivity gain from skill investment net of education costs is maximized at $\tau = -1/((\theta - 1) < 0$: a regressive system invites larger skill investments and the stronger is complementarity in production, the stronger is this force.

However, the government also cares about how the choice for $\tau$ impacts consumption dispersion both directly (via redistribution) and indirectly, via its effects on equilibrium prices.
and quantities of skills. See equations (22) and (23). The welfare cost of this consumption dispersion across skill types is

\[ \text{welfare cost of skill price dispersion} = -\log \left( 1 - \frac{1}{\theta} \right) - \left( 1 - \frac{\tau}{\theta} \right) \]  

(33)

or the term in the fourth line of (27). This cost is decreasing in \( \tau \) because higher progressivity reduces post-government earnings and consumption dispersion.

We learned that there are offsetting forces that determine the optimal level of progressivity with respect to skill acquisition: more progressivity diminishes aggregate productivity, but it also decreases consumption dispersion across skills. Which force is dominant? If we set \( \chi = 0, \sigma = \infty \) and \( v_\phi = v_\omega = v_\epsilon = 0 \) to isolate the skill investment channel, the answer depends entirely on \( \theta \).\(^{17}\) Figure 2 shows that the relationship between \( \tau^* \) and \( \theta \) is non-monotone and hump-shaped. As \( \theta \to \infty \) the economy converges to a representative agent economy, and there is no role for progressive taxation (\( \tau^* = 0 \)). For lower values for \( \theta \), the utilitarian government chooses \( \tau^* > 0 \) to reduce the welfare loss from consumption dispersion. However, as \( \theta \to 1 \) and skill complementarity increases in production, the distortion from progressivity on aggregate productivity becomes more important – because of the term \( \frac{1}{2(\theta - 1)} \log(1 - \tau) \) in (31)– and \( \tau^* \) falls. At the same time, a regressive tax scheme, which

\(^{17}\)Equation (31) shows that welfare is additively separable in \( \mu \) and \( \eta \), and hence these two parameters are irrelevant.
would raise productivity, would make consumption inequality explode: equation (33) shows that, as $\theta \to 1$ the term $-\log (\theta - 1 + \tau)$ goes to $-\infty$ as $\tau$ approaches zero. Overall, these two forces balance out, and as $\theta \to 1$, $\tau^\ast \to 0$—a flat tax system. In conclusion, the skill investment component of social welfare pushes $\tau^\ast$ towards progressivity, and this force is stronger for intermediate values of $\theta$. Note that $\tau^\ast$ is always positive in Figure 2. In Section 5.2, we formally prove this result.

### 5.1.3 Welfare from preference heterogeneity and uninsurable wage risk

The existence of heterogeneity in preference for leisure, through variation in the parameter $\varphi$, translates into dispersion in hours worked, earnings and consumption. We have that

$$\text{var}_\varphi (\log c) = (1 - \tau)^2 v_\varphi.$$  \hfill (34)

The fifth line of the social welfare function is the welfare cost due to this source of consumption dispersion. This term is the familiar Lucas representation of the welfare cost of consumption dispersion when the underlying shocks are lognormal, i.e., one half of the variance of log consumption times the coefficient of risk aversion, equal to one in our model.

Uninsurable shocks are another key source of consumption dispersion which shows up in the sixth line of the social welfare function:

$$\text{welfare cost of } \text{var}_\alpha (\log c) = \left[ (1 - \tau) \frac{\delta}{1 - \delta} \frac{v_\omega}{2} - \log \left( \frac{1 - \delta \exp \left( -\tau(1 - \tau) \frac{v_\omega}{2} \right)}{1 - \delta} \right) \right]$$  \hfill (35)

where the approximation (which uses $\log (1 + x) \simeq x$) is extremely accurate for plausible parameter values.\(^{18}\)

As can be seen from (34) and (35), a higher $\tau$ reduces consumption dispersion stemming from both preference heterogeneity and uninsurable risk. Since consumption inequality lowers welfare, these two forces push the optimal $\tau$ towards one, the value at which there would be zero ex-post consumption dispersion.

\(^{18}\)The variance $v_\alpha$ is the cross-sectional variance of the cumulated innovations $\omega$, or $v_\alpha = \delta \tau v_\omega$ where the $\delta$ in the numerator reflects the fact that we have assumed that wage uncertainty starts realizing at age $a - 1$.  

22
5.1.4 Welfare from insurable wage risk

The last two terms of the welfare expression are also easily interpretable. Note that:

\[
\log \left( \frac{N}{H} \right) = \frac{1}{\sigma} v_\varepsilon \quad (36)
\]

\[
\text{var}_\varepsilon (\log h) = \frac{1}{\sigma^2} v_\varepsilon. \quad (37)
\]

The first term is the log productivity gain from insurable wage variation. As explained when discussing the equilibrium allocations, more insurable wage dispersion is good news for welfare, as individual hours worked become more positively correlated with individual productivity and aggregate output increases. Hours dispersion is, however, costly in welfare terms because of the convexity in the disutility of hours. This cost is captured by the last term in the welfare expression where the cross-sectional variance of log hours due to insurable shocks is multiplied by \( \sigma \), which measures the degree of individual aversion to hours fluctuations.\(^{19}\)

These last term associated to insurable variation (productivity gain net of the disutility costs of hours fluctuations) is maximized at \( \tau = 0 \) because both progressivity and regressivity induce some misallocation in hours worked. Recall that a flat tax has no impact on labor supply with our preference specification. To conclude, larger insurable wage risk pushes the optimal \( \tau \) towards zero.

5.2 When should taxes be progressive?

By differentiating the expression for social welfare in equation (27) with respect to \( \tau \) one can obtain a necessary and sufficient parametric condition for the optimal tax system to be progressive.

**Proposition 5 [condition for progressivity].** The optimal value for \( \tau \) is strictly positive if and only if

\[
\frac{1}{2} \left( \frac{1}{\theta - 1} - \frac{1}{\theta} \right) + (v_\varphi + v_\alpha) > \chi \left( \frac{1}{2(\theta - 1)} + \frac{1}{\sigma + 1} \right). \quad (38)
\]

\(^{19}\)As for the productivity gain from skill investment, also the productivity gain from insurable risk is multiplied by \( (1 + \chi) \), a term that takes into account the additional value of an extra unit of output when government expenditures give agents utility.

23
The terms on the left hand side of (38) are the marginal benefits from increasing progressivity at \( \tau = 0 \), and the ones on the right-hand side are corresponding marginal costs.\(^{20}\) The first term is the gain from reducing consumption inequality across skill types net of the productivity loss from reduced skill investment (minus its costs). The second term is the gain from reducing consumption dispersion across agents with different idiosyncratic preferences for work and labor productivity.

The term in parentheses on the right hand side is the elasticity of aggregate output with respect to \( \tau \). The first component captures the loss in output associated with reduced skill investment, while the second reflects the loss from reduced labor supply: recall the discussion of Corollary 3.2. The welfare loss from lower output is proportional to \( \chi \), which captures the strength of the public good externality. The logic is that \( \chi \) captures the wedge between the social marginal value of additional output and the private marginal value. Note that if \( \chi = 0 \) then the optimal tax system is always progressive.

### 5.3 Optimal marginal tax rate at the top

One focus of the Mirrlees approach to optimal taxation has been characterizing the optimal marginal tax rate at the top of the income distribution. Assuming an unbounded Pareto right tail for *exogenous* labor productivity (and assuming the social welfare function puts zero weight on agents far in the tail), Saez (2001) shows that the marginal tax rate at the top converges to

\[
\hat{t} = \frac{1}{1 + \zeta^u + \zeta^c(\theta - 1)}
\]

(39)

where \( \zeta^u \) and \( \zeta^c \) are uncompensated and compensated labor supply elasticities.\(^{21}\) Thus, as Saez and others have noted, the thicker is the right tail of the productivity distribution (i.e. the smaller is the Pareto coefficient \( \theta \)) the higher is the optimal marginal tax rate at the top.

Our model with \( \chi = 0 \) and \( v_\rho = v_\omega = v_\epsilon = 0 \) is a version of the Mirlees environment, but with one important difference: the Pareto distribution for labor productivity is *endogenously*

---

\(^{20}\)This condition is independent of the insurable variance \( v_\epsilon \) because the term in welfare involving this component is maximized at \( \tau = 0 \), and thus the marginal welfare effect from a change in progressivity is zero at \( \tau = 0 \).

\(^{21}\)Given our utility function, as earnings increase, these elasticities converge to zero and \( 1/(1 + \sigma) \) respectively, and thus the efficient top marginal tax rate converges to \( \hat{t} = \frac{1+\sigma}{\sigma+\theta} \).
determined by skill investment and its Pareto parameter is the elasticity of substitution across skill types in production. As we showed in Section 5.1.2, the optimal choice for $\tau$ is non-monotonic in $\theta$.\footnote{Figure 2 is drawn for the case $\sigma = \infty$, but the non-monotonicity result is more general and applies with elastic labor supply.} In particular, there is a range of values for $\theta$ close to unity in which reducing $\theta$ lowers progressivity, and thus marginal tax rates at the top, in sharp contrast to the familiar Mirlees result.\footnote{Strictly speaking, for any $\tau > 0$, marginal tax rates converge to one under our functional form for taxes as earnings converge to infinity. But from equation (3) it is easy to show that marginal tax rates are strictly increasing in progressivity for any $y \in (\exp(-1/(1-\tau)), \infty)$.} Recall the logic for this result: the more complementary are skill types, the larger are the productivity gains from a more even skill distribution. Thus the more costly are high tax rates at the top because they discourage skill investment of high learning-ability (high $\kappa$) individuals.

We conclude that whether the Pareto right tail in the earnings distribution reflects exogenous luck or endogenous investments is quantitatively important in determining optimal top marginal tax rates.

## 6 Quantitative implications of the theory

After describing the model’s parameterization, we explore the quantitative implications of the theory. Next, we perform a robustness analysis with respect to (i) the weights used by the government in its social welfare function and (ii) how we model government expenditures.

### 6.1 Parameterization

Thanks to the closed-form solution for the allocations, we can derive analytical expressions for the cross-sectional moments of the joint equilibrium distribution of wages, hours, and consumption. The explicit analytical links between structural parameters and equilibrium moments enable us to show identification of all parameters and estimate the model based on empirical counterparts of these moments computed with commonly used micro-data on wages, hours worked, and consumption.

We begin by recognizing that, in survey data, hours and consumption are measured with error and hourly wages (computed as annual earnings divided by annual hours) inherit mea-
measurement error from both variables. Let $v_{\mu h}, v_{\mu c}, v_{\mu y}$ denote the variances of reporting error in hours, consumption and earnings, respectively, and assume measurement error is classic. If we tack on measurement error to log wages and the log allocations in (20) and (21), and compute cross-sectional moments of their joint distribution, we obtain the following set of moment conditions:

\[
\begin{align*}
\text{var} (\log w) &= \frac{1}{\theta^2} + v_\alpha + v_\varepsilon + v_{\mu y} + v_{\mu h} \\
\text{var} (\log h) &= v_\varepsilon + \frac{1}{\theta^2} v_\varepsilon + v_{\mu h} \\
\text{var} (\log c) &= (1 - \tau)^2 \left( v_\varepsilon + \frac{1}{\theta^2} + v_\alpha \right) + v_{\mu c} \\
\text{cov} (\log h, \log w) &= \frac{1}{\theta} v_\varepsilon - v_{\mu h} \\
\text{cov} (\log h, \log c) &= (1 - \tau) v_\varepsilon \\
\text{cov} (\log w, \log c) &= (1 - \tau) \left( \frac{1}{\theta^2} + v_\alpha \right)
\end{align*}
\]

These moments contain all the structural parameters of the model. The variance of the uninsurable innovation $v_\omega$ is implied by $v_\alpha$, given a value for $\delta$.\(^{24}\)

Based on our previous work (Heathcote, Storesletten, and Violante, 2013), we set (i) the variances of measurement error to $v_{\mu h} = 0.036$, $v_{\mu y} = 0$ and $v_{\mu c} = 0.040$; and (ii) $\sigma = 2.165$, a value broadly consistent with the microeconomic evidence on the Frisch elasticity (see, e.g., Keane, 2011). In light of our evidence on the progressivity of the US tax/transfer system described in Section 2, we set $\tau = 0.151$.

It is easy to see that $v_\varepsilon$, $v_\varepsilon$, and $(v_\alpha + 1/\theta^2)$ are over-identified by the set of moments in (40). To separately identify the cross-sectional variance of uninsurable risk, $v_\alpha$, from the cross-sectional variance of skill-prices, $1/\theta^2$, we use the cross-sectional moments $\text{var}^0 (\log w)$, $\text{var}^0 (\log c)$ and $\text{cov}^0 (\log w, \log c)$ at age $j = 0$ which reflect only variation in skills acquired before labor market entry, since $v_0^0 = 0$. In Section 6.2.1 we pursue an alternative strategy for separating $v_\alpha$ and $\theta$.

\(^{24}\)The variance of the uninsurable shocks at age $j$ is $v_\alpha^a = j v_\omega$, and the cross-sectional uninsurable variance in the model is

\[
v_\alpha = (1 - \delta) \sum_{a=0}^{\infty} \delta^a v_\alpha^a = \frac{\delta}{1 - \delta} v_\omega.
\]
Our data are drawn from from two surveys, the Panel Study of Income Dynamics (PSID) for year 2000, 2002, 2004, and 2006 and the Consumption Expenditure Survey (CEX) for years 2000-2006. We apply the same sample selection criteria outlined in Section 2. We first regress individual log wages, individual log hours, and household log consumption on year dummies, a quartic in age, and (for consumption) household composition dummies. We then use the residuals from these regressions to construct the empirical counterpart of the moments in (40) plus the three moments at age “zero” (an average of ages 25-29 in the data). The minimum distance estimation uses, therefore, 9 moments for 4 parameters \((\phi, \epsilon, \alpha, \theta)\). The estimated parameter values are summarized in Table 1, together with the other pre-determined parameter values.

We set \(\delta = 0.971\) to match an expected working life of 35 years, the same age span considered in the micro-data, and obtain \(\omega = 0.003\).

To interpret the estimates in Table 1 it is useful to report the values for some of the empirical moments. In the data, \(\text{var} (\log w) = 0.43\), \(\text{var} (\log h) = 0.11\), \(\text{var} (\log c) = 0.18\), and \(\text{var}^0 (\log c) = 0.15\).\(^{25}\) Our estimates imply that, net of measurement error, (i) the insurable component accounts for 40 percent of the variance of wages, the uninsurable component accounts for one quarter, and the heterogeneity in skills for the residual one third; (ii) cross-sectional dispersion in the disutility of work effort explains almost half of the hours variation, and insurable shocks explain the other half; (iii) one fifth of consumption inequality is due to dispersion in the disutility of work, and the residual is accounted for equally by uninsurable wage risk and skill heterogeneity; (iv) the growth in the variance of

\(^{25}\)The values of the other empirical moments used in the estimation are: \(\text{cov} (\log h, \log w) = -0.09\), \(\text{cov} (\log h, \log c) = 0.03\), \(\text{cov} (\log w, \log c) = 0.15\), \(\text{var}^0 (\log w) = 0.28\), and \(\text{cov}^0 (\log w, \log c) = 0.10\).
Figure 3: Social Welfare as a function of $\tau$ and welfare gain relative the the current US system (left panel). Decomposition of social welfare into various components described in Section 5.1. The optimal value of $\tau^{US}$ is 0.065.

log consumption over the life cycle (ages 25-60) is around 0.10. These findings are broadly in line with the results in Heathcote, Storesletten, and Violante (2013). Overall, our parameter values are consistent with the level of inequality in wages, hours, and consumption in the early 2000s in the US.

To set the value for $\chi$, the relative weight on the publicly provided good in preferences, we take the view that the fraction of output devoted to $G$ is chosen efficiently. In Section we show that if $g$ was the outcome of voting, the median voter would pick the efficient level $g^*$ which provides a strong foundation for this calibration choice. Note that, with log utility, even if a fraction of $G$ is wasted and does not enter agents’ utility, the optimality condition (29), as well as the median voter outcome, would not be affected. Since, over the period 2000-2006 $g = G/Y = 0.189$, we set $\chi = 0.233$. XXX WITH $g = 0.137$ (SO MEDIAN VOTER PREFERENCES $\tau = 0.151$), THE $\chi$ BECOMES $0.137 = \frac{\chi}{1+\chi} \Rightarrow \chi = 0.159$ XXX

The optimal level of progressivity is quite sensitive to this choice, and Section ?? discusses alternative scenarios.
6.2 Results

Once the optimality condition $g^* = -\chi$ is substituted into (27), and values have been assigned to all the structural model’s parameters, one obtains social welfare $\mathcal{W}(\tau)$ only as a function of $\tau$. Figure 3(a) plots this function. The optimal value of progressivity that maximizes social welfare is found to be $\tau^* = 0.065$. The welfare gain from reducing progressivity from the current value of $\tau^{US} = 0.151$ to $\tau^*$ is equivalent to 0.5 percent of lifetime consumption. How different are the actual from the optimal scheme? First, note that the ratio of the variance of log disposable income to pre-government income is $(1 - \tau)^2$. This ratio would drop from 0.72 to 0.87. Figure 3(b) plots the marginal tax/transfer schedule under the two systems. At average income, the marginal tax rate would drop from 28 to 21 percent.

The right panel reconstructs $\mathcal{W}(\tau)$ by adding, one by one, all its components. The first component is welfare of the representative agent. As discussed, it is maximized at $\tau = -\chi = -0.233$. Adding the skill-investment component (productivity gain from skill investment net of education costs minus the implied welfare loss from between-skill consumption inequality) pushes towards a more progressive system, and the optimal $\tau$ moves to the right to $-0.062$. The concern for additional consumption inequality induced by preference heterogeneity raises the optimal $\tau$ only slightly to $-0.02$. Uninsurable shocks are a stronger source of consumption dispersion, which is reflected in the substantial upward jump in $\tau$ to 0.076 when this latter component is incorporated. Finally, adding the productivity gain from insurable shocks pulls $\tau$ back towards zero to its final value of 0.065.

Consumption dispersion generated by preference heterogeneity, skill dispersion, and uninsurable risk induces the government to choose a progressive scheme. If this was the only government’s concern, $\tau$ would be optimally set to one, and the tax/transfer scheme would fully redistribute household income across households to equate post-government income and consumption. Beyond the provision of public goods, which we analyze in detail in Section 6.4 below, there are two forces that limit progressivity in the model: the distortion to skill investment and the distortion to labor supply.

To measure the strength of these two channels, we compute the optimal $\tau$ when (i) $\sigma = \infty$
and labor supply is inelastic, and when (ii) skills are exogenous.\footnote{This latter case is obtained by excluding from the welfare function the first two terms associated to the productivity gain from skill investment net of the education cost because, if skills are exogenously determined, they would not be affected by the value of \( \tau \).} In the case \( \sigma = \infty \), the optimal \( \tau \) is 0.221 and in the case of exogenous skill distribution, the optimal \( \tau \) is 0.214. Therefore, the endogeneity of labor supply and skill investment play, quantitatively, very similar roles in limiting progressivity, and in the absence of either one of these channels, optimal progressivity would be substantially higher.\footnote{In the absence of both forces, the optimal \( \tau \) is one.}

6.2.1 Alternative calibration of \( \theta \) and top marginal tax rates

Our model offers an alternative strategy to calibrate \( \theta \). Since the model’s distribution of income is Pareto with parameter \( \theta \), we have that

\[
\frac{E[p(s) | s > \bar{s}]}{p(\bar{s})} = \frac{\theta}{\theta - 1}.
\]

From our PSID sample, we estimate that the ratio in (41) is stable around 2 for income thresholds above $250,000, which implies a value of \( \theta = 2 \).\footnote{Data from tax returns on wage income tabulated by Piketty and Saez (2003, Table B3) indicate a value for \( \theta \) between 1.6 and 2.2 for the years 2000-2006, depending on the choice for the threshold \( \bar{s} \). Thus our estimate falls within this range. Since the benchmark of comparison for our normative results is \( \tau^{US} \) estimated on PSID data, we use PSID data for this alternative estimate of \( \theta \).} If we use this moment, instead of consumption dispersion at labor market entry, to calibrate \( \theta \), and re-estimate the other model parameters, we obtain the values in the second row of Table 1. Because the consumption dispersion implied by skill investment is so large with \( \theta = 2 \), the estimation sets the variance of the uninsurable lifetime shocks \( \nu_\omega \) to zero.\footnote{As a result, under this calibration, the model cannot generate a rise in life-cycle consumption inequality.} The model still calls for positive preference heterogeneity and insurable dispersion to account for the cross-sectional inequality of hours in the data. Under this alternative calibration, we obtain \( \tau^* = 0.037 \). Optimal progressivity is lower than in the baseline calibration, because there is no longer a role for social insurance against privately-uninsurable shocks.

We have also experimented with varying \( \theta \) holding fixed all other structural parameters at their baseline values. The welfare-maximizing value for \( \tau \) turns out to be largely insensitive to \( \theta \), for \( \theta \in (2, \infty) \) (i.e., for all quantitatively plausible values). The logic is that the terms
Table 2: Optimal progressivity under non-utilitarian welfare

<table>
<thead>
<tr>
<th></th>
<th>Utilitarian</th>
<th>κ-neutral</th>
<th>φ-neutral</th>
<th>Insurance-only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Redist. wrt κ</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Redist. wrt φ</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Insurance wrt ω</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>

| τ^*                | 0.065       | 0.006     | 0.037     | -0.022         |
| Welf. gain (pct of c) | 0.49       | 1.35      | 0.84      | 1.92           |

in social welfare involving θ do not vary much with τ for values for τ between 0 and 0.2. Because other terms in the social welfare expression are more sensitive to τ (for θ > 2) those other terms tend to shape the optimal choice for τ.

6.3 Non-utilitarian welfare criteria

The utilitarian social welfare function in (26) embeds both the desire to insure households against the privately uninsurable lifecycle shocks ω and to redistribute against income differentials due to initial heterogeneity in preferences (φ) and learning ability (κ). We now consider some alternative formulations for the social welfare function which retain the desire to insure against uninsurable shocks, but switch off the desire to redistribute with respect to initial conditions.

Proposition 6 [efficiency with χ = νω = τ = 0]. If χ = νω = 0 then competitive equilibrium allocations with τ = 0 are efficient. The corresponding planner weights are

\[ ζ(α, κ) = \frac{1}{θ-1} \exp \left( -φ + \frac{η}{θ}κ \right) \] (42)

When there is no need to raise revenue to finance public expenditure (χ = 0), and when all life-cycle productivity shocks are insurable (νω = 0) the welfare theorems apply, and thus the laissez-faire competitive equilibrium with λ = 1 and τ = 0 is efficient. However, this competitive equilibrium does not correspond to the allocations that would be chosen by a utilitarian planner, since relatively high ability agents (with high draws for κ) and relatively diligent agents (with low draws for φ) will enjoy relatively high consumption – the former
because they will make relatively high skill investments, the latter because they will work relatively long hours. Proposition 6 describes the Pareto weights for the planner’s problem whose solution coincides exactly with the competitive equilibrium with \( \tau = 0 \). These planner weights, by construction, do not incorporate any desire to redistribute with respect to income differentials reflecting fixed heterogeneity in preferences or learning ability. The social welfare function corresponding to these weights overweights agents with high ability and high diligence.

The proof of Proposition 6 is straightforward. Competitive equilibrium allocations with \( \chi = v_\omega = \tau = 0 \) are easily derived by substituting those parameter values into equations (20), (21) and (??). We then solve a planner’s problem, where the planner discounts at rate \( \beta \) and maximizes social welfare subject to the aggregate resource constraints (equation 5). We show that when the planner uses the weights described in equation (42) to weight agents with different initial draws for \( (\varphi, \kappa) \) the solution to the planner’s problem coincides exactly with the competitive equilibrium allocations.

COMPLETE PROOF FOR APPENDIX

We label the planner with preferences described by equation (42) the “\((\varphi, \kappa)\)-neutral” or “insurance-only” planner, because the only motive for progressivity with such weights is insuring against ex-post realizations of the \( \omega \) shock. Using similar logic, we also construct preference weights for a \( \varphi \)-neutral planner, who is indifferent to consumption inequality originating from heterogeneity in the taste for leisure, and for a \( \kappa \)-neutral planner who has no desire to respond to income inequality generated by the initial heterogeneity in learning ability \( \kappa \) and the ensuing skill inequality.

Evaluating equilibrium allocations given the weights in equation (42) we can compute social welfare under the insurance-only social welfare function.

**Corollary 6.1 [non-utilitarian welfare].** Social welfare for the insurance-only planner
\[is\]

\[W^{\text{ins}}(g, \tau) = W(g, \tau) \]
\[+ (1 - \tau)^2 \frac{v_\varphi}{2} + (1 - \tau^2) \frac{v_\varphi}{2} \]
\[- \frac{1}{2\theta}(1 - \tau) + \frac{1}{2(\theta - 1)}(1 - \tau).\]

The second line corrects the welfare component associated to preference heterogeneity which appears in the utilitarian welfare function. The new term that replaces (34) penalizes deviations from \(\tau = 0\). The third line corrects the welfare component associated to skill-investment (productivity gain, education cost and consumption inequality). As special cases, the first two lines in (43) give welfare for the \(\varphi\)-neutral planner, and the first and third line give welfare for the \(\kappa\)-neutral planner.

Table 2 summarizes the results. We find that the insurance-only planner would set \(\tau^* = -0.022\). The concern for social insurance against lifecycle wage shocks offsets almost entirely the desire for regressivity linked to public good provision. The \(\varphi\)-neutral planner would set progressivity to 0.037 and the \(\kappa\)-neutral planner to 0.006. Overall, these governments with limited taste for redistribution choose tax systems that are nearly proportional. Under these alternative welfare criteria, the welfare gains of switching from the current progressive system towards the optimal near-proportional system are much larger: the insurance-only criterion implies a gain close to 2 percent of lifetime consumption.

### 6.4 Alternative modelling of \(G\)

In the baseline model we assumed that (i) households derive some utility from government expenditures \((\chi > 0)\), and that (ii) the government chooses a fraction of output \(g\) to be transformed into the publicly-provided good. As shown in Section 5.1, as long as some portion of government expenditures is valued, the model has an intrinsic force that restrains the tax system from being too progressive.

We now examine the sensitivity of our findings to these two assumptions. We set \(\chi = 0\) throughout – which implies that the publicly provided good is not valued – and examine
Table 3: Optimal progressivity under different scenarios for G

<table>
<thead>
<tr>
<th>Case</th>
<th>$G \frac{\tau}{Y}$</th>
<th>Utilitarian $\tau^*$</th>
<th>Welf. gain</th>
<th>Insurance-only $\tau^*$</th>
<th>Welf. gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g^*$ endogenous</td>
<td>$\chi = 0.233$ 0.189</td>
<td>0.065 0.49%</td>
<td></td>
<td>-0.022 1.92%</td>
<td></td>
</tr>
<tr>
<td>$g^*$ endogenous</td>
<td>$\chi = 0$ $0$</td>
<td>0.184 0.07%</td>
<td></td>
<td>0.092 0.20%</td>
<td></td>
</tr>
<tr>
<td>$\bar{g}$ exogenous</td>
<td>$\chi = 0$ $0.189$</td>
<td>0.184 0.07%</td>
<td></td>
<td>0.092 0.20%</td>
<td></td>
</tr>
<tr>
<td>$\bar{G}$ exogenous</td>
<td>$\chi = 0$ $0.189$</td>
<td>0.071 0.45%</td>
<td></td>
<td>-0.011 1.78%</td>
<td></td>
</tr>
</tbody>
</table>

three cases. In the first, $g$ is chosen optimally, and therefore set to $g^* = 0$. In the second, $g$ is fixed exogenously to $\bar{g} = g^{US} = 0.189$. Thus a fraction of output equal to government purchases’ share in the baseline economy is wasted. In the third, the level of expenditures is fixed to $\bar{G} = G^{US} = g^{US} Y(\tau^{US})$.

Table 3 reports the optimal degree of progressivity in these various cases for the utilitarian and the insurance-only planners.

To understand the first case ($\chi = 0$ and $g^* = 0$), recall that when government expenditures are valued by households, there is a force for regressive taxation as a means to encourage labor supply and skill investment (which would otherwise be inefficiently low given that agents do not internalize that more output allows for increased government purchases).

With $\chi = 0$, this externality is absent, and thus the optimal $\tau$ rises to 0.184 for the utilitarian planner and to 0.092 for the insurance-only planner.

The second case ($\bar{g}$ fixed exogenously but positive) delivers exactly the same optimal $\tau$ as the case $g^* = 0$. The logic is that fixing $\bar{g}$ means that a fixed fraction of aggregate output is wasted, akin to an economy with lower TFP. But since the model is invariant to scale optimal progressivity is independent of the fraction of output that is wasted and identical to optimal progressivity when $g = g^* = 0$.

The third case (the level of $G$ is fixed) gives a similar value for optimal $\tau$ to the baseline model, and the logic is similar: the planner internalizes that less progressive taxation encourages labor supply, and makes it easier to finance public good provision. Note that cases two and three are superficially similar – in both cases an exogenous amount of non-valued government purchases must be financed – but the two models give very different predictions for optimal progressivity. In both models lowering progressivity raises output. However, in
the fixed $G$ model more output makes financing government purchases easier – which acts as an incentive to limit progressivity – while in the fixed $g$ model more output necessitates proportionately larger purchases.

6.5 Taking stock

To visualize the differences between the actual and the optimal tax system, we plot marginal and average tax rates in the current scheme, and in the optimal scheme under the utilitarian and insurance-only criteria, for the baseline model where $G$ is valued (left-panels of Figure 4) and for the model where $\chi = 0$ and $g$ is exogenous (right-panels of Figure 4).

When $G$ is non-valued, the differences between actual scheme and utilitarian planner are minor, but when $G$ is valued they are substantial. For example, at average income, the marginal tax rate is 7 percentage points lower in the optimal scheme. The optimal degree of progressivity for the insurance-only planner is always lower than in the data and, when...
7 Extensions

7.1 Skill bias in production

A number of papers have emphasized that wage dispersion due to differences in skills may have two different origins: heterogeneity due to differences in relative scarcity, which we have emphasized above, and heterogeneity due to a bias in production (e.g., that high skilled have more efficiency units). In Katz and Murphy (1992) both these sources are present. They interpret skill-biased technical change as an increase in the weight on the high skilled in the production function. We now generalize our analysis to incorporate both these sources in our analysis. To this end, we extend the production function in (4) to the following multi-skill version of Katz and Murphy’s skill aggregation function:

\[
Y = \left\{ \int \exp(\varrho s) \cdot [N(s) \cdot m(s)]^{\frac{\theta-1}{\sigma}} ds \right\}^{\frac{\theta}{\theta-1}},
\]

where the weight \( \exp(\varrho s) \) captures the skill bias in production for skill \( s \). The skill bias must be bounded to ensure finite production and we therefore make the following assumption.

**Assumption 1** Assume that \( \varrho < \frac{\theta-1}{\theta} / \sqrt{2} \)

It is straightforward to verify that all the individual decision rules are the same as above, up to the fact that the return to skill is different. Therefore, under Assumption 1 Propositions 1 and 2 continue to hold, where part 2 of the equilibrium definition in section 4 is amended to

\[
p(s) = \exp(\varrho s) \left( \frac{Y}{N(s) \cdot m(s)} \right)^{\frac{1}{\theta}}
\]

(44)

The following amended version of Proposition 3 summarizes how the skill prices and skill choices are affected by the presence of skill bias in production.

---

30 The welfare difference between these two schemes is only 0.03% of lifetime consumption for the insurance-only planner.
Proposition 7 [prices under skill bias]. In equilibrium, the skill price is given by

$$\log p(s; \tau) = \pi_0(\tau) + \pi_1(\tau) \cdot s(\kappa; \tau)$$  \hspace{1cm} (45)

where the marginal return to skill is $$\pi_1(\tau) = \frac{\theta}{2} + \sqrt{\left(\frac{\theta}{2}\right)^2 + \frac{1}{\theta(1-\tau)}}$$ and $$\pi_0(\tau) = -\frac{1}{\theta} \ln \left(\frac{\theta - 1}{\theta} \left(\frac{\theta}{2} + \sqrt{\left(\frac{\theta}{2}\right)^2 + \frac{1}{\theta(1-\tau)}}\right)\right)$$. Finally, the skill investment allocation is given by

$$s(\kappa; \tau) = \left(\frac{\theta}{2} + \sqrt{\left(\frac{\theta}{2}\right)^2 + \frac{1}{\theta(1-\tau)}}\right) \cdot \kappa$$  \hspace{1cm} (46)

and, thus, the equilibrium skill density $$m(s)$$ is exponential with parameter $$\frac{1}{1-\tau} / \left(\frac{\theta}{2} + \sqrt{\left(\frac{\theta}{2}\right)^2 + \frac{1}{\theta(1-\tau)}}\right)$$.

As is clear from Proposition 7, the model with skill bias in production preserves the result that in equilibrium, the price of skills is log-linear in $$s$$. This can be anticipated from the fact that the logarithm of the marginal product of labor is now the skill bias $$\phi$$ plus the scarcity term from before (see equation (44)). As before, the upper tail of the equilibrium wage distribution is Pareto, the marginal return to skill, $$\pi_1(\tau)$$, is increasing in $$\tau$$, and the base price $$\pi_0(\tau)$$ is falling in $$\tau$$. However, the sensitivity of $$\pi_1$$ and $$\pi_0$$ to $$\tau$$ is smaller the larger is the exogenous skill bias $$\phi$$. The logic is that price differences due to $$\phi$$ is independent of the relative supply of different skill types.

The standard deviation of log skill prices is now

$$\text{std} (\log p(s; \tau)) = (1-\tau) \left(\frac{\theta}{2} + \sqrt{\left(\frac{\theta}{2}\right)^2 + \frac{1}{\theta(1-\tau)}}\right)^2.$$  \hspace{1cm} (47)

We conclude that the variance is strictly falling in $$\tau$$ whenever $$\phi > 0$$. Thus, progressive taxation reduces dispersion in pre-tax wages as well as after-tax wages. [perhaps refer to Guvenen?]

The social welfare function is unchanged up to the skill-related terms in $$\mathcal{W}$$. In the appendix we provide exact expressions for how these terms change when $$\phi > 0$$.

Note that all the model moments we used for estimation (in Section 6) are the same as before, except for the standard deviation of skill prices (47). The identification of the model parameters is therefore the same, except for the estimate of $$\theta$$. From equation (47) it is clear that skill price dispersion can stem from either a large complementarity between skills, i.e., a low $$\theta$$, or a large skill bias in production, i.e., a large $$\phi$$. We derive the possible combinations...
Figure 5: The figure shows the combinations of $\theta$ and $\varrho$ which generate a cross-sectional variance of skill prices of 0.1012.

of $(\varrho, \theta)$ by imposing $\tau_{US} = 0.151$, $\text{std}(\log p(s; \tau)) = \sqrt{0.1012}$, and $\varrho \geq 0$ in equation (47). This implies $\theta \geq 3.1$ and $\varrho \in [0, 0.61]$, and the combinations are plotted here.\footnote{Note that Assumption 1 is satisfied under these parameter values.}

We now recompute the optimal $\tau$ for these combinations of $(\varrho, \theta)$. To this end, we substitute out $\theta$ for the calibrated value of $\varrho$ in the welfare expression and calculate the optimal $\tau$ for the relevant range of $\varrho$. The optimal $\tau$ for these calibrated values of $(\varrho, \theta)$ are plotted below in figure 7.1.
Figure 6: The figure shows the optimal degree of progressivity ($\tau$) under all combinations of $\theta$ and $\rho > 0$ consistent with a cross-sectional variance of skill prices of 0.1012.

### 7.2 Optimal progressivity when past skill investment is sunk

In our baseline model we assumed fully reversible investment. This implied that (i) the optimal choice for $\tau$ is independent of the pre-existing distribution of skills, (ii) transition following any tax reform is instantaneous, and (iii) the optimal choice for $\tau$ is largely independent of the relative weights that the planner places on cohorts born before versus after the tax reform.$^{32}$

$^{32}$Our baseline social welfare function assumes the planner discounts lifetime utility for future cohorts at rate $\beta$. If we were to alternatively assume that the planner puts equal weight on all past and future cohorts, the expression for social welfare would be identical to expected lifetime utility for a newborn agent in steady state, instead of average period utility in cross-section. However, the quantitative implications of this difference are very small: the optimal $\tau$ changes from 0.062 to 0.040.
However, the assumption that skill investment is fully reversible is not very realistic. We therefore now consider the opposite extreme assumption, namely that skill investment is chosen once and for all at age zero, and can never be adjusted thereafter. This introduces an additional force in the direction of more progressivity, since by making the tax system more progressive the social planner can immediately reduce consumption inequality for all agents, without simultaneously reducing skill investments for agents who entered the economy in the past. The strength of this temptation towards greater progressivity will depend on the planner’s relative Pareto weights on current versus future generations in social welfare.

We now consider an unanticipated once-and-for-all change in the tax system, defined as permanent changes in the economy-wide values for $\tau$ and $g$. Prior to the tax reform the economy is assumed to be in a steady state defined by constant fiscal parameters $(\tau_{-1}, g_{-1})$. We evaluate welfare assuming the planner discounts future generations at rate $\gamma$.

Note that with irreversible investment transition is no longer instantaneous because, given an unanticipated change in $\tau$, output gradually evolves over time as the population share of agents who make skill investments under the new tax regime rises.

Our baseline model is tractable, because the distribution of skills is exponential. To maintain tractability in the model with irreversible investment we introduce an additional assumption on the production technology, namely that production is segregated by age. Because each cohort makes skill investments simultaneously, expecting a fixed future $\tau$, the distribution of skills within agents of any given age is then always exponential, even though the aggregate economy-wide distribution is not.

It is straightforward to show that the optimal choice for $g$ remains constant throughout transition, and equal to $\psi/(1 + \psi)$. Given a constant value for $g$ and a constant (by assumption) value for $\tau$, the budget balancing value for $\lambda$ is now time-varying during transition.

Assuming $\gamma = \beta$, the contributions to social welfare from consumption, hours and skill investment at date $t$ are now given by the following expression:
The terms involving inequality are largely familiar from our baseline social welfare expression. The key difference is in the second line, which capture the average wage at date \( t \), and reflects the fractions of the population that made skill investments before and after the tax reform.

7. When we set gamma = beta, and assume that the initial steady state corresponds to our estimated value for tau in the US (0.151) we find that the optimal new permanent choice for tau is 0.141. This is larger than the optimal tau in our baseline model with fully reversible investment (tau = 0.065). The intuition is analogous to the temptation to tax initial capital in the familiar dynamic taxation problem in the growth model.

8. Note, however, that now the optimal new tau is much more sensitive to the choice for gamma. With gamma = 0.99, the optimal new tau is 0.065, as in our baseline model. The logic is simply that as the planner becomes more patient the short term gains from exploiting sunk skill investments play a smaller role in the overall welfare calculus.

9. Note also, that the model with fully fixed skill investments is also quite extreme. Perhaps a reasonable compromise between the two models we have considered thus far would be a model in which it is impossible for agents to reduce past skill investments, but in which previous investments can be costlessly supplemented. In this hybrid model, the welfare effects of unanticipated progressivity reductions would be identical to those in the fully flexible model, while the welfare effects of progressivity increases would be identical to those in the fully fixed model. In this case, under our baseline calibration with gamma = beta, the op-
timal choice for tau here would again be tau = 0.065. The logic is that in the fully rigid model, leaving progressivity unchanged is preferred to increasing progressivity (recall that the optimal choice in that model involves a modest reduction in progressivity) while in the fully flexible model reducing tau to 0.065 is the welfare maximizing policy.

10. We conclude that reductions in tax progressivity remain welfare-improving even in models in which it is difficult to adjust past skill investments, and thus when there is a short-run temptation to increase progressivity. Moreover, our baseline optimal value for tau appears to be quantitatively robust.

11. Two directions are worth exploring for future research. First, the assumption on technology we made for the purpose of retaining closed form – that production is segregated by age – is rather unappealing, and could be relaxed at the cost of characterizing the skill distribution numerically. Second, and more substantively, it would be nice to solve the full Ramsey problem, in which the degree of progressivity is allowed to change freely over time, as opposed to changing in an unanticipated fashion once and for all. One reason this more general problem is challenging in our baseline model is that given time-varying progressivity, agents will want to borrow and lend inter-temporally.

7.3 Progressive consumption taxation

The comparison between income and consumption taxes has a long tradition in public finance and macroeconomics. The main argument set forth by the literature in favor of consumption taxes is productive efficiency. As explained by Nishiyama and Smetters (2005), a tax on consumption is an income tax where saving is exempt and, as a result, it imposes a smaller distortion on capital accumulation compared to an income tax. Correia (2010) shows that when a flat consumption tax is augmented with a universal lump-sum transfer, productive efficiency can be improved without any welfare loss in terms of higher inequality.

Here, we put forth a novel argument in favor of progressive consumption taxes.\textsuperscript{33} We show that, compared to progressive earnings taxation, progressive consumption taxation reduces distortions to the distribution of labor supply and allows stronger protection against

\textsuperscript{33}Implementation in our model would follow McCaffery’s (2002) proposal of taxing income progressively with an exemption for savings.
lifecycle earnings shocks that are privately uninsurable.

In our model, with a consumption expenditure tax/transfer scheme, the household budget constraint (15) of household \(i\) of age \(a\) becomes:

\[
\lambda_{cia}(\varepsilon)^{1/\sigma} + \delta q b_{i,a+1} = p(s_i) \exp{(a_{ia} + \varepsilon_{ia})} h_{ia} + B(\varepsilon_{ia})
\]  

(48)

where, as before, \(\tau > 0\) \((\tau < 0)\) denotes a progressive \((\text{regressive})\) scheme. Progressive consumption taxation changes the hours allocation which now becomes

\[
\log h^* (\varphi, \varepsilon; \tau) = \log H^{RA}(\tau) - \varphi + \frac{1}{\sigma} \varepsilon - \frac{1}{\sigma} \mathcal{M}(v_\varepsilon; 0).
\]  

(49)

Compared to (20), we note two important differences. First, the pass-through from insurable shocks to hours is undistorted by taxes and equal to the Frisch elasticity \(1/\sigma\), not the tax modified elasticity \(1/\sigma\). Second, also the productivity gain from insurable dispersion \(v_\varepsilon\) is unaffected by \(\tau\). The new productivity gain term \(\mathcal{M}(v_\varepsilon; 0) = \frac{1}{\sigma} \frac{v_\varepsilon}{2}\) is obtained, as the notation suggests, by simply setting \(\tau = 0\) in \(\mathcal{M}(v_\varepsilon; \tau)\).\(^{34}\) To understand these differences, recall that the consumption allocation is independent of \(\varepsilon\). Therefore, taxing consumption, instead of earnings, is a form of taxation that is specifically targeted towards the uninsurable shocks \((\text{which pass through to consumption})\) and that does not distort how hours worked respond to the insurable shocks. In sum, progressive consumption taxation is more efficient than progressive income taxation. The next proposition summarizes this result by stating the form of the social welfare function.

**Proposition 7 [welfare with progressive consumption tax].** Under a progressive consumption tax/transfer system, the social welfare function is

\[
\mathcal{W}^{cons}(g, \tau) = \mathcal{W}(g, \tau) - (1 + \chi) \left[ \frac{1}{\sigma} v_\varepsilon - \frac{1}{\sigma^2} \frac{v_\varepsilon}{2} \right] + (1 + \chi) \frac{1}{\sigma} \frac{v_\varepsilon}{2}.
\]

The productivity gain term \(\frac{1}{\sigma} v_\varepsilon\) becomes \(\frac{1}{\sigma} v_\varepsilon\) and the welfare loss from hours fluctuation \(-\frac{1}{\sigma^2} \frac{v_\varepsilon}{2}\) becomes \(\frac{1}{\sigma} \frac{v_\varepsilon}{2}\). Both new terms correspond to the old ones with \(\tau\) set to zero.

\(^{34}\)The consumption allocation is given by

\[
\log c^* (\varphi, \alpha, s; g, \tau) = \log C^{RA}(g, \tau) + (1 - \tau) \mathcal{M}(v_\varepsilon; 0) - (1 - \tau) \varphi + (1 - \tau) \log p(s; \tau) + (1 - \tau) \alpha
\]

where \(\log C^{RA}(g, \tau) = \log \lambda^*(g, \tau) + \frac{1}{(1 + \sigma)} \log(1 - \tau).\)
The degree of progressivity of consumption taxes which maximizes $W^{cons}(g, \tau)$ is 0.076. Because this tax system does not distort the distribution of labor supply across households, the planner chooses a slightly higher value for $\tau$ than under the baseline earnings tax system, thereby providing better insurance against life-cycle earnings shocks. The welfare gain associated with switching from the current earnings tax system to this progressive consumption tax system is 0.54 percent, slightly larger than the baseline welfare gain (0.49 percent).

7.4 Political-economic determination of progressivity

After our extensive characterization of optimality, it is natural to ask the following question: if $(g, \tau)$ were determined through a political-economic mechanism, how would the outcome of voting differ from the one chosen by a utilitarian government? To maintain symmetry with our normative analysis, we restrict ourselves to voting once and for all and retain the assumption that the human capital accumulation decision is reversible, so the transition to a new steady state is immediate.\(^{35}\)

The challenge in analyzing a political-economic version of our model is twofold. First, voting has two dimensions, $(g, \tau)$. Second, there are multiple sources of heterogeneity across households which, potentially, means that preferences over fiscal variables may not be single-peaked. In what follows, we show that (i) irrespective of the choice for $\tau$, every agent agrees on the amount of $G$ to be provided, and this amount is fraction $\chi / (1 + \chi)$ of aggregate output; (ii) notwithstanding multidimensional heterogeneity, the attitude of individual agents towards progressivity $\tau$ can be summarized by a single summary statistic, so voters are effectively heterogeneous in only one dimension. As a consequence, the median voter theorem applies.

We begin by proving that everyone, in our economy, agrees on the optimal size of government.

Proposition 8 [agreement on $G$]. When voting over $g$, every agent chooses $\hat{g} = g^* = \chi / (1 + \chi)$, independently of the choice for $\tau$.

\(^{35}\)As explained previously, a time-varying $\tau$ would break the no-bond-trade result, and the wealth distribution would become a relevant part of the state space. If investment were irreversible, the skill distribution would evolve gradually over time in response to a tax policy changes.
To understand this result, note that the choice of $g$ of agent $i$ obeys the first-order condition
\[
\chi \frac{1}{g} = \frac{c_i}{1 - g} \cdot \frac{1}{c_i}
\] (50)
The left hand side is the benefit from a marginal increase in the share of output devoted to publicly-provided goods, which, given separable preferences, is identical across agents. The right hand side is the cost associated to a marginal increase in $g$. Since $c_i(g, \tau) = \lambda(g) \bar{c}_i(\tau) = (1 - g) \bar{\Lambda} \bar{c}_i(\tau)$ where $\bar{\Lambda}$ and $\bar{c}_i(\tau)$ are independent of $g$, the derivative of individual consumption with respect to $g$ is (minus) the first term in the right-hand side of (50). The second term is the marginal utility of private consumption. The key to the result in Proposition 8 is that also the cost of increasing $g$ is the same for every individual. A low productivity (and low earnings) agent has a high marginal utility of private consumption, but also has little to gain from reducing average tax rates. The two effects exactly cancel out, and all agents agree on the optimal choice for $g$.

We now characterize households’ preferences over $\tau$. Here, we make an additional simplifying assumption: voting occurs before the realization of the inurable shock $\varepsilon$. The assumption is rather innocuous since $\varepsilon$ is i.i.d. and we should not expect its value to affect much the choice of progressivity forever after. The individual state vector relevant for a voter is, therefore, $(\varphi, \alpha, \kappa)$.\textsuperscript{36}

**Proposition 9 [Median voter’s preferences over $\tau$].** The preferences of the median voter over $\tau$ are given by
\[
U^{med}(g, \tau) = \mathcal{U}(g, \tau) - (1 - \tau) \left( \frac{\beta \delta}{1 - \beta \delta} \frac{v_\omega}{2} - (1 - \tau) \left( - \frac{\delta}{1 - \delta} \frac{v_\omega}{2} - \frac{v_\varphi}{2} + \frac{1}{2\theta} \right) + (1 - \tau) x^{med} \right)
\] (51)
where $x^{med}$ is the median of $x = \alpha - \varphi + \frac{n}{2\theta} \kappa$, an Exponentially-Modified Gaussian random variable.

The first important result is that, even though voters differ in the three-dimensional vector $(\varphi, \alpha, \kappa)$, their heterogeneity can be collapsed into the new random variable $x$ which is a

\textsuperscript{36}The variable $\kappa$ is a state because we let agents reoptimize their choice of human capital after the change in $\tau$.\textsuperscript{45}
linear combination of Normal \((\varphi, \alpha)\) and Exponential \((\kappa)\) variables, and hence is distributed as an Exponentially-Modified Gaussian (EMG) variable.

We find that the median voter (i.e. the agent with the median value for \(x\)) would choose \(\tau^{med} = 0.083\). The difference relative to the utilitarian planner’s choice \(\tau^* = 0.062\) is not large: for example, the implied marginal rate tax at the average income level is only 1.5 percentage points higher in the political-economic equilibrium.

To understand the difference between the two values, note first that in the limiting case \(\theta \to \infty\) and \(v_\omega = 0\), \(U^{med}(g, \tau) = W(g, \tau)\) and thus the median voter would choose exactly the same degree of progressivity as the utilitarian planner. The logic for this result is that the idiosyncratic preference term appears in individual log consumption in the form \(- (1 - \tau) \varphi\) (see equation 21). Given a symmetric normal distribution for \(\varphi\) and utility that is logarithmic in consumption, average utility is then equal to utility of the median \(\varphi\) agent. Similar logic would apply to fixed heterogeneity in initial labor productivity \(\alpha_0\). Thus the reason the median voter prefers a higher value for \(\tau\) has to do with the existence of permanent uninsurable shocks \((v_\omega > 0)\) and heterogeneity in skill prices \((\theta < \infty)\). In regard to the former, the median voter wants a higher \(\tau\) as insurance against future uninsurable shocks (the second term on the right-hand side of equation ??) but, unlike the planner, does not care about the fact that future generations will enter with zero initial dispersion in the uninsurable component of productivity. In regard to skill heterogeneity, the agent with median ability \(\kappa\) has less than average ability, because \(\kappa\) is exponentially distributed. Thus the median \(\kappa\) agent has more to gain from progressive taxation, and would choose a higher value for \(\tau\).

8 Conclusions

——— FOR THE APPENDIX:

Welfare expression in Section 7.1

In the case of skill bias in production \((\varrho > 0)\), the terms in the equally-weighted social welfare function that involve skills are as follows,
\[
(1 + \chi) \left( -\ln \left( (1 - \tau) \left( \frac{\phi}{2} + \sqrt{\frac{\phi}{2}^2 + \frac{1}{\theta (1 - \tau)}} \right) \right) - \frac{\theta}{\theta - 1} \ln \left( \frac{(\theta - 1)}{\theta} \frac{1}{(1 - \tau) \left( \frac{\phi}{2} + \sqrt{\frac{\phi}{2}^2 + \frac{1}{\theta (1 - \tau)}} \right) \right) \right)
\]

log of productivity: \( \ln(Y/N) \)

\[
+ \ln \left( 1 - (1 - \tau)^2 \left( \frac{\phi}{2} + \sqrt{\left( \frac{\phi}{2} \right)^2 + \frac{1}{\theta (1 - \tau)}} \right)^2 \right) + (1 - \tau)^2 \left( \frac{\phi}{2} + \sqrt{\left( \frac{\phi}{2} \right)^2 + \frac{1}{\theta (1 - \tau)}} \right)^2
\]

utility cost from consumption dispersion across skill groups

\[
- \frac{1}{2} (1 - \tau)^2 \left( \frac{\phi}{2} + \sqrt{\left( \frac{\phi}{2} \right)^2 + \frac{1}{\theta (1 - \tau)}} \right)^2
\]

avg. education cost

Note: in order for the social welfare to be finite (i.e., larger than \(-\infty\)) it must be the case that consumption dispersion is not too large. In particular, it must be the case that the term in the logarithm (in the second line) is positive, i.e., a lower bound on \( \tau \) given \( \theta \) and \( \varrho \):

\[
0 < 1 - (1 - \tau)^2 \left( \frac{\phi}{2} + \sqrt{\left( \frac{\phi}{2} \right)^2 + \frac{1}{\theta (1 - \tau)}} \right)^2 \Rightarrow \\
\tau > 1 - \frac{1}{\theta + \varrho}
\]

Finally, consider the special case when \( \theta \to \infty \), in which case the terms in the welfare function become

\[
(1 + \chi) \left( -\ln \left( (1 - \tau) \varrho - \ln \left( \frac{1}{(1 - \tau) \varrho - \varrho} \right) \right) \right)
\]

log of productivity: \( \ln(Y/N) \)

\[
+ \ln \left( 1 - (1 - \tau)^2 \varrho^2 + (1 - \tau)^2 \varrho^2 \right)
\]

utility cost from consumption dispersion across skill groups

\[
- \frac{1}{2} (1 - \tau)^2 \varrho^2
\]

avg. education cost
Appendix

This Appendix proves all the results in the main body of the paper.

A Proof of Proposition 1

To simplify the exposition, here we conjecture that the consumption allocation for an agent with state $(\phi, \alpha, \varepsilon, s)$ has the form

$$ c(\phi, \alpha, s; g, \tau) = \exp \left[ c(g, \tau) + (1 - \tau) \alpha + f(\phi, s; \tau) \right] $$  \hspace{1cm} (A1)

where $c(g, \tau)$ is a constant that does not depend on any individual state variables, and $f$ is a common function of individual age-invariant states $(\phi, s)$. We show that this class of allocations implies that agents do not desire to trade the bond. In the proof of the next proposition, we show that, under the guess of no bond-trade, the allocations have that form, which together with the result proved here verifies the guess.

To prove absence of bond trading, consider first the marginal rate of substitution for an agent between state $z = (\alpha, \varepsilon)$ and $z' = (\alpha', \varepsilon')$

$$ MRS_{z, z'} = \beta \frac{\exp \left[ c(g, \tau) + (1 - \tau) \alpha + f(\phi, s; \tau) \right]}{\exp \left[ c(g, \tau) + (1 - \tau) \alpha' + f(\phi, s; \tau) \right]} = \exp \left[ - (1 - \tau) \omega \right]. $$

The expected marginal rate of substitution between is

$$ E_{z' | z} [MRS_{z, z'}] = \beta \mathbb{E} [\exp (- (1 - \tau) \omega)] = \beta \exp \left[ (1 - \tau) \left[ (1 - \tau) + 1 \frac{\nu_{\omega}}{2} \right] \right] $$

which is common across all agents. As a result, there are no gains from trading a non-state contingent bond across agents with different individual states $(\phi, \alpha, \varepsilon, s)$. The bond price that supports this equilibrium is, precisely

$$ q = \beta \exp \left[ (1 - \tau) \left[ (1 - \tau) + 1 \frac{\nu_{\omega}}{2} \right] \right] $$

as stated in Proposition 1 (in log terms).

A Proof of Proposition 2

We follow the line of proof in Heathcote, Storesletten and Violante (2013). We first guess that there is no-bond trade in equilibrium. Given no bond trade across agents, the only security that is traded is claims against $\varepsilon$ shocks, and markets are complete. Without loss of generality, we can therefore think of our economy as an “island economy” where each island is populated by agents indexed by their fixed effects $(\phi, s)$ and their uninsurable wage

48
component \(\alpha\). On each island, there are complete markets with respect to \(\varepsilon\) so the competitive equilibrium allocation can be computed as the outcome of an island-specific social planner problem. Since agents on an island are ex-ante identical, the planner weights must be equal for all agents. Moreover, since each island by assumption transfers zero net financial wealth between periods and preferences are time-separable, the island-specific planner problem is static.

The island planner’s problem, taking the aggregate fiscal variables \((G, \lambda, \tau)\) and the skill price \(p(s)\) as given is

\[
\max_{\{c(\varepsilon), h(\varepsilon)\}} \int_E \left\{ \log c(\varepsilon) - \frac{\exp[(1 + \sigma) \varphi]}{1 + \sigma} h(\varepsilon)^{1+\sigma} + \chi \ln G \right\} dF_\varepsilon
\]

subject to the resource constraint

\[
\int_E c(\varepsilon) dF_\varepsilon = \lambda \int_E \exp [(1 - \tau) (p(s) + \alpha + \varepsilon)] h(\varepsilon)^{1-\tau} dF_\varepsilon \quad (A2)
\]

After taking the first-order conditions of this problem, and substituting in the resource constraint, one easily obtains the allocations of Proposition 2. The consumption allocation (20) has the form in (A1), which confirms the no bond trade guess.

A Proof of Proposition 3 and Corollaries 3.1 and 3.2

Recall, from equation (13) in the main text that the optimality condition for skill investment for agent \(i\) is

\[
\frac{1}{\kappa \mu} s = (1 - \beta \delta) \mathbb{E}_0 \sum_{a=0}^{\infty} (\beta \delta)^a \frac{\partial u(c(\varphi, \alpha, s; g, \tau), h(\varphi, \varepsilon; \tau), g)}{\partial s} \quad (A3)
\]

The skill level \(s\) affects only the consumption allocation (and not the hours allocation) through its price \(p(s)\), fixed over time, and hence (13) can be simplified as

\[
\frac{1}{\kappa \mu} s = (1 - \tau) \frac{\partial \log p(s)}{\partial s}.
\]

We now guess that the log-price function has the form

\[
\log p(s) = \pi_0 + \pi_1 s \quad (A4)
\]

which implies that the skill allocation has the form

\[
s = (1 - \tau) \mu \pi_1 \cdot \kappa. \quad (A5)
\]
and therefore, since the exponential distribution is closed under scaling, the skill density $m(s)$ inherits the exponential shape from $\kappa$, with parameter $\zeta \equiv \frac{\eta}{(1-\tau)\mu \pi_1}$, i.e., $m(s) = \zeta \exp(-\zeta s)$.

We now turn to the production side of the economy. Effective hours worked of type $s$ are given by

$$N(s) = N = \int \exp(\alpha_i + \varepsilon_i) h_i di = (1 - \tau)^{\frac{1}{1+\sigma}} \exp \left( \frac{1}{\sigma^2} (\tau + \hat{\sigma} + \hat{\sigma} \tau) \frac{v_\varepsilon}{2} \right)$$

and note that, since the hours allocation is independent of skill levels, so is $N$. In particular, by integrating the hours allocation with respect to $\left( \phi, \varepsilon \right)$ we obtain that average hours are

$$H = (1 - \tau)^{\frac{1}{1+\sigma}} \exp \left( \frac{\tau (1 + \hat{\sigma}) v_\varepsilon}{\sigma^2} \right).$$

Aggregate output is

$$Y = N \cdot \left[ \int_0^{\infty} m(s)^{\frac{\theta-1}{\theta}} ds \right]^{\frac{\theta}{\theta-1}}$$

and, therefore, independent of the dispersion in uninsurable risk $v_\alpha$.

The (log of the) hourly skill price $p(s)$ is the (log of the) marginal product of an extra effective hour supplied by a worker with skill $s$, or

$$\log p(s) = \log \left[ \frac{\partial Y}{\partial (Nm(s))} \right] = \frac{1}{\theta} \log \left( \frac{Y}{N} \right) - \frac{1}{\theta} \log [m(s)]. \quad \text{(A6)}$$

Since $Y/N$ is independent of $s$ and $\frac{1}{\theta} \log [m(s)]$ is linear in $s$, we obtain that $\pi_1 = \frac{1}{\theta} \left( \frac{\eta}{\mu (1 - \tau) \pi_1} \right)$ which verifies our guess and yields

$$\pi_1 = \sqrt{\frac{\eta}{\theta \mu (1 - \tau)}}. \quad \text{(A7)}$$

From our guess (A5) and (A6) one can easily derive that the base skill price is

$$\pi_0 = \frac{-1}{2(\theta - 1)} \left( \log \left( \frac{\eta}{\mu} \right) - \log (1 - \tau) + \log (\theta) \right) + \frac{1}{(\theta - 1)} \ln \left( \frac{\theta}{\theta - 1} \right) \quad \text{(A8)}$$

The implied equilibrium skill density has the exponential form:

$$m(s) = \sqrt{\frac{\eta \theta}{(1 - \tau) \mu}} \exp \left( -\sqrt{\frac{\eta \theta}{(1 - \tau) \mu}} s \right). \quad \text{(A9)}$$

Finally, one can solve for aggregate output:

$$Y = N \cdot \left[ (1 - \tau)^{\frac{1}{\theta(\theta-1)}} \left( \frac{\eta \theta}{\mu} \right)^{\frac{1}{\theta(\theta-1)}} \cdot \left( \frac{\theta}{\theta - 1} \right)^{\frac{\theta}{\theta-1}} \right] \quad \text{(A10)}$$
We now report a number of useful results about the distribution of skill prices, discussed in the main text, which can be easily derived.

The mean of log skill prices is
\[
E[\log p(s)] = \int (\pi_0 + \pi_1 s) \cdot m(s) \, ds = \pi_0 + \frac{1}{\theta},
\]
and the variance of log skill prices is
\[
\text{var}[\log p(s)] = E \left[ (\log p(s) - E[\log p(s)])^2 \right] = E \left[ \left( \pi_1 s - \frac{1}{\theta} \right)^2 \right] = \frac{1}{\theta^2}.
\]

Therefore, the log skill price is a shifted exponential distribution with parameter \(\theta\) whose support starts at \(\pi_0\), i.e., \(m(\log p(s)) = \theta \exp[-\theta (\log p(s) - \pi_0)]\). This density can be written as \(\exp(\pi_0) \theta \exp[-\theta (\log p(s))]\). The exponent of such random variable is Pareto with parameter \(\theta\) and minimum \(\exp(\pi_0)\). Hence, the skill price and, as a result, the income distribution is Pareto with parameter \(\theta\).

**A Proof of Corollary 3.3**

First best allocations in the representative agent economy are the solutions to the following planner’s problem:

\[
\max_{C, g, H} \left\{ \log C - \frac{H^{1+\sigma}}{1+\sigma} + \chi \log G \right\}
\]
subject to
\[
G = gH, \\
C = (1 - g) \cdot H.
\]

which yields the following solutions for \(g\) and \(H\):
\[
g = \frac{\chi}{1 + \chi}, \\
H = (1 + \chi)^{\frac{1}{1+\sigma}}.
\]

Substituting the candidate solution \(\tau_{RA}^* = -\chi\) into the competitive equilibrium allocation (equation 18) gives \(H_{RA}^* (\tau_{RA}^*) = (1 + \chi)^{\frac{1}{1+\sigma}}\) which is the first best value for hours.
A Proof of Lemma 1

To derive this result, we aggregate, one by one, period utility at ages 0, 1, ... for all existing and future cohorts. Let $u_0$ period utility at age 0 and $W_0$ be the age 0 component of the social welfare function. Then:

$$W_0 = \frac{1 - \beta}{1 - \beta \delta} \cdot (1 - \delta) \cdot (1 + \beta + \beta^2 + ...) \cdot (1 - \beta \delta) \mathbb{E} [u_0] = (1 - \delta) \cdot \mathbb{E} [u_0]$$

where the term $(1 - \beta) / (1 - \beta \delta)$ is an innocuous scaling of the welfare function that simplifies the final expression, $(1 - \delta)$ is the size of the population at age zero, $(1 + \beta + \beta^2 + ...)$ reflects the weights the planner puts on current and future cohorts of age zero, and $(1 - \beta \delta)$ is the scaling variable in expected utility $().$

The age 1 component is given by:

$$W_1 = \frac{1 - \beta}{1 - \beta \delta} \cdot (1 - \delta) \cdot (1 + \beta + \beta^2 + ...) \cdot (1 - \beta \delta) \cdot \mathbb{E} [u_1] = (1 - \delta) \delta \mathbb{E} [u_1]$$

where the term $(1 - \delta) \delta$ is the size of the population at age one. And so on. Now we need to compute how education costs factor into social welfare. If every agent is allowed to revert her decision every period, then for the existing cohorts we have

$$W_\kappa = \frac{1 - \beta}{1 - \beta \delta} \cdot \left\{ (1 - \delta) \sum_{a=1}^{\infty} \delta^n \mathbb{E} \left[ \frac{1}{\kappa} \frac{s^2}{2 \mu} - \frac{1}{\kappa} \frac{s_0^2}{2 \mu} \right] + (1 - \delta) \left(1 + \beta + \beta^2 + ...\right) \mathbb{E} \left[ \frac{1}{\kappa} \frac{s^2}{2 \mu} \right] \right\}$$

$$= \frac{1 - \beta}{1 - \beta \delta} \cdot \left\{ -\delta \mathbb{E} \left[ \frac{1}{\kappa} \frac{s_0^2}{2 \mu} \right] + \delta \left(1 + \beta + \beta^2 + ...\right) \mathbb{E} \left[ \frac{1}{\kappa} \frac{s^2}{2 \mu} \right] \right\}$$

$$= -\delta \frac{1 - \beta}{1 - \beta \delta} \mathbb{E} \left[ \frac{1}{\kappa} \frac{s_0^2}{2 \mu} \right] + \mathbb{E} \left[ \frac{1}{\kappa} \frac{s^2}{2 \mu} \right]$$

The first row accounts for the fact that, since skills are reversible, the existing generations can adjust their level of $s$. The term in $s_0$ in the last row is irrelevant for welfare, since it reflects past decisions, and can be omitted.

Adding up all the components $\sum_{a=1}^{\infty} W_a + W_\kappa$ yields the social welfare function in $()$ in the main text.

A Proof of Proposition 4

We prove this proposition in two steps. First, we show how to derive a closed-form solution for the residual fiscal variable $\lambda$. Second, we substitute the allocations in the social welfare function $()$ and show how to obtain expression $()$. 

52
Step 1. If we let \( \tilde{Y} = \int y_i^{1-\tau} di \), we have

\[
\lambda = \frac{(1-g)Y}{\bar{Y}}.
\] (A11)

To compute \( \tilde{Y} \), it is useful to aggregate by age group. Let

\[
\tilde{Y}^a = \int (y(s, \varphi, \epsilon, \alpha))^{1-\tau} m(s) ds dF^a d\varphi d\epsilon = \int (h(\epsilon) \exp(p(s) + \alpha_a + \epsilon))^{1-\tau} m(s) ds dF^a d\varphi d\epsilon.
\]

Substituting the hours allocation in (1), the expression for the skill price in (1), the density function \( m(s) \) in (A9) and integrating we arrive at

\[
\tilde{Y}^a = (1-\tau)^{\frac{\tau}{\bar{\sigma}}} \exp\left(-\frac{1}{\sigma} \mathcal{M}\right) \exp\left(\left(\frac{(1-\tau)(1+\bar{\sigma})}{\bar{\sigma}} - 1\right) \frac{\nu_\varphi}{2}\right) \exp((1-\tau)\pi_0) \exp(-1).
\]

Now we sum across age groups to obtain

\[
\tilde{Y} = (1-\delta) \sum_{a=0}^{\infty} \delta^a \tilde{Y}^a = -\frac{\theta}{(1-\tau) - \theta} (1-\tau)^{\frac{\tau}{\bar{\sigma}}} \exp\left(-\frac{1}{\sigma} \mathcal{M}\right) \exp\left(\left(\frac{(1-\tau)(1+\bar{\sigma})}{\bar{\sigma}} - 1\right) \frac{\nu_\varphi}{2}\right)
\cdot (1-\delta) \sum_{a=0}^{\infty} \delta^a \exp\left(-\tau (1-\tau) \frac{v_\alpha}{2}\right)
\]

Note that:

\[
\sum_{a=0}^{\infty} \delta^a \exp\left(-\tau (1-\tau) \frac{v_\alpha}{2}\right) = \frac{\exp\left(-\tau (1-\tau) \frac{v_\alpha^0}{2}\right)}{1 - \delta \exp(\nu_\omega)}
\]

Therefore, we conclude that

\[
\tilde{Y} = -(1-\tau)^{\frac{\tau}{\bar{\sigma}}} \exp\left(-\frac{1}{\sigma} \mathcal{M}\right) \exp\left(\left(\frac{(1-\tau)(1+\bar{\sigma})}{\bar{\sigma}} - 1\right) \frac{\nu_\varphi}{2}\right)
\times \exp((1-\tau)\pi_0) \exp\left(-\tau (1-\tau) \frac{v_\alpha^0}{2}\right)
\times \frac{(1-\delta) \exp\left(-\tau (1-\tau) \frac{v_\alpha}{2}\right)}{1 - \delta \exp\left(-\tau (1-\tau) \frac{v_\alpha}{2}\right)} \times \frac{\theta}{(1-\tau) - \theta}
\]

where, recall that

\[
\pi_0 = \frac{-1}{2(\theta - 1)} \left(\log \left(\frac{\eta}{\mu}\right) - \log(1-\tau) + \log(\theta)\right) + \frac{1}{(\theta - 1)} \ln \left(\frac{\theta}{\theta - 1}\right)
\]

\[
\mathcal{M} = \frac{(1-\tau)(1-\tau(1+\bar{\sigma}))}{\bar{\sigma}} \frac{v_\varphi}{2}
\]

53
Substituting (A10) and (A12) into (A11), and simplifying, we arrive at a solution for the equilibrium value of $\lambda$ which, in logs, is:

$$\log \lambda = \ln(1 - g) + \frac{\tau (1 - \tau)}{\sigma + \tau} \left( \frac{1 + \sigma}{\sigma + \tau} + 2 + \sigma \right) \frac{v_\varphi}{2} + \frac{\tau}{1 + \sigma} \log (1 - \tau) + \frac{\tau (1 - \tau)}{2} \frac{v_\varphi}{2}$$

$$- \frac{1}{\sigma} \log (1 - \delta) + \tau (1 - \tau) \frac{v_\theta}{2} + \log \left[ 1 - \delta \exp \left( \frac{-\tau (1 - \tau)}{2} v_\omega \right) \right]$$

$$+ \frac{1}{2} \frac{\tau}{\theta - 1} \log \left( \frac{1 - \tau}{\theta} \right) + \frac{1}{2} \frac{\tau}{\theta - 1} \log \left( \frac{\mu}{\eta} \right) + \frac{\theta - 1 + \tau}{\theta - 1} \log \left( \frac{\theta}{\theta - 1} \right) + \log \left( \frac{\theta - 1 + \tau}{\theta} \right)$$

Step 2. Substituting the equilibrium allocations into period utility at age $a \geq 0$, we have

$$u(c_a, h, G) = \log \lambda + \frac{1 - \tau}{1 + \sigma} \log (1 - \tau) - (1 - \tau) \varphi + (1 - \tau) \alpha_a + \frac{1 - \tau}{2 (\theta - 1)} \left( \log (1 - \tau) + \log \theta - \log \left( 1 - \frac{\tau}{\theta - 1} \right) \right)$$

$$+ \mathcal{M} - \exp \left( -\frac{1 + \sigma}{\sigma (1 - \tau)} \mathcal{M} \right) \exp \left( \frac{1 + \sigma}{\sigma} \varepsilon \right) \frac{(1 - \tau)}{(1 + \sigma)} + (1 - \tau) \frac{\eta}{\theta} + \chi \ln G$$

The disutility cost at $a = 0$ from investing in education is

$$v(s(\kappa)) = -\frac{1}{\kappa} \frac{s^2}{2 \mu} = -\frac{\kappa \eta (1 - \tau)}{2 \theta}$$

(A14)

Average cross-sectional welfare at age $a$ is

$$\int u(c_a, h, G) dF_\kappa dF_\varphi dF_\alpha = \log \lambda + \frac{1 - \tau}{1 + \sigma} \log (1 - \tau) - (1 - \tau) \frac{v_\varphi}{2} - (1 - \tau) \frac{v_\alpha}{2} + \frac{1 - \tau}{2 (\theta - 1)} \left( \log (1 - \tau) + \log \theta - \log \left( 1 - \frac{\tau}{\theta - 1} \right) \right)$$

$$+ \mathcal{M} - \frac{1 - \tau}{1 + \sigma} + \frac{1 - \tau}{\theta} + \chi \log (gY)$$

where, it is useful to note that

$$\int \exp \left( -\frac{1 + \sigma}{\sigma (1 - \tau)} \mathcal{M} \right) \exp \left( \frac{1 + \sigma}{\sigma} \varepsilon \right) = 1$$

$$\int (1 - \tau) \frac{\eta}{\theta} dF_\kappa = \frac{1 - \tau}{\theta}.$$
Finally, the average cost of education is obtained by integrating (A14)

\[ \int v(s(\kappa)) \, dF_\kappa = \int \frac{1}{\kappa} \frac{\kappa^2 \eta \mu^{1-\tau}}{2\mu} dF_\kappa = \frac{1-\tau}{2\theta} \]

Social welfare is, therefore,

\[ W = (1-\delta) \int \delta^\alpha u(c, h, G) \, dF_\kappa dF_\varepsilon dF_\phi dF_\alpha - \int v(s(\kappa)) \, dF_\kappa \]

\[ = \log \lambda + \frac{1-\tau}{1+\sigma} \log (1-\tau) - (1-\tau) \frac{v_\varepsilon}{2} - (1-\tau) \frac{v_\alpha}{2} + \frac{1-\tau}{2(\theta-1)} \left( \log(1-\tau) + \log \theta - \log \left( \frac{\eta}{\mu} \right) \right) \]

\[ + \mathcal{M} - \frac{1-\tau}{1+\sigma} + \frac{1-\tau}{2\theta} + \chi \log (gY) \]

Substituting in the expression for \( \lambda \) in (A15), the expression for aggregate output in (A10) and rearranging, we arrive at the expression in Proposition 4.

Differentiating the expression for social welfare twice with respect to \( \tau \) shows that welfare of the representative agent \( W_{RA} \) is concave in \( \tau \) and so are (i) the productivity gain from skill investment net of the education cost, and (ii) the welfare components due to uninsurable wage risk and preference heterogeneity. It can also be derived that the term in insurable risk has second derivative equal to

\[ - (1+\chi) \frac{\sigma - 2\tau}{(\sigma+\tau)^2} \frac{(1+\sigma)^2}{2} v_\varepsilon \]

which is less than or equal to zero if \( \sigma \geq 2 \). The latter is, therefore, a sufficient condition for global strict concavity of the social welfare function.

## A Proof of Corollary 4.1

The planner’s problem in the representative agent economy is

\[ \max_{C,H,g} \log C - \frac{H^{1+\sigma}}{1+\sigma} + \chi \log (g \cdot Y) \]

\[ \text{s.t.} \]

\[ Y = H \]

\[ C = (1-g)Y \]

where the first constraint is the production technology and the second is the aggregate feasibility constraint. After substituting these two constraints into the objective function, it is easy to derive that \( g = \chi / (1+\chi) \) and \( H = (1+\chi)^{1+\sigma} \). Comparing these planner’s choices with the competitive equilibrium allocations derived in Section XYZ, it follows that the value \( \tau = -\chi \) implements the first best.
A Proof of Proposition 5

The derivative of social welfare (equation 27) with respect to \( \tau \) is

\[
\frac{\partial W}{\partial \tau} = -\frac{(1 + \chi)}{2(\theta - 1)} \cdot \frac{1}{(1 - \tau)} + \frac{1}{\theta - 1 + \tau} - \frac{1}{2\theta} \cdot \frac{(\tau + \chi)}{(\sigma + 1)(1 - \tau)} + 2(1 - \tau) \left[ \frac{v_\varphi}{2} + \frac{v_\alpha}{2} \right] - 2 \tau \frac{(1 + \sigma)^2}{(\sigma + \tau)^3} v_\varepsilon (1 + \chi)
\]

We want to sign this derivative at \( \tau = 0 \) in order to ascertain whether a marginal increase in progressivity is welfare-improving.

\[
\frac{\partial W}{\partial \tau} \bigg|_{\tau=0} = -\frac{1 + \chi}{2(\theta - 1)} + \frac{1}{\theta - 1} - \frac{1}{2\theta} \cdot \frac{\chi}{\sigma + 1} + v_\varphi + v_\alpha
\]

It is immediate that this derivative is positive if and only if the condition in Proposition 5 is satisfied.

A Proof of Proposition 6

A Proof of Proposition 7

To obtain the allocations when the government uses progressive consumption taxation, it suffices to follow the steps in the Proof of Proposition 2, with the difference that the island-level resource constraint (A2) now becomes

\[
\lambda \int_{E} \epsilon^{\frac{1-\tau}{1-\gamma}} dF_\epsilon = \int_{E} \exp \left[ (p(s) + \alpha + \varepsilon) \right] h_\epsilon(\varepsilon) dF_\epsilon.
\]

(A17)

With the allocations in hand, one can replicate all the steps of the proof of Proposition 4 to obtain the new social welfare function.

A Proof of Proposition 8

The expected utility of an individual with state \((\varphi, \alpha, \kappa, \varepsilon)\) can be written as

\[
U(\varphi, \alpha, \kappa, \varepsilon; g, \tau) = -\frac{1}{\kappa} \frac{s^*(\kappa; \tau)^2}{2\mu} + (1 - \beta\delta) \sum_{a=0}^{\infty} (\beta\delta)^a \left[ \log c^*(\varphi, \alpha, s^*(\kappa; \tau); g, \tau) - \exp((1 + \sigma)\varphi) \frac{h^*(\varphi)}{1 - \beta\delta} \right]
\]

(A18)
where we have made explicit the fact that the choice of $s^*$ and aggregate output $Y$ are independent of $g$. We can rewrite expected utility as

$$U(\varphi, \alpha, \kappa; g, \tau) = -\frac{1}{\kappa_i}\frac{s^*(\kappa; \tau)^2}{2\mu} + \log c^*(\varphi, \alpha, s^*(\kappa; \tau); g, \tau) - \frac{\beta \delta}{1 - \beta \delta} \cdot \frac{\nu_\omega}{2} \cdot \exp((1 + \sigma) \varphi) \cdot \frac{h^*(\varphi, \varepsilon; \tau)}{1 + \sigma}$$

where we have used the fact that $g$ affects $c^*$ only through the residual fiscal variable $\lambda^*$. From (), we can write $\lambda^*(g) = (1 - g) \bar{A}$ where $\bar{A}$ is independent of $g$. Moreover, from the consumption allocation we obtain that $c^*(\varphi, \alpha, s^*(\kappa; \tau); g, \tau) = \lambda^*(g) \bar{c}(\varphi, \alpha, s^*(\kappa; \tau); \tau)$.

Rearranging (A19) yields

$$\frac{\bar{c}(\varphi, \alpha, s^*(\kappa; \tau); \tau)}{c^*(\varphi, \alpha, s^*(\kappa; \tau); g, \tau)} \left(\frac{\lambda^*(g)}{1 - g}\right) = \frac{\chi}{g}$$

which is the desired result.

### A Proof of Proposition 9

Substituting the allocations $s^*(\kappa; \tau)$, $h^*(\varphi, \varepsilon; \tau)$ and $c^*(\varphi, \alpha, s^*; g, \tau)$ into expected utility (A18) yields

$$U(\varphi, \alpha, \kappa, \varepsilon; g, \tau) = -(1 - \tau) \frac{\eta \kappa}{2\theta} + (1 - \beta \delta) \mathbb{E} \sum_{a=0}^{\infty} (\beta \delta)^a \left\{ \log \lambda^*(g, \tau) + \frac{\log(1 - \tau)}{1 + \delta} + \mathcal{M} + (1 - \tau) \left( \alpha_1 - \frac{(1 - \tau)}{1 + \sigma} \exp \left( -\frac{1 + \sigma}{\delta (1 - \tau)} \mathcal{M} + \frac{1 + \sigma}{\delta} \varepsilon \right) + \chi \log G \right) \right\}$$

Now, suppose that the choice of $\tau$ is made before observing $\varepsilon$. Then, the term in the second row becomes $-\frac{(1 - \tau)}{1 + \sigma}$ and $U$ does not depend on $\varepsilon$ any longer, since $\varepsilon$ is an i.i.d. random variable. Rearranging:

$$U(\varphi, \alpha, \kappa; g, \tau) = \log \lambda^*(g, \tau) + \frac{\log(1 - \tau)}{1 + \delta} + \mathcal{M} + \frac{1 - \tau}{2(\theta - 1)} \left( \ln(1 - \tau) + \ln \theta - \ln \left( \frac{\eta}{\mu} \right) - 2 \ln(\theta - 1) \right)$$

$$- (1 - \tau) \varphi + (1 - \tau) \kappa \frac{\eta}{2\theta}$$

$$+ (1 - \beta \delta) \mathbb{E} \left[ \sum_{a=0}^{\infty} (\beta \delta)^a (1 - \tau) \alpha_a \right]$$
Solving for the last term gives

\[
U(\varphi, \alpha, \kappa; g, \tau) = \log \lambda^*(g, \tau) + \frac{\log(1 - \tau)}{(1 + \sigma)} + \mathcal{M} + \frac{(1 - \tau)}{2(\theta - 1)} \left( \ln(1 - \tau) + \ln \theta - \ln \left( \frac{\eta}{\mu} \right) - 2 \ln(\theta - 1) \right)
- (1 - \tau) \varphi + (1 - \tau) \kappa \frac{\eta}{2\theta}
+ (1 - \tau) \left[ \alpha - \beta \delta \frac{v_\omega}{2} \right]
\]

Recall that the baseline social welfare function is

\[
W = \ln \lambda^*(g, \tau) + \frac{\log(1 - \tau)}{(1 + \sigma)} + \mathcal{M} + \frac{(1 - \tau)}{2(\theta - 1)} \left( \ln(1 - \tau) + \ln \theta - \ln \left( \frac{\eta}{\mu} \right) - 2 \ln(\theta - 1) \right) - \frac{(1 - \tau)}{(1 + \sigma)}
- (1 - \tau) \varphi + \frac{(1 - \tau)v_\omega}{2\theta}
- (1 - \tau) \left( \frac{\delta - v_\omega}{1 - \delta} \right)
\]

Therefore, we can express expected utility as

\[
U(\varphi, \alpha, \kappa; g, \tau) = W(\tau) + (1 - \tau) \frac{v_\omega}{2} - \frac{(1 - \tau)}{2\theta}
+ (1 - \tau) \left[ \frac{\delta - v_\omega}{1 - \delta} - \frac{\beta \delta - v_\omega}{2} \right]
+ (1 - \tau) \left[ \alpha - \varphi + \frac{\kappa \eta}{2\theta} \right]
\]

Note that since \( W(\tau) \) is strictly concave in \( \tau \), so is \( U(\varphi, \alpha, \kappa; g, \tau) \) because the additional terms are linear.

We need to determine the median voter. A useful property is that the three individual states \((\alpha, \varphi, \kappa)\) enter as a linear combination. Let

\[
x = \alpha - \varphi + \frac{\eta}{2\theta} \kappa
\]

then, all we need is to determine the median value of \( x \). Since \( \alpha \) and \( \varphi \) are normally distributed and \( \kappa \) is exponentially distributed, the random variable \( z \) is an Exponentially Modified Gaussian distribution.