



Optimal Tax Progressivity: An Analytical Framework

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

Staff Report 496

Revised December 2016

Keywords: Tax progressivity; Income distribution; Skill investment; Labor supply; Partial insurance;

Government expenditures; Welfare; Cross-country evidence

JEL classification: D30, E20, H20, H40, J22, J24

The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

Federal Reserve Bank of Minneapolis • 90 Hennepin Avenue • Minneapolis, MN 55480-0291

<https://www.minneapolisfed.org/research/>

OPTIMAL TAX PROGRESSIVITY: AN ANALYTICAL FRAMEWORK*

Jonathan Heathcote

Kjetil Storesletten

Giovanni L. Violante

THIS VERSION: DECEMBER 21, 2016

Abstract

What shapes the optimal degree of progressivity of the tax and transfer system? On the one hand, a progressive tax system can counteract inequality in initial conditions and substitute for imperfect private insurance against idiosyncratic earnings risk. On the other hand, progressivity reduces incentives to work and to invest in skills, distortions that are especially costly when the government must finance public goods. We develop a tractable equilibrium model that features all of these trade-offs. The analytical expressions we derive for social welfare deliver a transparent understanding of how preference, technology, and market structure parameters influence the optimal degree of progressivity. A calibration for the U.S. economy indicates that endogenous skill investment, flexible labor supply, and the desire to finance government purchases play quantitatively similar roles in limiting optimal progressivity. In a version of the model where poverty constrains skill investment, optimal progressivity is close to the U.S. value. An empirical analysis on cross-country data offers support to the theory.

Keywords: Tax Progressivity, Income Distribution, Skill Investment, Labor Supply, Partial Insurance, Government Expenditures, Welfare, Cross-country Evidence.

JEL Codes: D30, E20, H20, H40, J22, J24

*Formerly titled “Redistributive Taxation in a Partial Insurance Economy.” The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System. Corresponding author: Jonathan Heathcote. Mailing address: Federal Reserve Bank of Minneapolis, Research Department, 90 Hennepin Ave., Minneapolis MN 55401, USA. Fax: 612 294 5515. Tel: 612 204 6385. Email: heathcote@minneapolisfed.org

I. INTRODUCTION

In determining how progressive to make the tax and transfer system, governments face a difficult trade-off. The classic argument in favor of progressivity is that private risk sharing is incomplete. Empirical estimates of the extent of pass-through from life cycle earnings shocks into consumption indicate limited private risk sharing (e.g., Cochrane 1991; Attanasio and Davis 1996). Perhaps more importantly, there are no markets to hedge against initial endowments that induce low future earnings. A progressive tax system offers both social insurance against labor market uncertainty (e.g., Eaton and Rosen 1980; Varian 1980) and redistribution with respect to initial conditions.

At the same time, governments are hesitant to push progressivity too far because of distortions to labor supply and skill investment. A tax schedule with increasing marginal rates reduces both the returns to working more hours and the returns to acquiring human capital (e.g., Heckman, Lochner, and Taber 1998; Krueger and Ludwig 2013; Guvenen, Kuruscu, and Ozkan 2014). Moreover, if the equilibrium skill premium responds to skill scarcity, a more progressive tax system, by depressing skill investment, may exacerbate inequality in pretax wages and undermine the original redistributive intent (e.g., Feldstein 1973; Stiglitz 1985).

An additional factor that has an impact on desired progressivity is the need to finance government purchases of goods and services. Individuals do not internalize that the additional output associated with working more hours or acquiring more skills allows the government to supply more public goods. This free-riding problem increases the social cost of a progressive tax system.

In this paper we develop an analytically tractable equilibrium model that features all of the forces shaping the optimal degree of progressivity described above. The environment is an extension of the partial insurance framework developed in Heathcote, Storesletten, and Violante (2014a). The economy is populated by households who choose how much to work and who face idiosyncratic labor market shocks. Some shocks are privately insurable and do not transmit to consumption, whereas others are uninsurable and induce consumption volatility. Individuals differ ex ante with respect to two characteristics: learning ability and the disutility of work effort. Those with higher learning ability invest more in skills prior to entering the labor market, whereas more diligent individuals work and earn more. An aggregate production technology with imperfect substitutability across skill types determines the marginal product and equilibrium price of each skill

type.

The government uses a nonlinear income tax and transfer system to provide social insurance and to finance publicly provided goods and services. Net taxes as a function of individual earnings y are given by the function $T(y) = y - \lambda y^{1-\tau}$, where the parameter τ indexes the progressivity of the system. The parameter λ determines net tax revenue and thus the share of output g devoted to public goods. One contribution of our paper is to show that this functional form offers a good approximation of the actual tax and transfer system in the United States.

We derive a closed-form expression for social welfare as a function of τ and g and the structural parameters of the model describing preferences, technology, and households' access to private consumption insurance. We then ask what degree of progressivity τ would be chosen by a benevolent planner. The planner's desire to provide social insurance against privately uninsurable wage shocks calls for $\tau > 0$ (i.e., marginal tax rates that rise with earnings). Similarly, heterogeneity in innate learning ability and in diligence translates into consumption dispersion that a utilitarian planner would like to counteract via a progressive tax and transfer system. However, the planner understands that more progressive taxes will lower labor supply and skill investment. The skill investment distortion depends on both the behavioral investment response to after-tax skill premia and the equilibrium response of pretax skill premia to changes in relative supplies of different skill types. The presence of government purchases constitutes a force toward regressive taxation ($\tau < 0$): the planner internalizes that a less progressive system encourages labor supply and skill investment, and makes it easier to finance expenditure.

After qualitatively inspecting these channels, we investigate their relative quantitative impacts on the optimal degree of progressivity. The model delivers closed-form solutions for the cross-sectional (co)variances of wages, hours, and consumption. We use empirical counterparts to these moments to estimate the structural model parameters. Under our baseline parameterization, a utilitarian government chooses less progressivity than is currently embedded in the U.S. tax/transfer system. The optimal value for τ is 0.084, whereas the current estimated value is 0.181. These values for progressivity translate into average (income-weighted) marginal tax rates of 26% and 34%, respectively. Switching to the optimal τ yields welfare gains of half a percent of lifetime consumption. Endogenous labor supply, endogenous skill investment, and the need to finance government purchases play quantitatively similar roles in limiting optimal progressivity. In the absence of any

one of these three channels, the optimal τ is substantially higher.

One way to rationalize the empirical degree of progressivity is to posit a planner who is averse to inequality in lifetime utility from consumption. Such a planner would choose the current U.S. τ given relative risk aversion over lifetime utility consumption equivalents slightly below 2 (this risk aversion coefficient is 1 under our baseline utilitarian objective). We also explore casting the choice for progressivity in a political-economic framework. Here we show that the model features a well-defined median voter and that this voter would choose notably higher progressivity than a utilitarian planner: $\tau = 0.144$ versus 0.084.

We consider two extensions of the baseline model in which we introduce frictions to skill investment. In the first, we prevent existing cohorts from adjusting their skill choice following a tax reform, which effectively allows the planner to expropriate past investment. In the second, we introduce a “poverty trap” constraint that prevents poor households from acquiring skills. Progressive taxation can now expand human capital investment on the extensive margin while still reducing investment for those unconstrained. In both of these versions of the model, optimal progressivity is similar to that observed in the United States.

Our paper contributes to the Ramsey-style literature that investigates the determinants of optimal progressivity in heterogeneous-agents incomplete-markets economies. A closely related study is Bénabou (2002), with whom we share the tax/transfer function. Also common to both models is the absence of trade in noncontingent bonds (a result that can be derived as an equilibrium outcome in our setting), which helps deliver analytical tractability. Key elements that differentiate our framework are our multiskill production technology, the partial insurance structure, heterogeneity in the taste for work, and the presence of public goods.

Other influential studies in the literature are Conesa and Krueger (2006) and Krueger and Ludwig (2013). Our environment is richer than those papers along some dimensions (preference heterogeneity, public goods, policy effects on skill prices) and more stylized in others (notably, the fact that wealth is in zero net supply). The advantage of our framework is that it is tractable, so the mechanics of how progressivity affects allocations and welfare are transparent.

Our normative analysis, in the spirit of Ramsey (1927), restricts the search for optimal progressivity within a given parametric class of tax/transfer schemes. In contrast, the Mirrlees (1971) approach to optimal taxation imposes no constraints on the form of the tax schedule and focuses

on the informational frictions that prevent the planner from implementing the first best allocation. Solving for constrained-efficient allocations is quite difficult outside simple static environments. Researchers have only recently incorporated persistent labor productivity shocks (Farhi and Werning 2013; Golosov, Troshkin, and Tsyvinski 2016; Gorry and Oberfield 2012), human capital accumulation (Stantcheva 2015), and imperfect substitutability across worker types (Rothschild and Scheuer 2013). Our model embeds all of these ingredients yet remains tractable, at the cost of exogenously restricting the class of tax schedules available to the planner. However, we will argue that our parametric specification is sufficiently flexible that the welfare gains from moving to a constrained-efficient Mirrleesian tax schedule are likely to be small.

Even though ours is primarily a normative exercise, in the last section of the paper we investigate the positive content of the theory by exploring whether it can help to account for observed cross-country variation in tax progressivity. Consistent with the theory's prescriptions for optimal policy, we find that tax progressivity falls with government purchases of goods and services, and rises with appropriate measures of income inequality.

II. TAX FUNCTION

Let $T(y)$ be net tax revenues at income level y . We study the optimal degree of progressivity within the class of tax and transfer policies defined by the function

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

This specification has a long tradition in public finance, starting from Feldstein (1969). More recently, Persson (1983) and Bénabou (2000, 2002) introduced this class of policies into dynamic macroeconomic models with heterogeneous agents.

The parameter τ determines the degree of progressivity of the tax system and is the key object of interest in our analysis. There are two ways to see why τ is a natural index of progressivity. First, equation (1) implies the following mapping between disposable (postgovernment) earnings \tilde{y}_i and pregovernment earnings y_i :

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

Thus, $(1 - \tau)$ measures the elasticity of posttax to pretax income.¹ Second, a tax scheme is commonly labeled progressive (regressive) if the ratio of marginal to average tax rates is larger (smaller) than 1 for every level of income y_i . Within our class, we have

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

When $\tau > 0$, marginal rates always exceed average rates, and the tax system is therefore progressive. Conversely, the tax system is regressive when $\tau < 0$. The case $\tau = 0$ implies that marginal and average tax rates are equal: the system is a flat tax with rate $1 - \lambda$.

Given τ , the second parameter, λ , shifts the tax function and determines the average level of taxation in the economy. At the break-even income level $y^0 = \lambda^{\frac{1}{\tau}} > 0$, the average tax rate is zero and the marginal tax rate is τ . If the system is progressive (regressive), then at every income level below (above) y^0 , the average tax rate is negative and households obtain a net transfer from the government. Thus, this function is best seen as a *tax and transfer* schedule, a property that has implications for the empirical measurement of τ .

Let g denote the fraction of output devoted to public consumption. Assuming a balanced budget, so that $gY = \int T(y_i) di$, the average income-weighted marginal tax rate is simply

$$(4) \quad \int T'(y_i) \left(\frac{y_i}{\bar{Y}} \right) di = 1 - (1 - \tau)(1 - g).$$

From equation (4) it is immediate that the average income-weighted marginal tax rate is increasing in both progressivity τ and government's share of output g . When $g = 0$, the average income-weighted marginal tax rate is exactly τ .

Empirical fit: We now demonstrate that this functional form offers a remarkably good representation of the actual tax and transfer system in the United States.

Because the U.S. tax system allows for numerous deductions and exemptions from gross income, it is important to distinguish between statutory and effective tax rates. The appropriate empirical counterpart to model progressivity τ is the degree of progressivity of *statutory* rates. In Online Appendix A, we show that (i) it is statutory rates that affect individual consumption/leisure

¹Musgrave (1959) refers to $1 - \tau$ as the coefficient of residual income progression. As discussed in Bénabou (2000), it has been proven that the posttax income distribution induced by one fiscal scheme Lorenz-dominates (i.e., displays less inequality than) the one induced by an alternative scheme if and only if the first scheme's progression coefficient ($1 - \tau$) is smaller everywhere. See, for example, Kakwani (1977).

and skill investment choices, and (ii) if we use taxable income (i.e., gross income net of deductions) as our empirical measure of income y , then equation (2) can be used to deliver an estimate for τ that captures precisely the progressivity of statutory tax rates.

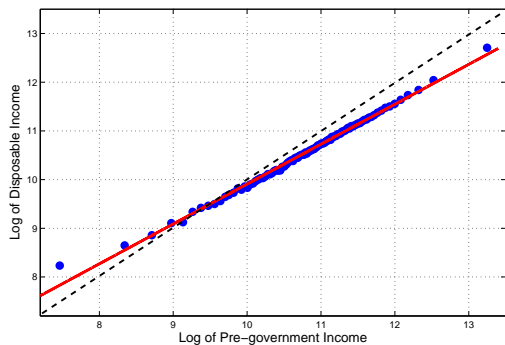
For our empirical exercise, we use data from the Panel Study of Income Dynamics (PSID) for survey years 2000, 2002, 2004, and 2006, in combination with the NBER's TAXSIM program (Feenberg and Coutts 1993).²

Pregovernment gross household income includes labor earnings, self-employment income, private transfers (alimony, child support, help from relatives, miscellaneous transfers, private retirement income, annuities, and other retirement income), plus income from interest, dividends, and rents. Taxable income is gross income minus deductions. For each household in the data, we compute the four main deductible expenses in the U.S. tax code: medical expenses, mortgage interest, state taxes paid, and charitable contributions. The first three categories are available from PSID data, while we use an imputation procedure to estimate charitable contributions at the household level (see Online Appendix A for details). Given tax-deductible expenses, TAXSIM calculates whether each household would be better off itemizing or taking the standard deduction. To obtain our final measure of pregovernment taxable income, we add the employer share (50%) of the Federal Insurance Contribution Act (FICA) tax—the sum of Social Security and Medicare taxes, computed directly by TAXSIM.³

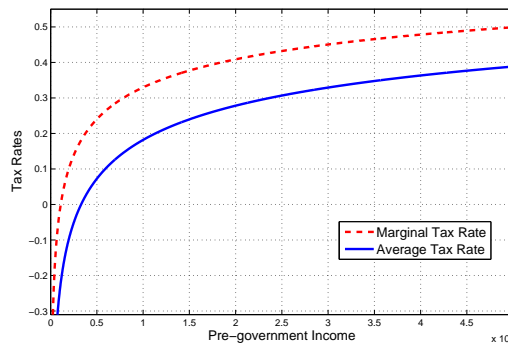
Postgovernment income \tilde{y} equals pregovernment income minus taxes plus transfers. Taxes include federal and state income taxes as well as the total FICA tax (both employer and employee shares), all of which we obtain from TAXSIM. Transfers include public cash transfers (AFDC/TANF, SSI and other welfare receipts, unemployment benefits, workers' compensation, and veterans' pensions). These transfers are measured directly from the PSID. Since we subtract Social Security taxes from household income in each year they are paid, we make an imputation for the corresponding marginal Social Security benefits gained by working that year and include those benefits in measured transfers for that year (see Online Appendix A for details). We do not

²The sample selection criteria follow closely those applied by Heathcote, Perri, and Violante (2010). In particular, we restrict attention to households aged 25-60 where at least one adult in the household earns more than the equivalent of part-time work at the minimum wage, in order to focus on the active population. The choice of the period 2000-2006 is motivated by the desire to use recent data while acknowledging that government transfers to U.S. households were abnormally large during the Great Recession.

³Whenever reported earnings in the PSID include some business income, we reduce this imputation by the portion of income coming from self-employment.



(a) Statistical fit on U.S. data



(b) Implied average and marginal tax rates

FIGURE I: Representation of the actual U.S. tax/transfer system through our tax/transfer function. The estimated value for progressivity is $\tau^{US} = 0.181$. Source: PSID 2000-06 and TAXSIM.

make any imputation for Medicare benefits because such entitlements are only conditional on age, not on years of work.

We estimate τ^{US} by ordinary least squares using equation (2) in log form. The point estimate is $\tau^{US} = 0.181$ ($S.E. = 0.002$). This simple model fits the empirical relationship between pre- and postgovernment income distributions remarkably well, yielding an R^2 of 0.91. In Figure I(a) we collapse our 12,875 observations into percentiles.⁴ Figure I(b) plots the average and marginal tax rates implied by our tax/transfer scheme evaluated at τ^{US} .⁵ The implied income-weighted average marginal tax rate is 0.34. For the same period, Barro and Redlick (2011) report a slightly higher income-weighted average marginal tax rate—around 37%.

In Figure I(a), the dots corresponding to the lowest percentiles of the income distribution (roughly below \$10,000) lie above the line predicted by the model, suggesting that our tax/transfer scheme tends to underestimate marginal tax rates at low income levels. Kosar and Moffitt (2016) estimate marginal tax rates faced by low-income families participating in various welfare programs and find that they generally face low or negative marginal rates. However, marginal rates vary substantially across households, and some households simultaneously enrolled in multiple welfare

⁴The coordinates of each circle in the figure are the mean of a particular percentile of the pregovernment income distribution (x axis) and the mean of postgovernment income within the households in that same percentile (y axis).

⁵Guner, Kaygusuz, and Ventura (2014) estimate this same function on a large cross-sectional dataset from the U.S. Internal Revenue Service (the “IRS Public Use Tax File”). They estimate smaller values for progressivity, reflecting the fact that IRS data do not include government transfers. The same caveat applies to the estimate in Chen and Guo (2013).

programs face high marginal tax rates where benefits are phased out. While our parametric functional form cannot capture this variation in tax rates at low income levels, we note that families with less than \$10,000 of taxable income represent less than 10% of all taxpayers aged 25-60 in the United States, and less than 2% of our sample of households active in the labor market.

Robustness: The PSID data have three potential limitations for the purposes of estimating progressivity: (i) the PSID undersamples the very rich, (ii) PSID taxes are imputed through TAXSIM, and (iii) the PSID covers only a subset of in-kind benefits. The Congressional Budget Office (CBO) publishes tables reporting household income, federal taxes paid, and federal transfers received for various quantiles of the entire distribution of before-tax income, including all the top earners.⁶ The CBO measure of transfers is more comprehensive than the one we construct from the PSID. It includes the value of Food Stamps, school lunches, housing and energy assistance, and benefits provided by Medicare and Medicaid. Moreover, the CBO also includes employer-paid health insurance premiums in its measure of pregovernment income.⁷

From the CBO tables we construct pre- and postgovernment income for the first, second, third, and fourth quintiles of the pregovernment income distribution, and for the 81st-90th, the 91st-95th, and the 96th-99th percentiles, and the top 1%. We use these moments to estimate the progressivity parameter τ^{US} for the 2000-2006 period and obtain $\tau^{US} = 0.200$, which is slightly higher than our PSID estimate for the same years.⁸ Since we use the PSID to estimate other model parameters in Section VI., we will use the PSID-based estimate ($\tau^{US} = 0.181$) in our baseline analysis.

Discussion: One way to think about our exercise is as follows. Given the form of the tax system that is currently in place, we ask how much more or less progressive taxes should be, and what would be the associated welfare gains? Although the functional form in equation (1) offers a good positive account of the U.S. tax system, it is potentially restrictive from a purely normative perspective. Two key restrictions are implicit in $T(y_i)$. First, it is either globally convex in income,

⁶The CBO analysis draws its information on income from two primary sources. The core data come from the Statistics of Income (SOI), a nationally representative sample of individual income tax returns collected by the IRS. The CBO supplements that information with data on transfers from the Annual Social and Economic Supplement to the Census Bureau's Current Population Survey (CPS).

⁷There are two limitations of the CBO data from the perspective of measuring the overall progressivity of the U.S. tax and transfer system. First, the CBO data exclude state-level taxes and transfers. Second, the reported statistics are aggregates across a cross section of households of all ages, and the CBO does not attempt to measure the extent to which Social Security transfers received when old are tied to Social Security taxes paid when young, as we do.

⁸We make one adjustment to the CBO measure of postgovernment income to better align it with our own measure, which is that we exclude the value of Medicare transfers.

if $\tau > 0$, or globally concave, if $\tau < 0$. As a result, marginal tax rates are monotonic in income. Second, it does not allow for lump-sum cash transfers, since $T(0) = 0$.⁹

Heathcote and Tsujiyama (2016) consider the welfare gains of moving from tax systems of the type described by equation (1) to the fully optimal nonparametric Mirrlees tax schedule. Their environment is a stripped-down version of the model developed here. They find that the size of the welfare gains of moving from the tax system described above with $\tau = \tau^{US}$ to the constrained-efficient Mirrlees system is sensitive to the taste for redistribution embedded in the planner’s social welfare function. However, for a wide range of alternative welfare functions, the best policy in the class described by equation (1) delivers the vast majority of the maximum potential welfare gains from tax reform.¹⁰ Thus, the restrictions implicit in the system described by (1) are not particularly important from a normative standpoint.

III. ECONOMIC ENVIRONMENT

We describe the economy in steady state and omit time subscripts.

Demographics: We adopt the Yaari “perpetual youth” structure. At every age a , an agent survives to the next period with constant probability $\delta < 1$. Each period a cohort of newborn agents of size $1 - \delta$ enters the economy. There are no intergenerational links.¹¹ We index agents by $i \in [0, 1]$.

Life cycle: The life of every individual i starts with an initial investment in skills. After choosing skill level s_i at age $a = 0$, the individual enters the labor market and starts facing random fluctuations in labor productivity z_i . Every period she chooses market hours of work $h_i \geq 0$ and consumption of a private good c_i .

Technology: Output Y is a constant elasticity of substitution aggregate of effective hours supplied by the continuum of skill types $s \in [0, \infty)$,

$$(5) \quad Y = \left(\int_0^\infty [N(s) \cdot m(s)]^{\frac{\theta-1}{\theta}} ds \right)^{\frac{\theta}{\theta-1}},$$

⁹Our model can in principle capture (as part of the public good G) lump-sum transfers in the form of in-kind goods or services, as long as these are imperfectly substitutable with private consumption (e.g., public education and health care).

¹⁰Assuming a utilitarian objective, for example, the best policy in this class delivers 84% of the maximum potential welfare gains from tax reform. See Heathcote and Tsujiyama (2016) for details.

¹¹Private bequests could provide a form of insurance against a bad draw of initial conditions, which might reduce optimal progressivity. For a model along these lines, see Bakış, Kaymak, and Poschke (2015).

where $\theta > 1$ is the elasticity of substitution across skill types, $N(s)$ denotes average effective hours worked by individuals of skill type s , and $m(s)$ is the density of individuals with skill type s . Note that all skill levels enter symmetrically in the production technology, and thus any equilibrium differences in skill prices will reflect relative scarcity of the corresponding skill types.¹²

Output is used for private consumption and public consumption G . The rate of transformation between the two forms of consumption is 1, and thus the aggregate resource constraint is

$$(6) \quad Y = \int_0^1 c_i di + G.$$

Preferences: Preferences over private consumption, hours worked, publicly provided goods, and skill investment effort for individual i are given by

$$(7) \quad U_i = -v_i(s_i) + (1 - \beta\delta) \mathbb{E}_0 \sum_{a=0}^{\infty} (\beta\delta)^a u_i(c_{ia}, h_{ia}, G),$$

where $\beta < 1$ is the pure 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 form

$$(8) \quad v_i(s_i) = \frac{(\kappa_i)^{-1/\psi}}{1 + 1/\psi} (s_i)^{1+1/\psi},$$

where the parameter $\psi \geq 0$ determines the elasticity of skill investment with respect to the return to skill, and $\kappa_i \geq 0$ is an individual-specific parameter that determines the utility cost of acquiring skills. The larger is κ_i , the smaller is the cost, so one can think of κ_i as indexing innate learning ability. We assume that $\kappa_i \sim \text{Exp}(\eta)$, an exponential distribution with parameter η . As we demonstrate below, exponentially distributed ability yields Pareto right tails in the equilibrium wage and earnings distributions.

The period utility function u_i is specified as

$$(9) \quad u_i(c_{ia}, h_{ia}, G) = \log c_{ia} - \frac{\exp[(1 + \sigma)\varphi_i]}{1 + \sigma} (h_{ia})^{1+\sigma} + \chi \log G,$$

where $\exp[(1 + \sigma)\varphi_i]$ measures the disutility of work effort. The individual-specific parameter φ_i is normally distributed: $\varphi_i \sim N\left(\frac{v_\varphi}{2}, v_\varphi\right)$, where v_φ denotes the cross-sectional variance.¹³

¹²In Online Appendix D.1, we compare our constant elasticity of substitution skill aggregator with the one assumed by Bénabou (2005) and discuss the implications of these alternative specifications for the relation between optimal progressivity and the elasticity of substitution parameter θ .

¹³Introducing additional weighting parameters (common across all households) on the utility terms defining the costs of skill investment and labor supply would have no impact on the shape of the welfare-maximizing policy.

We assume that κ_i and φ_i are uncorrelated. The parameter $\sigma > 0$ determines aversion to hours fluctuations. It is useful to define the *tax-modified* Frisch elasticity:

$$(10) \quad \frac{1}{\hat{\sigma}} = \frac{1 - \tau}{\sigma + \tau}.$$

Below we show that $\hat{\sigma}^{-1}$ measures the elasticity of hours worked to a transitory shock to the pretax wage. Finally, $\chi \geq 0$ measures the taste for the public good G relative to private consumption.

Labor productivity and earnings: Log individual labor efficiency z_{ia} is the sum of two orthogonal components, α_{ia} and ε_{ia} :

$$(11) \quad \log z_{ia} = \alpha_{ia} + \varepsilon_{ia}.$$

The first component α_{ia} follows the unit root process $\alpha_{ia} = \alpha_{i,a-1} + \omega_{ia}$ with i.i.d. innovation $\omega_{ia} \sim N\left(-\frac{v_\omega}{2}, v_\omega\right)$, and with initial condition $\alpha_{i0} = 0$.¹⁴ The second component is an i.i.d. shock, $\varepsilon_{ia} \sim N\left(-\frac{v_\varepsilon}{2}, v_\varepsilon\right)$. 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 standard law of large numbers ensures that idiosyncratic shocks induce no aggregate uncertainty.

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

$$(12) \quad y_{ia} = \underbrace{p(s_i)}_{\text{skill price}} \times \underbrace{\exp(\alpha_{ia} + \varepsilon_{ia})}_{\text{labor market shocks}} \times \underbrace{h_{ia}}_{\text{hours}}.$$

The first component $p(s_i)$ is the equilibrium price for the type of labor supplied by an individual with skills s_i , the second component is individual stochastic labor efficiency, and the third component is the number of hours worked by the individual. Thus, individual earnings are determined by (i) skills accumulated before labor market entry, in turn reflecting innate learning ability κ_i ; (ii) fortune in labor market outcomes determined by the realization of idiosyncratic efficiency shocks; and (iii) work effort, reflecting, in part, innate taste for leisure, measured by φ_i .

Because idiosyncratic productivity shocks are exogenous, the two channels via which taxation will have an impact on the equilibrium pretax earnings distribution are by changing skill investment choices, and thus skill prices, and by changing labor supply decisions.

Financial assets: We adopt a simplified version of the partial insurance structure developed in Heathcote, Storesletten, and Violante (2014a). There is a full set of state-contingent claims

¹⁴Thus, all earnings inequality among newborn agents reflects heterogeneous skill levels.

indexed by the ε shock —and thus the ε shocks are fully insurable—whereas the α shocks, by assumption, cannot be insured through markets or smoothed via storage.¹⁵ Let $B(\mathbf{E})$ and $Q(\mathbf{E})$ denote the quantity and the price, respectively, of insurance claims purchased that pay one unit of consumption if and only if $\varepsilon \in \mathbf{E} \subseteq \mathbb{R}$. Insurance claims are in zero net supply, and newborn agents start with zero initial holdings of such claims.¹⁶ Our model spans the entire range between autarky ($v_\varepsilon = 0$) and full insurance ($v_\omega = 0$). In general, when both $v_\omega > 0$ and $v_\varepsilon > 0$, ours is a *partial insurance* economy, in the language of Blundell, Pistaferri, and Preston (2008).¹⁷

Markets: The final consumption good, all types of labor services, and financial claims are traded in competitive markets. The publicly provided good G cannot be purchased privately. The final good is the numéraire of the economy.

Government: The government runs the tax/transfer scheme described in Section II. and funds expenditure G on public consumption. Let g denote government consumption as a fraction of aggregate output (i.e., $G = gY$). Since we abstract from public debt, the government budget constraint holds period by period and reads as

$$(13) \quad g \int_0^1 y_i di = \int_0^1 (y_i - \lambda y_i^{1-\tau}) di.$$

The government chooses the pair (g, τ) , with λ being determined residually by equation (13).

III.A. Agent's Problem

At age $a = 0$, the agent chooses a skill level, given her idiosyncratic draw (κ_i, φ_i) . Combining equations (7) and (8), the first-order necessary and sufficient condition for the skill choice is

$$(14) \quad \frac{\partial v_i(s)}{\partial s} = \left(\frac{s}{\kappa_i} \right)^{\frac{1}{\psi}} = (1 - \beta\delta) \mathbb{E}_0 \sum_{a=0}^{\infty} (\beta\delta)^a \frac{\partial u_i(c_{ia}, h_{ia}, G)}{\partial s}.$$

¹⁵The environment in Heathcote, Storesletten, and Violante (2014a) is more general because it also includes a noncontingent bond market. In a previous version of this paper (Heathcote, Storesletten, and Violante 2014b), we prove that, even if households were allowed to trade this additional asset, they would choose not to do so in our model. Thus, α shocks would remain uninsured in equilibrium, and equilibrium allocations for consumption, hours worked, and skill choices would be identical to the allocation we study here.

¹⁶The complete markets assumption with respect to ε implies that it is straightforward to introduce a richer statistical process for the ε shocks. For example, in Heathcote, Storesletten, and Violante (2014a), we add a unit root component to the insurable component of wages. As we show below, all that matters for the analysis of optimal taxation is the *cross-sectional* variance of insurable wage risk, which can be estimated independently of the *time-series* process for ε . Therefore, to simplify the exposition, in this paper we maintain the assumption that ε is an i.i.d. shock.

¹⁷Tax progressivity provides public risk sharing and reduces the equilibrium demand for private insurance. However, since the market structure is exogenous, tax progressivity does not affect the *supply* of private insurance. In contrast, public insurance can crowd out private risk sharing in models with moral hazard or limited enforcement.

Thus, the marginal disutility of skill investment for an individual with learning ability κ_i must equal the discounted present value of the corresponding higher expected lifetime wages.

The timing of the agent's problem during her subsequent working life is as follows. At the beginning of every period a , the innovation ω_{ia} to the random walk shock α_{ia} is realized. Then, the insurance markets against the ε shocks open and the individual buys insurance claims $B(\cdot)$. Finally, ε_{ia} is realized and the individual chooses hours h_{ia} , receives wage payments, and chooses consumption expenditures c_{ia} . Thus, the individual budget constraint in the middle of the period, when the insurance purchases are made, is

$$(15) \quad \int_E Q(\varepsilon) B(\varepsilon) d\varepsilon = 0,$$

and the budget constraint at the end of the period, after the realization of ε_{ia} , is

$$(16) \quad c_{ia} = \lambda [p(s_i) \exp(\alpha_{ia} + \varepsilon_{ia}) h_{ia}]^{1-\tau} + B(\varepsilon_{ia}).$$

Given an initial skill choice, the problem for an agent is to choose sequences of consumption and hours worked in order to maximize (7) subject to sequences of budget constraints of the form (15)-(16), taking as given the wage process described in equation (11). In addition, agents face non-negativity constraints on consumption and hours worked.

III.A.i. 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 dispersion in the taste for leisure or in labor productivity. Since skill levels are perfect substitutes in production, there is no skill investment either, so the economy collapses to a representative agent model. The representative agent's problem is static:

$$(17) \quad \max_{C,H} \left\{ \log C - \frac{H^{1+\sigma}}{1+\sigma} + \chi \log G \right\}$$

s.t.

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

Taking the fiscal variables (λ, g, τ) as given, the optimal choices for the representative agent are

$$(18) \quad \log H^{RA}(\tau) = \frac{1}{1+\sigma} \log(1-\tau),$$

$$(19) \quad \log C^{RA}(g, \tau) = \log \lambda(g, \tau) + \frac{1-\tau}{1+\sigma} \log(1-\tau).$$

The production technology simplifies to $Y = H$, implying $G = gH$. Solving for $\lambda(g, \tau)$ from the government budget constraint, $G = H - \lambda H^{1-\tau}$, and substituting into (19) gives

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

These expressions show that a more progressive tax system (a higher value for τ) reduces labor supply and, therefore, equilibrium consumption. The reason is that higher progressivity raises the marginal tax rate faced by the representative agent. As $\tau \rightarrow 1$, $H^{RA}(\tau) \rightarrow 0$. Note that, with logarithmic utility, the tax level parameter λ has no impact on labor supply, which explains why hours worked (and output) are independent of the government consumption to output ratio g .

IV. EQUILIBRIUM

We now adopt a recursive formulation to define a stationary competitive equilibrium for our economy. The individual state vector for the skill accumulation decision at age $a = 0$ is just the fixed individual effects (κ, φ) . At subsequent ages, the state vector for the beginning-of-the-period decision when insurance claims are purchased is (φ, α, s) . The state vector for the end-of-period consumption and labor supply decisions is $(\varphi, \alpha, \varepsilon, s, \bar{B})$, where $\bar{B} = B(\varepsilon; \varphi, \alpha, s)$ are state-contingent insurance payouts.¹⁸ Because of the perpetual youth structure, age is not a state variable.

Given (g, τ) , a *stationary recursive competitive equilibrium* for our economy is a tax level λ , asset prices $Q(\cdot)$, skill prices $p(s)$, decision rules $s(\kappa, \varphi)$, $c(\varphi, \alpha, \varepsilon, s)$, $h(\varphi, \alpha, \varepsilon, s)$, and $B(\cdot; \varphi, \alpha, s)$, and aggregate quantities $N(s)$ such that:

1. Households solve the problem described in Section III.A., and $s(\kappa, \varphi)$, $c(\varphi, \alpha, \varepsilon, s)$, $h(\varphi, \alpha, \varepsilon, s)$, and $B(\cdot; \varphi, \alpha, s)$ 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 : $p(s) = (Y / [N(s) \cdot m(s)])^{1/\theta}$.
3. Asset markets clear, and the prices $Q(\cdot)$ of insurance claims are actuarially fair.
4. The government budget is balanced: λ satisfies equation (13).

¹⁸Since in equilibrium \bar{B} is a function of $(\varphi, \alpha, \varepsilon, s)$, in what follows we omit \bar{B} from the state vector.

Propositions 1 and 2 describe the equilibrium allocations and skill prices in closed form. The payoff from tractability will be especially evident in Proposition 4, where we derive an analytical solution for social welfare. In what follows, we make explicit the dependence of equilibrium allocations and prices on (g, τ) in preparation for our analysis of the optimal taxation problem.

Proposition 1 [hours and consumption]. *The equilibrium hours-worked allocation is given by*

$$(20) \quad \log h(\varphi, \varepsilon; \tau) = \log H^{RA}(\tau) - \varphi + \frac{1}{\hat{\sigma}}\varepsilon - \frac{1}{\hat{\sigma}(1-\tau)}\mathcal{M}(v_\varepsilon; \tau),$$

where H^{RA} are hours for the “representative agent” in equation (18) and

$\mathcal{M}(v_\varepsilon; \tau) = (1-\tau)(1-\tau(1+\hat{\sigma}))/\hat{\sigma} \cdot v_\varepsilon/2$. The consumption allocation is given by

$$(21) \quad \log c(\varphi, \alpha, s; g, \tau) = \log [C^{RA}(g, \tau) \vartheta(\tau)] + (1-\tau) [\log p(s; \tau) + \alpha - \varphi] + \mathcal{M}(v_\varepsilon; \tau),$$

where C^{RA} is consumption of the “representative agent” in equation (19) and $\vartheta(\tau)$ is common across agents.

With logarithmic utility and zero wealth, the income and substitution effects on labor supply from differences in uninsurable shocks α and skill levels s exactly offset, and hours worked are independent of (s, α) . The hours allocation is composed of four terms. The first is hours of the representative agent, which, as explained above, fall with progressivity. The second term captures the fact that a higher idiosyncratic disutility of work leads an agent to choose lower hours. The third term shows that the response of hours worked to an insurable shock ε (which has no income effect precisely because it is insurable) is mediated by the tax-modified Frisch elasticity $1/\hat{\sigma}$. Progressivity lowers this elasticity. The fourth term captures the welfare-improving effect of insurable wage variation. As shown in Heathcote, Storesletten, and Violante (2008), larger dispersion of insurable shocks allows agents to work more when they are more productive and take more leisure when they are less productive, thereby raising average productivity, average leisure, and welfare. Progressivity weakens this channel because it dampens the efficient response of hours to insurable wage shocks.

Consumption is additive in five separate components. The first component is (rescaled) consumption of the representative agent, described in Section III.A.i.¹⁹ Consumption is increasing in

¹⁹The rescaling constant $\vartheta(\tau)$ reflects the fact that the equilibrium balanced-budget function $\lambda(g, \tau)$ is different in the heterogeneous agent and representative agent versions of the model.

the skill level s (because skill prices are increasing in skills) and in the uninsurable component of wages α . Since hours worked are decreasing in the disutility of work φ , so are earnings and consumption. The redistributive role of progressive taxation is evident from the fact that a larger τ shrinks the pass-through to consumption from heterogeneity in fixed effects s and φ and from realizations of uninsurable wage shocks α . The final component captures the fact that insurable variation in productivity has a positive level effect on average consumption in addition to average leisure. Again, higher progressivity weakens this effect. Because of the assumed separability between consumption and leisure in preferences, consumption is independent of the insurable shock ε .

Proposition 2 [skill price and skill choice]. *In equilibrium, skill prices are given by*

$$(22) \quad \log p(s; \tau) = \pi_0(\tau) + \pi_1(\tau) \cdot s(\kappa; \tau), \quad \text{where}$$

$$(23) \quad \pi_1(\tau) = \left(\frac{\eta}{\theta}\right)^{\frac{1}{1+\psi}} (1-\tau)^{-\frac{\psi}{1+\psi}}$$

$$(24) \quad \pi_0(\tau) = \frac{1}{\theta-1} \left\{ \frac{1}{1+\psi} \left[\psi \log\left(\frac{1-\tau}{\theta}\right) - \log(\eta) \right] + \log\left(\frac{\theta}{\theta-1}\right) \right\}.$$

The skill investment allocation is given by

$$(25) \quad s(\kappa; \tau) = [(1-\tau)\pi_1(\tau)]^\psi \cdot \kappa = \left[\frac{\eta}{\theta}(1-\tau)\right]^{\frac{\psi}{1+\psi}} \cdot \kappa$$

and the equilibrium skill density $m(s)$ is exponential with parameter $(\eta)^{\frac{1}{1+\psi}} [\theta/(1-\tau)]^{\frac{\psi}{1+\psi}}$.

Note, first, that the log of the equilibrium skill price takes a ‘‘Mincerian’’ form (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 pretax marginal return to skill.

From the skill investment rule (25), it is clear that the parameter ψ defines the elasticity of skill investment to the after-tax return to skills, $(1-\tau)\pi_1(\tau)$. Equation (25) also clarifies that higher progressivity τ reduces the after-tax return to investing in skills and depresses skill investment. In the limit as $\tau \rightarrow 1$, $s \rightarrow 0$ at every κ : there is no incentive to boost wages by investing in skills if all the excess returns will be taxed away.

Equation (23) indicates that higher τ increases the equilibrium *pretax* marginal return $\pi_1(\tau)$. The logic is that increasing τ compresses the skill distribution toward zero and, as high skill types

become scarce, imperfect substitutability in production drives up the pretax return to skill. Thus, our model features a *Stiglitz effect* (Stiglitz 1985). The larger is ψ , the more sensitive is skill investment to a given increase in τ , and the larger is the increase in the pretax skill premium.

Note that the skill investment decision is independent of φ (and it would also be independent of α_0 if there was heterogeneity in initial labor productivity within skill types). The logic is that, with log utility, the welfare gain from additional skill investment is proportional to the log change in wages the investment would induce, which is unaffected by the *level* of wages or hours.

Corollary 2.1 [distribution of skill prices]. *The distribution of log skill premia $\pi_1(\tau) \cdot s(\kappa; \tau)$ is exponential with parameter θ . Thus, the variance of log skill prices is*

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

The distribution of skill prices $p(s; \tau)$ in levels is Pareto with scale (lower bound) parameter $\exp(\pi_0(\tau))$ and Pareto parameter θ .

Log skill premia are exponentially distributed because the log skill price is affine in skill s (equation 22) and skills retain the exponential shape of the distribution of learning ability κ (equation 25). It is interesting that inequality in skill prices is independent of τ . 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 Stiglitz effect on prices). On the other hand, higher progressivity compresses the distribution of skills (the quantity effect). These two forces exactly cancel out under our baseline utility specification.

Since the exponential of an exponentially distributed random variable is Pareto, the distribution of skill prices in levels is Pareto with parameter θ . The other stochastic components of wages (and hours worked) are lognormal. Because the Pareto component dominates at the top, the equilibrium distributions of wages and earnings have Pareto right tails, a robust feature of their empirical counterparts (see, e.g., Atkinson, Piketty, and Saez 2011). It also follows that the distribution of consumption has a Pareto tail, consistent with the empirical evidence in Toda (2016). Finally, the distributions of log wages and log earnings 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 VI.F.

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

Corollary 2.2 [aggregate quantities]. *Average hours worked $H(\tau)$ and average effective hours $N(\tau)$ are independent of skill type s . $H(\tau)$, $N(\tau)$, and output $Y(\tau)$ are given by*

$$(26) \quad H(\tau) = \mathbb{E}[h(\varphi, \varepsilon; \tau)] = (1 - \tau)^{\frac{1}{1+\sigma}} \cdot \exp \left[\left(\frac{\tau(1 + \hat{\sigma})}{\hat{\sigma}^2} - \frac{1}{\hat{\sigma}} \right) \frac{v_\varepsilon}{2} \right],$$

$$(27) \quad N(\tau) = \mathbb{E}[\exp(\alpha + \varepsilon)h(\varphi, \varepsilon; \tau)] = H(\tau) \cdot \exp \left(\frac{1}{\hat{\sigma}} v_\varepsilon \right).$$

$$(28) \quad Y(\tau) = \mathbb{E}[p(s; \tau) \exp(\alpha + \varepsilon)h(\varphi, \varepsilon; \tau)] = N(\tau) \cdot \mathbb{E}[p(s; \tau)],$$

where $\mathbb{E}[p(s; \tau)] = \exp(\pi_0(\tau)) \cdot \theta / (\theta - 1)$.

Aggregate labor productivity is

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

Progressivity affects aggregate output through two channels: labor supply and skill investment choices. From equation (26), the elasticity of aggregate effective hours $N(\tau)$ with respect to τ at $\tau = 0$ is $-1/(1 + \sigma)$. The elasticity of output per effective hour ($Y(\tau)/N(\tau) = \mathbb{E}[p(s; \tau)]$) with respect to τ , which reflects skill investment, is $-\psi / [(1 + \psi)(\theta - 1)]$ (see equation (34) below). We will return to these two elasticities in Section V.C. when characterizing the conditions under which the optimal tax system is progressive.

IV.A. Efficiency

Before turning to the characterization of the optimal degree of progressivity, we briefly discuss the efficiency properties of the competitive equilibrium. The equilibrium with $\tau = 0$ is generally not efficient in our environment for two reasons. The first is that there are no private markets for insuring the ω shock. The second is that for a given g (and corresponding λ), there is a free-riding problem. If all agents worked more, the quantity of the valued public good provided would increase, but from the perspective of an atomistic single agent, the supply of the public good is exogenous. Because the marginal social gain from work exceeds the marginal private gain, labor supply is inefficiently low. The following proposition states that when the economy features complete markets with respect to wage shocks ($v_\omega = 0$) and does not feature this free-riding problem ($\chi = 0$), the equilibrium with $\tau = 0$ is efficient.

Proposition 3 [efficiency with $\chi = v_\omega = \tau = 0$]. *If $\chi = v_\omega = 0$, then the competitive equilibrium allocation with $\tau = 0$ is efficient. This allocation is the solution to a planner's problem with Pareto weights proportional to $\exp(-\varphi + \eta/\theta \cdot \kappa)$.*

Because individuals with high learning ability κ or low disutility of work effort φ enjoy relatively high consumption in the competitive equilibrium with $\tau = 0$, these agents must receive relatively large Pareto weights in the planner's problem that delivers the same allocation. Note that the competitive equilibrium with $\tau = 0$ cannot deliver the allocation that would be chosen by a utilitarian social planner who weights all agents equally. Such a planner would want to redistribute against income differentials because of heterogeneity in κ and φ , and would therefore choose $\tau > 0$.

V. WELFARE EFFECTS OF TAX REFORM

We imagine the economy starting out in a steady state corresponding to a policy pair (g_{-1}, τ_{-1}) and consider permanent unanticipated policy changes at date 0 to a new policy (g, τ) . The presence of skill investment in the model raises two related issues when contemplating tax reform. First, if past investment decisions are irreversible, then the government is tempted to tax returns to skill because such taxation is not distortionary ex post. This result is analogous to the temptation to tax initial physical capital in the growth model. Second, if the distribution of skills adjusts slowly following a change in the tax system, then even permanent policy changes will induce transitional dynamics. In our benchmark analysis, we sidestep both of these issues by making the assumption that the choice of skills is fully reversible at any point. This assumption implies that transition following a tax reform is instantaneous: given a choice for the new pair (g, τ) , the economy immediately converges to the steady-state distribution of skills associated with this policy.

In Section VI.C. we generalize our characterization of optimal progressivity by making the polar opposite assumption that skills are fully irreversible. In this alternative version of the model, there are transitional dynamics between the initial and final steady states, and the motive to tax the existing sunk stock of skills affects the optimal choice of progressivity.

V.A. Social Welfare Function

The baseline utilitarian social welfare function we use to evaluate alternative policies puts equal weight on all agents within a cohort. In our context, where agents have different relative weights on consumption versus work effort, we define equal weights to mean that the planner cares equally about the utility from consumption of all agents. Thus, the contribution to social welfare from any given cohort is the within-cohort average value for remaining expected lifetime utility, where equation (7) defines expected lifetime utility at age zero. In Section VI.E., we generalize the objective function to consider cases in which the planner is more or less averse to cross-sectional inequality.

The overlapping generations structure requires us to take a stand on how the government weighs cohorts that enter the economy at different dates. We assume that the planner discounts the lifetime utility of future generations at rate γ . Social welfare evaluated as of date 0 is then given by

$$(29) \quad \mathcal{W}(g, \tau; \tau_{-1}) \equiv (1 - \gamma)\Gamma \sum_{j=-\infty}^{\infty} \gamma^j U_{j,0}(g, \tau; \tau_{-1}),$$

where $U_{j,0}(g, \tau; \tau_{-1})$ is remaining expected lifetime utility (discounted back to date of birth) as of date 0 for the cohort that entered the economy at date j .²⁰ The constant $\Gamma = (\gamma - \beta\delta)/[\gamma(1 - \beta\delta)]$ premultiplying the summation is a convenient normalization.²¹

The next proposition expresses social welfare as a function of the two policy instruments (g, τ) .

Proposition 4 [closed-form social welfare]. *In the model with fully reversible investment, when*

²⁰Remaining lifetime utility depends on the lagged value for progressivity τ_{-1} because the difference between τ_{-1} and τ will determine (the cost of) net new investment in skills for cohorts who entered the economy prior to date 0.

²¹Following Calvo and Obstfeld (1988), we assume that the planner discounts each individual's welfare back to their birth dates using the agent's discount factor β . This ensures that the planner's objective function is time consistent. Thus, the relative weight the planner places at a particular date on the felicity of one agent who is s years older than another is $(\beta/\gamma)^s$.

the social welfare function is given by equation (29), welfare from implementing policy (g, τ) is

$$\begin{aligned}
(30) \quad \mathcal{W}(g, \tau; \tau_{-1}) = & \log(1-g) + \chi \log g + (1+\chi) \frac{\log(1-\tau)}{(1+\hat{\sigma})(1-\tau)} - \frac{1}{1+\hat{\sigma}} \\
& + (1+\chi) \frac{1}{(1+\psi)(\theta-1)} \left[\psi \log(1-\tau) + \log \left(\frac{1}{\eta \theta^\psi} \left(\frac{\theta}{\theta-1} \right)^{\theta(1+\psi)} \right) \right] \\
& - \frac{\psi}{(1+\psi)\theta} \left[(1-\tau) - \frac{\beta\delta}{\gamma} \frac{(1-\gamma)}{(1-\beta\delta)} (1-\tau_{-1}) \right] \\
& - \left[-\log \left(1 - \left(\frac{1-\tau}{\theta} \right) \right) - \left(\frac{1-\tau}{\theta} \right) \right] \\
& - (1-\tau)^2 \frac{v_\varphi}{2} \\
& - \left[(1-\tau) \left(\frac{\beta\delta}{\gamma - \beta\delta} \right) \frac{v_\omega}{2} - \log \left(\frac{1 - \delta \exp \left(\frac{-\tau(1-\tau)}{2} v_\omega \right)}{1-\delta} \right) \right] \\
& + (1+\chi) \left[\frac{1}{\hat{\sigma}} v_\varepsilon - \sigma \frac{1}{\hat{\sigma}^2} \frac{v_\varepsilon}{2} \right].
\end{aligned}$$

To obtain the expression in equation (30), we first solve for the value $\lambda(g, \tau)$ that balances the government budget. Next, plugging the consumption, hours, and skill allocations into equation (29), we can express social welfare as a function of primitive preference, technology, and policy parameters.

Corollary 4.1 [independence from past choices]. *The optimal pair (g, τ) is independent of τ_{-1} .*

In equation (30), τ_{-1} appears in an additively separable term that does not involve any other policy parameters. This result depends on the reversible skill investment assumption. Because the welfare impact of alternative choices for (g, τ) is independent of τ_{-1} , we will henceforth denote social welfare $\mathcal{W}(g, \tau)$ and omit the inconsequential terms involving τ_{-1} .

Corollary 4.2 [concavity of social welfare]. *Social welfare $\mathcal{W}(g, \tau)$ is globally concave in g and, if $\sigma \geq 2$, is also globally concave in τ .*

As we show in Online Appendix B.7, aside from the term multiplying v_ε in the last row of equation (30), social welfare is globally concave in τ for any $\sigma \geq 0$. The term involving v_ε is also globally concave in τ if $\sigma \geq 2$, a condition that is satisfied in the calibration. Establishing concavity is useful since it means that a first-order approach is sufficient to compute the optimal τ and g .

Corollary 4.3 [independence of policy instruments]. *The welfare-maximizing value for τ is independent of g .*

The two policy parameters τ and g do not appear jointly in any one of the additively separable terms in (30). Thus, the welfare-maximizing choice for τ must be independent of the value for g , irrespective of whether the choice for g is welfare maximizing. However, the welfare-maximizing choice for τ will depend on the parameter χ that defines the taste for publicly provided goods.

Corollary 4.4 [Samuelson condition]. *The welfare-maximizing value for g is given by*

$$(31) \quad g^* = \frac{\chi}{1 + \chi}.$$

The optimal choice for public goods in the economy is obtained by taking the first-order condition of equation (30) with respect to g . The optimal fraction of output to devote to public goods is independent of how much inequality there is in the economy and independent of the progressivity of the tax system. It depends only on households' relative taste for the public good χ .

To understand this result, note that the choice for g does not appear in the equilibrium allocations for hours worked or skill investment (eqs. 20 and 25). Thus, changing g will not change aggregate income or the distribution of income. It follows that the government's only concern in setting g is to optimally divide output between private and public consumption. The optimal split turns out to be exactly the one that equates the marginal rate of substitution between private and public consumption to the marginal rate of transformation between the two goods in a representative agent version of the model (see Section V.B.i.). We call this the "Samuelson condition" (Samuelson 1954).

Corollary 4.5 [irrelevance of η]. *The welfare-maximizing value for progressivity τ is independent of the learning ability distribution parameter η .*

Even though the exponential parameter η has an impact on welfare through aggregate productivity, this effect is independent of τ .²²

Corollary 4.6 [$\gamma = \beta$ case]. *If the government discounts the lifetime utility of future cohorts at rate $\gamma = \beta$, then social welfare $\mathcal{W}(g, \tau)$ is equal (up to an additive constant) to average period*

²²The logic behind this result is the same as that of Corollary 2.1: in general equilibrium, there are offsetting forces. If skill prices were exogenous (π_1 fixed), η and τ would interact. See the equations for the skill price and skill quantity in Proposition 2.

utility in the cross section

(32)

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

where the first expectation is taken with respect to the equilibrium cross-sectional distribution of $(\varphi, \alpha_a, s, \varepsilon)$ and the second expectation with respect to the cross-sectional distribution of (s, κ) .²³

Note that this result hinges on skill investment being fully reversible.

V.B. Decomposition of the Social Welfare Function

We now demonstrate that every term in (30) has an economic interpretation and captures one of the forces determining the optimal degree of progressivity. Because these terms are additively separable in the expression for social welfare, the distinct roles of various economic forces are easy to differentiate and quantify. For this decomposition, we focus on the special case $\gamma = \beta$.

V.B.i. Welfare of the Representative Agent

Substituting allocations (18) and (19) into the objective function of the representative agent problem of Section III.A.i. (after solving for λ), one obtains welfare for the representative agent,

$$(33) \quad \mathcal{W}^{RA}(g, \tau) = \log(1 - g) + \chi \log g + (1 + \chi) \frac{\log(1 - \tau)}{(1 + \hat{\sigma})(1 - \tau)} - \frac{1}{1 + \hat{\sigma}},$$

which is precisely the first line of the social welfare expression in (30).

What does equation (33) imply for optimal policy? Differentiation of (33) with respect to g yields the Samuelson condition (31). This value for g equates the marginal rate of substitution between private and public consumption for the representative agent (equal to $\chi g / (1 - g)$) to the technological rate of transformation between the two goods (equal to 1).

Differentiation of (33) with respect to τ yields $\tau^* = -\chi$. Thus, a benevolent government in the representative agent economy would choose *regressive* taxes, with the extent of regressivity proportional to the fraction of output devoted to the public good. To understand why, it is instructive to consider what would go wrong if the government were to instead choose proportional taxes, by

²³We index the uninsurable component of the log wage α by age a to reflect the fact that the presence of permanent shocks implies an age-varying distribution for α .

setting $\tau = 0$. In that case, each agent would not fully internalize the social return to working harder on the margin: although the fraction $(1 - \lambda)$ of additional income paid to the government through taxes benefits everyone in the form of higher public consumption, this gain is neglected by each individual agent, who views himself as atomistic and takes G as exogenous. A regressive tax increases the private return to work, on the margin, and thereby increases labor supply, as is clear from the hours allocation (18). This tax scheme equates social and private returns, a result we restate in the following proposition.

Proposition 5 [RA model]. *If $v_\varphi = v_\omega = v_\varepsilon = 0$ and $\theta = \infty$ (i.e., in a representative agent economy), then $g^* = \chi/(1 + \chi)$ and $\tau^* = -g^*/(1 - g^*) = -\chi$ implement the first best.*

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 . Given $\tau = -g/(1 - g)$, the marginal tax rate at the equilibrium level of income $H^{RA}(\tau)$ is exactly equal to zero (just as it would be with lump-sum taxes). Moreover, given $g = \chi/(1 + \chi)$, the average tax rate is exactly sufficient to finance the optimal level of public goods. Thus, the system replicates a lump-sum tax. Furthermore, as we show in the proof of Proposition 5, this efficiency result extends to a broader class of constant relative risk aversion utility functions of which our baseline logarithmic specification is a special case.

V.B.ii. Welfare from Skill Investment

The second, third, and fourth lines in equation (30) are all related to the skill investment choice. To begin with, from equations (22) and (28), the term that multiplies $(1 + \chi)$ in the second line is the log of aggregate productivity (output per efficiency unit of labor) in the economy:

$$(34) \quad \log \left(\frac{Y(\tau)}{N(\tau)} \right) = \frac{1}{(1 + \psi)(\theta - 1)} \left[\psi \log(1 - \tau) + \log \left(\frac{1}{\eta \theta^\psi} \left(\frac{\theta}{\theta - 1} \right)^{\theta(1 + \psi)} \right) \right].$$

Given imperfect substitutability across skill types in the production technology, the more uniformly dispersed are skills, the higher is aggregate productivity. Higher productivity means higher wages, consumption, and, hence, welfare. Equation (34) indicates that higher progressivity reduces productivity and welfare because it reduces skill investment and thereby compresses the skill distribution toward zero. In terms of its impact on welfare, productivity is multiplied by $(1 + \chi)$

because higher productivity and output boost public consumption (valued in proportion to χ) in addition to private consumption.

Skill investment is not costless. The third line of (30) is the welfare contribution from skill investment costs. The average cost for all past and future cohorts who readjust their skill level is

$$(35) \quad \mathbb{E}[v(s(\kappa; \tau), \kappa)] = \frac{\psi}{1 + \psi} \frac{1}{\theta} (1 - \tau),$$

which is the first term on the third line.²⁴ Skill investment costs decrease in τ because more progressivity reduces skill acquisition. Combining (34) and (35), one can show that the productivity gain from skill investment net of education costs is maximized at $\tau = -1/(\theta - 1) < 0$.

The government also cares about how the choice for τ affects consumption dispersion both directly (via redistribution) and indirectly, via its impact on equilibrium skill prices and quantities. The welfare cost of consumption dispersion across skill types is

$$(36) \quad \text{welfare cost of skill price dispersion} = -\log \left(1 - \left(\frac{1 - \tau}{\theta} \right) \right) - \left(\frac{1 - \tau}{\theta} \right),$$

or the term in the fourth line of (30). This term differs from the familiar Lucas expression for the cost of consumption inequality because consumption dispersion induced by skill heterogeneity follows a Pareto rather than a lognormal distribution. This cost is decreasing in τ because higher progressivity reduces dispersion in after-tax earnings and consumption.

We have learned that offsetting forces determine the optimal level of progressivity with respect to skill acquisition: more progressivity diminishes aggregate productivity but also decreases consumption dispersion across skill types. Which force dominates?

Figure II plots optimal tax progressivity, τ^* , against the elasticity of substitution between skills in production, θ , for three different values for the elasticity of skill investment with respect to the after-tax return to skill, ψ . The first is $\psi = 0$, which corresponds to completely inelastic skill investment and an exogenous distribution of skills identical to the distribution for learning ability κ (see equation 25). The second, $\psi = 0.65$, is our baseline. The third, $\psi = 2$, is a case in which skill investment is highly sensitive to tax progressivity. The plots are constructed assuming that (i) learning ability κ is the only source of heterogeneity in the model, (ii) there is zero taste for public goods ($\chi = 0$), and (iii) the labor supply elasticity parameter, σ , is equal to 2. By virtue of

²⁴If older cohorts already have some skills at the time of the tax reform, they only need to pay net new investment costs, which accounts for the separable term in τ_{-1} in the welfare expression. See Online Appendix B.6 for details.

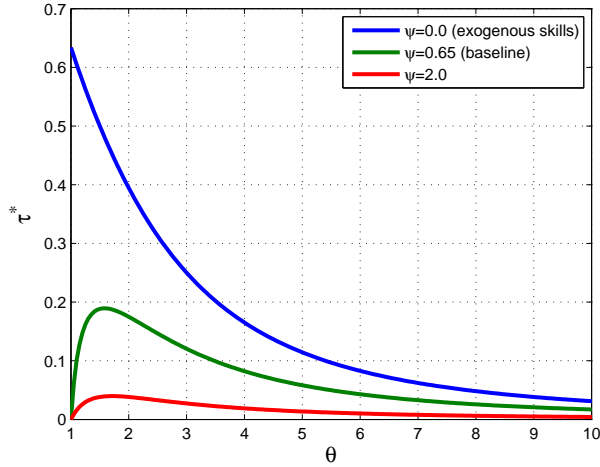


FIGURE II: Optimal τ as a Function of θ in the Special Case $v_\varphi = v_\omega = v_\varepsilon = \chi = 0$ and $\sigma = 2$

Corollary 2.1, these three models deliver identical inequality in wages, hours, and consumption, since the equilibrium distribution for skill prices is Pareto with parameter θ irrespective of ψ .

The figure reveals that skill heterogeneity always dictates positive progressivity ($\tau^* \geq 0$), a result we prove formally in Section V.C. As $\theta \rightarrow \infty$, the economy converges to a representative agent economy with no desire for public goods, so $\tau^* \rightarrow 0$. When we compare the three lines, it is clear that τ^* is decreasing in ψ , so more elastic skill investment implies lower progressivity.

When skills are exogenous ($\psi = 0$), τ^* declines monotonically with θ . The logic is simply that a larger value for θ implies less inequality in wages and a smaller role for redistributive taxation. In contrast, in the cases with $\psi > 0$, the optimal degree of progressivity is nonmonotone in θ . For high values for θ , reducing θ implies a higher optimal value for τ , driven, as in the exogenous skills case, by the utilitarian planner's desire to reduce between-skill consumption dispersion. Now, however, that the planner is also concerned about progressive taxation reducing skill investment and thereby reducing aggregate productivity via the term $\psi / [(1 + \psi)(\theta - 1)] \cdot \log(1 - \tau)$ in (34). This concern becomes quantitatively more important the lower is θ and the more complementary are different skill types in production. As $\theta \rightarrow 1$, the consumption dispersion and progressivity forces exactly balance out, and $\tau^* \rightarrow 0$ —a flat tax system. Thus, optimal progressivity is largest for intermediate values for θ . The finding that the optimal value for τ is a hump-shaped function of θ is quite general: it also applies when labor supply is inelastic ($\sigma \rightarrow \infty$) and extends to the case in which the planner has a Rawlsian *maximin* rather than a utilitarian social welfare function.

V.B.iii. Welfare from Preference Heterogeneity and Uninsurable Wage Risk

The existence of heterogeneity in the preference for leisure, through cross-sectional variation in φ , translates into dispersion in hours worked, earnings, and consumption. The fifth line of the social welfare expression reflects this source of consumption dispersion:

$$(37) \quad \text{welfare cost of } var_{\varphi}(\log c) = (1 - \tau)^2 \frac{v_{\varphi}}{2}.$$

This term is the familiar Lucas expression for the welfare cost of consumption dispersion when the underlying shocks are lognormal: one-half of the φ -driven variance of log consumption times the coefficient of risk aversion, which is equal to 1 in our model.

Uninsurable shocks are another key source of consumption dispersion, and their contribution to social welfare shows up in the sixth line of equation (30):

$$(38) \quad \text{welf. cost of } var_{\alpha}(\log c) = \left[(1 - \tau) \frac{\beta\delta}{\gamma - \beta\delta} \frac{v_{\omega}}{2} - \log \left(\frac{1 - \delta \exp\left(\frac{-\tau(1-\tau)}{2} v_{\omega}\right)}{1 - \delta} \right) \right]$$

When $\gamma = \beta$, this expression is approximately equal to $(1 - \tau)^2 \frac{v_{\alpha}}{2}$, which, symmetrically with the cost of preference heterogeneity, is one-half the α -driven variance of log consumption in the cross section.²⁵ As can be seen from (37) and (38), a higher value for τ reduces consumption dispersion stemming from both preference heterogeneity and uninsurable risk. Since consumption inequality lowers welfare, these two forces push the optimal τ toward 1, the value at which there would be zero consumption dispersion.

V.B.iv. Welfare from Insurable Wage Risk

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

$$(39) \quad \log \left(\frac{N(\tau)}{H(\tau)} \right) = \frac{1}{\hat{\sigma}} v_{\varepsilon}$$

$$(40) \quad var_{\varepsilon}(\log h) = \frac{1}{\hat{\sigma}^2} v_{\varepsilon}.$$

The first term is the log productivity gain from insurable wage variation. As we explained when discussing the equilibrium allocations, more insurable wage dispersion improves welfare because individual hours worked become more positively correlated with individual productivity

²⁵The approximation is extremely accurate for plausible parameter values (see Online Appendix B.6).

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, the cross-sectional variance of log hours due to insurable shocks multiplied by σ , which measures aversion to hours fluctuations.²⁶ The sum of these two terms (the productivity gain net of the disutility costs of hours fluctuations) is maximized at $\tau = 0$, the value at which hours worked respond efficiently to insurable shocks. Thus, greater insurable wage risk will push τ^* toward zero, a flat tax.

V.C. When Should Taxes be Progressive?

By differentiating the expression for social welfare in equation (30) with respect to τ , one can obtain a necessary and sufficient parametric condition for the optimal tax system to be progressive.

Proposition 6 [condition for optimal progressivity]. τ^* is strictly positive if and only if

$$(41) \quad \underbrace{\frac{\psi}{(1+\psi)\theta}}_{\text{lower investment cost}} + \underbrace{\frac{1}{(\theta-1)\theta}}_{\text{less skill price inequality}} + \underbrace{\frac{(v_\varphi + v_\alpha)}{\theta}}_{\text{less } \alpha \text{ and } \varphi \text{ inequality}} > \underbrace{\frac{\psi}{(1+\psi)(\theta-1)}}_{\text{lower productivity due to less inv.}} + \chi \underbrace{\left(\frac{1}{1+\sigma} + \frac{\psi}{(1+\psi)(\theta-1)} \right)}_{\text{lower } G \text{ due to less hours and less inv.}}$$

The terms on the left-hand side of (41) are the marginal benefits from increasing progressivity at $\tau = 0$, and the ones on the right-hand side are the corresponding marginal costs.²⁷ The first term on the left-hand side is the utility gain from lower skill investment costs, the second is the gain from reducing consumption inequality across skill types, and the third is the gain from reducing consumption inequality due to differences in uninsurable productivity shocks and preference heterogeneity. On the right-hand side, the first term is the cost of lower output associated with depressed skill investment. The second term reflects the diminished utility from public consumption due to the decrease in G caused by lower hours worked and lower skill investment.

It is easy to verify that if $\chi = 0$, then the condition for taxes to be progressive is always satisfied. The larger is χ , the larger are the costs of progressivity, making it more difficult to satisfy the condition for $\tau^* > 0$. A higher value for the investment elasticity parameter ψ also makes the

²⁶As with the productivity gain from skill investment, the productivity gain from insurable risk is multiplied by $(1 + \chi)$, reflecting the additional value of an extra unit of output when agents value government expenditures.

²⁷This condition is independent of the insurable variance v_ε because the term in equation (30) involving this component is maximized at $\tau = 0$, so the marginal welfare effect from a change in τ is zero at $\tau = 0$.

condition for $\tau^* > 0$ harder to satisfy, as does a larger labor supply elasticity (i.e., a lower value for σ). In both cases, a stronger behavioral response limits the temptation to compress inequality by increasing progressivity. Conversely, the larger is uninsurable risk, $v_\alpha + v_\varphi$, the more likely it is that $\tau^* > 0$. The effect of θ on optimal progressivity is ambiguous, as discussed in Section V.B.ii.

V.D. Optimal Marginal Tax Rate at the Top

One focus of the Mirrleesian 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 optimal marginal tax rate at the top converges to

$$(42) \quad \bar{t} = \frac{1}{1 + \zeta^u + \zeta^c(\theta - 1)},$$

where ζ^u and ζ^c are uncompensated and compensated labor supply elasticities.²⁸ Thus, the thicker is the right tail of the productivity distribution (i.e., the smaller is the Pareto coefficient θ), the higher is the optimal marginal tax rate at the top.

Now consider a version of our model with $\chi = 0$ and $v_\varphi = v_\omega = v_\varepsilon = 0$, so that heterogeneity in skills reflecting heterogeneity in ability κ is the only motive for taxation. This model is observationally equivalent to the case considered by Saez: there is a Pareto distribution for wages, with Pareto parameter θ . However, the underlying structure of the two models is quite different: wage heterogeneity reflects endogenous dispersion in skill choices in our model but exogenous differences in productivity in the standard Mirrlees setup.

We have not tackled the fully optimal Mirrlees tax problem in our environment with skill investment, so we have no theoretical analogue to equation (42). But within the parametric class of policies described by equation (1) we can compare the optimal policy in our environment in which wages reflect skill investments to the optimal policy when wages are exogenous and taxation only distorts labor supply. Recall that in our model skills are exogenous when $\psi = 0$. Figure II shows that with $\psi = 0$, reducing θ and thereby increasing inequality always implies higher optimal progressivity, and thus higher marginal tax rates at high income levels.²⁹ This result mirrors the

²⁸Given our utility function, as earnings increase, these elasticities converge to zero and $1/(1 + \sigma)$, respectively, and thus the efficient top marginal tax rate—given exogenous wages—would converge to $\bar{t} = (1 + \sigma)/(\sigma + \theta)$.

²⁹From equation (3), the marginal tax rates are strictly increasing in progressivity for any $y \in (\exp(-1/(1-\tau)), \infty)$.

familiar Mirrleesian prescription that top marginal rates should be higher the heavier is the Pareto tail. With endogenous skill choice ($\psi > 0$), in contrast, there is a range of values for θ close to unity in which reducing θ and increasing wage inequality lowers optimal progressivity and thus marginal tax rates at high income levels. Recall the logic for this result: the more complementary are skill types, the larger are the productivity gains from a more uniformly dispersed skill distribution and thus the more costly are high tax rates at the top that discourage skill investment by high ability individuals.

Our finding that more inequality at the top reduces optimal progressivity, in contrast to the standard Mirrlees prescription, has an interesting parallel in Scheuer and Werning (2015), who show that introducing superstar effects in a Mirrleesian framework reduces optimal marginal tax rates at the top, even though these effects increase inequality.

VI. QUANTITATIVE ANALYSIS

In this section, we describe the model parametrization and explore the quantitative implications of the theory. Next, we perform a robustness analysis with respect to (i) the assumption that past skill investment is flexible, (ii) the size of government purchases, (iii) the degree of inequality aversion embedded in the welfare function, and (iv) an alternative political-economic driver of policy.

VI.A. Parameterization

Thanks to the closed-form solution for 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 allow us to prove identification of all parameters and to estimate the model given empirical counterparts of these moments computed from commonly used micro data.

We begin by recognizing that in survey data, hours worked and consumption are measured with error, and hourly wages (computed as annual earnings divided by annual hours) inherit 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 classical. If we tack on measurement error to log wages and the log allocations in (20) and (21), and compute variances

and co-variances of their joint cross-sectional distribution, we obtain:

$$\begin{aligned}
(43) \quad \text{var}(\log w) &= \frac{1}{\theta^2} + v_\alpha + v_\varepsilon + v_{\mu y} + v_{\mu h} \\
\text{var}(\log h) &= v_\varphi + \frac{1}{\hat{\sigma}^2} v_\varepsilon + v_{\mu h} \\
\text{var}(\log c) &= (1 - \tau)^2 \left(v_\varphi + \frac{1}{\theta^2} + v_\alpha \right) + v_{\mu c} \\
\text{cov}(\log h, \log w) &= \frac{1}{\hat{\sigma}} v_\varepsilon - v_{\mu h} \\
\text{cov}(\log h, \log c) &= (1 - \tau) v_\varphi \\
\text{cov}(\log w, \log c) &= (1 - \tau) \left(\frac{1}{\theta^2} + v_\alpha \right).
\end{aligned}$$

These moments contain most of the structural parameters of the model. The variance of the uninsurable innovation v_ω is implied by v_α , given a value for δ .³⁰

Based on our previous work (Heathcote, Storesletten, and Violante, 2014a), we set $\sigma = 2$, a value broadly consistent with the microeconomic evidence on the Frisch elasticity (see, e.g., Keane, 2011). From the same paper we set the variances of measurement error to $v_{\mu h} = 0.036$, $v_{\mu y} = 0$, and $v_{\mu c} = 0.040$. In light of our estimate of the progressivity of the U.S. tax/transfer system above, we set $\tau = 0.181$.

It is easy to see that v_φ , v_ε , and $(v_\alpha + 1/\theta^2)$ are overidentified by the set of moments in (43). To separately identify the cross-sectional variance of uninsurable risk, v_α , 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 $a = 0$, which reflect only variation in skills acquired before labor market entry, since $v_\alpha^0 = 0$ by assumption.

Our data are drawn from two surveys, the Panel Study of Income Dynamics (PSID) for years 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 II. 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 eqs. (43) plus the three moments at age “zero” (an average of ages 25–29 in the data).³¹ The minimum distance procedure

³⁰The variance of α at age a is $v_\alpha^a = av_\omega$ so the cross-sectional uninsurable variance in the model is $v_\alpha = (1 - \delta) \sum_{a=0}^{\infty} \delta^a v_\alpha^a = v_\omega \delta / (1 - \delta)$.

³¹The resulting empirical moments used in the estimation are $\text{var}(\log w) = 0.43$, $\text{var}(\log h) = 0.11$, $\text{var}(\log c) =$

TABLE I: Parameterization

	δ	χ	σ	$v_{\mu h}$	$v_{\mu y}$	$v_{\mu c}$	v_{φ}	v_{ε}	v_{α}	θ	v_{ω}
Baseline	0.971	0.233	2.00	0.036	0	0.040	0.036	0.164	0.098	3.124	0.003
	–	–	–	–	–	–	(0.002)	(0.010)	(0.007)	(0.114)	(0.0003)
Alternative	0.971	0.233	2.00	0.036	0	0.040	0.023	0.139	0	2.0	0
	–	–	–	–	–	–	(0.002)	(0.012)	(0.000)	–	(0.000)

Source: PSID and CEX, 2000-2006. See the main text for details. Bootstrapped standard errors based on 500 replications in parentheses.

therefore uses nine moments to estimate four parameters $(v_{\varphi}, v_{\varepsilon}, v_{\alpha}, \theta)$. We set $\delta = 0.971$ to match an expected working life of 35 years, the same age span considered in the micro-data. All parameter values, including those that are predetermined, are summarized in Table I.

The estimates in Table I imply that (i) the insurable component accounts for 45% of the model variance of wages, while the uninsurable component and heterogeneity in skills each account for just over 25%; (ii) cross-sectional dispersion in the disutility of work effort explains 60% of model hours variation, while insurable shocks explain the remaining 40%;³² (iii) dispersion in the disutility of work accounts for 15% of consumption inequality, while uninsurable wage risk and skill heterogeneity are equally important and account for the remaining 85%; (iv) the growth in the variance of log consumption over the life cycle (ages 25-60) is around 0.10. These findings are in line with those of Heathcote, Storesletten, and Violante (2014a).

Our estimated value for θ is just above 3. An alternative way to calibrate θ would be to exploit the fact that the top end of the model income distribution is approximately Pareto with parameter θ , so that $E[p(s) | s > \bar{s}] / p(\bar{s}) = \theta / (\theta - 1)$. From our PSID sample, we estimate that this ratio is stable and around 2 for income thresholds above \$250,000, which points toward $\theta = 2$.³³ Table I reports how parameter estimates change when we impose $\theta = 2$ and re-estimate other parameters. We will examine how our results are affected by this alternative parameterization.

The parameter ψ controls the elasticity of the skill premium to τ and θ , where the skill premium

0.18, $cov(\log h, \log w) = -0.01$, $cov(\log h, \log c) = 0.03$, $cov(\log w, \log c) = 0.15$, $var^0(\log w) = 0.28$, $var^0(\log c) = 0.15$, and $cov^0(\log w, \log c) = 0.10$.

³²A key reason to allow for heterogeneity in φ is to be able to account for the large empirical dispersion in hours worked.

³³Tax return data on wage income tabulated by Piketty and Saez (2003, Table B3) indicate a value for θ between 1.6 and 2.2 for the years 2000-2006, depending on the threshold \bar{s} . Thus, our estimate falls within this range.

is increasing in τ and decreasing in θ (see equation 23). Thus, if we can measure the changes over time in the factors that control the relative demand and relative supply of different skill types, $\Delta \log \theta$ and $\Delta \log(1 - \tau)$, as well as the corresponding change in the skill premium, $\Delta \log(\pi_1)$, we can identify ψ . Taking time differences of logs and rearranging expression (23), we arrive at

$$(44) \quad \psi = -\frac{\Delta \log \theta + \Delta \log \pi_1}{\Delta \log(1 - \tau) + \Delta \log \pi_1}.$$

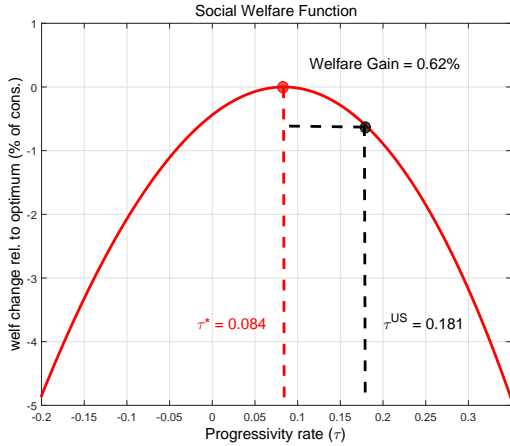
Inferring a value of ψ from this equation requires estimates for changes in the three variables on the right-hand side. We measure these three changes between 1970-1976 and 2000-2006.

When we estimate τ for these two periods, we obtain $\Delta \log(1 - \tau) = 0.034$.³⁴ To estimate the change in the return to skill, we need to take a stand on a measure of skills in the data. The most natural and easily measurable is years of education. We therefore estimate returns to skill on our PSID sample through a Mincerian regression of log hourly wages on years of education, controlling for race and a cubic function of age. We run two separate regressions, one for men and one for women, and average returns to education between the two groups. We obtain $\pi_1^{70-76} = 0.088$ and $\pi_1^{00-06} = 0.118$ and thus $\Delta \log \pi_1 = 0.29$. We cannot use the change in consumption inequality at age of labor market entry in order to estimate $\Delta \log \theta$, since CEX data are not available for the 1970s. Instead, we exploit the property that θ is the Pareto-tail coefficient of the income distribution. We find $\theta^{70-76} = 3.3$ and $\theta^{00-06} = 2.0$, which implies $\Delta \log \theta = -0.50$.³⁵ Plugging these numbers into equation (44) yields $\psi = 0.65$.

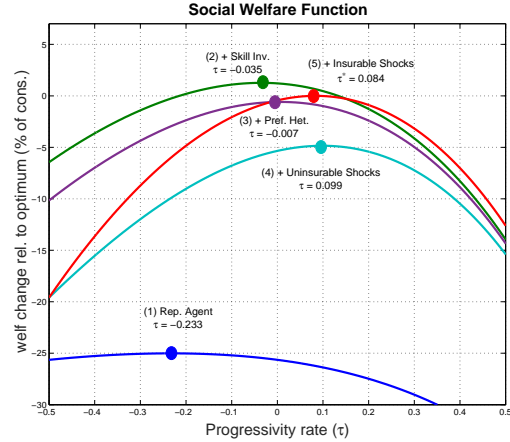
Finally, to set a value for χ , the relative weight on the government-provided good in preferences, we take the view that the fraction of output devoted to publicly provided goods g is chosen efficiently. In Section VI.F. we show that all agents agree that g should be set to the efficient level $g^* = \chi/(1 + \chi)$. This provides a theoretical motivation for our calibration choice. Over the period

³⁴We modify slightly our estimation strategy for τ outlined in Section II. The reasons are that (i) TAXSIM does not accept the state ID before 1977, and thus one cannot compute state taxes paid, and (ii) medical expenses are only available in the PSID after 1999. To preserve consistency across the earlier and later time periods, we therefore (i) exclude state taxes from tax liabilities in both periods and (ii) use gross income instead of taxable income in both periods. The new estimate of τ^{00-06} is 0.145. For the period 1970-1976, we find higher progressivity: $\tau^{70-76} = 0.174$ ($SE = 0.002$) with $R^2 = 0.92$, based on a regression on 12,977 observations.

³⁵Note that we should expect to find a decline in θ over time. Suppose, to the contrary, that θ was constant. Then the observed reduction in tax progressivity should have induced an increase in the supply of relatively high skills and a decline in the pretax skill premium. But in the data, the skill premium increased. Thus, we should expect to find a relatively large decline in θ such that the resulting increase in the relative demand for high skills dominates the effect of an increase in relative supply (via less progressive taxation) in terms of the net effect on the skill premium.



(a) Social welfare function



(b) Decomposition of the social welfare function

FIGURE III: Panel (a): Social welfare as a function of τ and welfare gain relative to the current U.S. system. Panel (b): Decomposition of social welfare into the components described in Section V.B.

2000-2006, $g = G/Y = 0.189$, and we therefore set $\chi = 0.233$. Because the optimal τ is sensitive to the level of government purchases, in Section VI.D. we discuss alternative scenarios.

VI.B. Results

Once the optimality condition $g^* = -\chi$ is substituted into (30) and values have been assigned to all the structural parameters, one obtains social welfare $\mathcal{W}(\tau)$ as a function of τ only. Figure III(a) plots this function, assuming $\gamma = \beta$. The value of progressivity that maximizes social welfare is $\tau^* = 0.084$. The average welfare gain from reducing progressivity from the current value of $\tau^{US} = 0.181$ to τ^* is equivalent to 0.63% of lifetime consumption.

How different are the actual and optimal schemes? Note that the ratio of the variance of log disposable income to pregovernment income is $(1 - \tau)^2$. Moving to the optimal scheme would increase this ratio from 0.67 to 0.84. The average income-weighted marginal tax rate would drop from 34% to 26%. Figure D2 in Online Appendix D plots average and marginal tax rates for the actual and optimal tax schemes.

Figure III(b) reconstructs $\mathcal{W}(\tau)$ by sequentially adding all of its components. The first component is welfare of the representative agent. As discussed, this is maximized at $\tau = -\chi = -0.233$. Adding the skill investment component (the productivity gain from skill investment net of edu-

cation costs minus the welfare loss from between-skill consumption inequality) pushes toward a more progressive system, and the optimal τ increases to -0.035 . The concern for additional consumption inequality induced by preference heterogeneity further raises the optimal τ to -0.007 . Uninsurable shocks are a stronger source of consumption dispersion, which is reflected in the substantial upward jump in τ to 0.099 when this component is incorporated. Finally, adding the productivity gain from insurable shocks pulls τ back toward zero to its final value of 0.084 .

If consumption inequality was the government's only concern, τ would be optimally set to one, and the tax/transfer scheme would equate postgovernment income and consumption across households. Besides the need to fund public consumption, which we analyze in Section VI.D., two forces limit progressivity: the distortion to labor supply and the distortion to skill investment.

To measure the strength of these two channels, we compute the optimal τ in two special cases, one in which labor supply is inelastic and a second in which the distribution of skills and skill prices is exogenous.³⁶ With inelastic hours the optimal τ is 0.281 , whereas with inelastic skills the optimal τ is 0.202 . Thus, the endogeneities of labor supply and skill investment both play quantitatively important roles in limiting optimal progressivity. Note, in particular, that absent endogenous skill investment, the welfare-maximizing policy would imply an increase in progressivity relative to the current tax system, whereas reducing progressivity is optimal once this margin is incorporated.

The alternative model calibration described in Table I reinforces this latter result. With θ fixed at 2, skill heterogeneity generates more wage and consumption dispersion, and the estimation therefore sets the variance of the uninsurable lifetime shocks v_ω to zero. A larger role for endogenous skills in generating wage dispersion translates into lower optimal progressivity: $\tau^* = 0.034$ corresponding to an average marginal rate of 22%.

VI.C. Progressivity When Past Skill Investment Is Fixed

Our baseline model assumes that past investments can be freely reversed or supplemented if the tax system changes. This assumption implies that (i) the optimal choice for τ is independent of the preexisting distribution of skills, and (ii) transition following any tax reform is instantaneous. We now consider the opposite extreme assumption, namely that skill investment is chosen once

³⁶This latter case is obtained by excluding from the welfare function the first two terms associated with the productivity gain from skill investment net of the education cost.

and for all at age zero and can never be adjusted thereafter. This introduces an additional force in the direction of more progressivity: the planner can now reduce consumption inequality without distorting skill investments for agents who entered the economy in the past.

Consider an unanticipated once-and-for-all change in the tax system from (τ_{-1}, g_{-1}) to (τ, g) . With skill investment fixed at age 0, output gradually evolves over time as the population share of agents who make skill investments under the new tax regime rises. Because the policy parameters (τ, g) are assumed constant during transition, the budget-balancing value for λ is time varying. To maintain tractability in this version of the model, we slightly modify the production structure by assuming that production is segregated by age groups. This assumption preserves tractability because the density of skills within a cohort remains exponential during the transition, whereas the economy-wide distribution does not. See Online Appendix D.3 for details.

When we set $\gamma = \beta$ and assume that the initial steady state corresponds to our estimated progressivity value, $\tau_{-1} = \tau^{US} = 0.181$, we find that the optimal permanent choice for τ is now 0.146, compared to 0.084 in the baseline model with flexible investment. The optimal choice for g is unchanged. In the fixed investment model, the planner can use progressive taxation to effectively expropriate the returns to past skill investments made by currently living individuals. The importance of this effect depends on the planner's intergenerational discount factor, γ . As $\gamma \rightarrow 1$, the fact that past skill investments can be expropriated becomes irrelevant, and in both the fixed and flexible investment models, the expression for social welfare corresponds to expected lifetime utility for a newborn agent in steady state, implying a common welfare-maximizing value for τ of 0.061.³⁷ The lower is γ , and thus the less concerned is the planner about the future, the larger is the incentive for expropriation in the fixed investment specification. Thus, reducing γ widens the difference in optimal progressivity relative to the baseline flexible investment model in which past investments can be costlessly reversed and thus cannot be expropriated (see Figure D3 in Online Appendix D.3).

³⁷This value for τ is smaller than the baseline optimum ($\tau = 0.084$) because putting more relative weight on future generations (via a higher γ) weakens the planner's desire to compress consumption inequality among existing older agents (see the sixth line of equation 30).

VI.D. Modeling Public Consumption

A theme of this paper is that there is an important interaction between the size of government and the desired progressivity of the tax and transfer system. Holding fixed all other structural parameters and the choice of the social welfare function, the optimal tax and transfer system should be less progressive the larger is desired government consumption (see Sections V.B.i. and V.C.). In contrast to our paper, most previous analyses of tax design abstract from the choice of public good provision and simply assume that an exogenous level of nonvalued expenditure must be financed. We therefore now consider a version of the model in which the level of G is exogenous, denoted \bar{G} , but where the government still chooses τ .

It turns out that increasing the exogenous value \bar{G} works just like increasing the preference parameter χ in the baseline endogenous G model: higher \bar{G} implies lower optimal progressivity, denoted by $\tau_{\bar{G}}^*$. Again, the planner internalizes that less progressive taxation encourages labor supply and skill investment, and makes it easier to finance expenditure. If \bar{G} is equal to the level that is optimally chosen in our baseline endogenous G model, so that $\bar{G} = g^{US}Y(\tau^*)$, then $\tau_{\bar{G}}^*$ is equal to the optimal choice $\tau^* = 0.084$ in the baseline model.

However, the results are quite different if instead of fixing G exogenously in levels, we model expenditure as a fixed exogenous share \bar{g} of output. We maintain the assumption that government consumption is nonvalued and set $\chi = 0$. From Corollary 4.3, the optimal choice for τ in the exogenous expenditure model is independent of \bar{g} . In fact, for any value for \bar{g} the optimal value $\tau_{\bar{g}}^*$ is equal to 0.20, which is optimal degree of progressivity in the baseline endogenous G model in which agents derive no utility from public goods ($\chi = 0$).

Table II summarizes all these different cases, while Figure D2 in Online Appendix D plots the corresponding average and marginal tax rates. The key takeaway is that differences in how government spending is modeled can have a large impact on optimal progressivity. We conclude that government consumption and the distribution of the associated tax burden are interrelated policy choices that should be analyzed within a single integrated framework.

VI.E. Inequality Aversion

Our baseline analysis assumes a utilitarian social welfare function. In that case, the planner is just as motivated to use progressive taxation to reduce permanent consumption differentials across

TABLE II: Optimal Progressivity under Alternative Models for Government Spending

	Endogenous G	Exogenous G	Exogenous g
	τ^*	$\tau_{\bar{G}}^*$	$\tau_{\bar{g}}^*$
$\chi = 0$	0.200	$G = 0$	$\bar{g} = 0$
$\chi = 0.233$	0.084	$\bar{G} = g^{US}Y(\tau^*)$	$\bar{g} = g^{US}$

Note: In the exogenous G and exogenous g cases we always set $\chi = 0$.

individuals as it is to dampen consumption fluctuations for an individual over time. We now explore a more general formulation, building on Bénabou (2002), which allows us to vary the planner's concern for redistribution with respect to preexisting inequality.

Let $U(\kappa, \varphi, \alpha; g, \tau)$ denote expected lifetime utility for an agent with characteristics $(\kappa, \varphi, \alpha)$ prior to drawing the current period insurable shock (see equation 45), and let $\bar{c}(\kappa, \varphi, \alpha; g, \tau)$ denote the certainty equivalent value for consumption that delivers $U(\kappa, \varphi, \alpha; g, \tau)$, assuming equilibrium decision rules for hours and skill investment. Thus,

$$U(\kappa, \varphi, \alpha; g, \tau) = \log \bar{c}(\kappa, \varphi, \alpha; g, \tau) - \frac{1 - \tau}{1 + \sigma} + \chi \log G(g, \tau) - [(1 - \tau) - \mathbb{I}_{\{a > 0\}}(1 - \tau_{-1})] \frac{\psi}{(1 + \psi)\theta} \eta \kappa,$$

where the second term on the right-hand side is lifetime expected disutility from labor supply, the third is lifetime utility from public consumption, and the fourth is the skill investment cost, which depends on whether the agent is newborn or an older individual revising her investment decision given the change in progressivity.

The contribution to welfare from all agents of age a at the onset of the tax reform, $\mathcal{V}_a(g, \tau)$, is

$$\begin{aligned} \mathcal{V}_a(g, \tau) &= \log \left(\int \int \int \bar{c}(\kappa, \varphi, \alpha; g, \tau)^{1-\nu} dF_\kappa dF_\varphi dF_\alpha^a \right)^{\frac{1}{1-\nu}} - \frac{1 - \tau}{1 + \sigma} + \chi \log G(g, \tau) \\ &\quad - [(1 - \tau) - \mathbb{I}_{\{a > 0\}}(1 - \tau_{-1})] \frac{\psi \eta}{(1 + \psi)\theta} \int \kappa dF_\kappa, \end{aligned}$$

where $\nu \in (0, \infty)$ is a new parameter that defines the planner's aversion to within-cohort inequality in expected lifetime utility from consumption due to heterogeneity in $(\kappa, \varphi, \alpha)$. The larger is ν , the less substitutable across agents is lifetime utility, and the stronger is the planner's desire to equalize consumption through redistribution. Note that while the inequality-averse planner's objective function allows for a flexible degree of aversion to inequality in lifetime utility from consumption, we have assumed that the planner remains utilitarian with respect to skill investment costs and disutility from labor supply, which are simply averaged in the usual way.

In terms of how the planner weights different cohorts, we assume, as before, a linearly additive formulation and set $\gamma = \beta$ so the planner discounts across generations at the same rate as agents discount over time. Current and future newborn agents are treated symmetrically and contribute $\mathcal{V}_0(g, \tau)$ to social welfare (recall that transition is immediate). Thus, the planner's objective is

$$\mathcal{W}^\nu(g, \tau) = (1 - \beta) \frac{1 - \delta}{1 - \beta\delta} \left[\sum_{a=1}^{\infty} \delta^a \mathcal{V}_a(g, \tau) + \sum_{j=0}^{\infty} \beta^j \mathcal{V}_0(g, \tau) \right].$$

Following algebra similar to that for the baseline utilitarian welfare expression, we can solve for $\mathcal{W}^\nu(g, \tau)$ in closed form.

Proposition 7 [inequality aversion]. *In the model with reversible investment and $\gamma = \beta$, social welfare from implementing policy (g, τ) for a planner with inequality aversion parameter ν is*

$$\mathcal{W}^\nu(g, \tau) = \begin{cases} \log(1 - g) + \chi \log g + (1 + \chi) \frac{\log(1 - \tau)}{(1 + \delta)(1 - \tau)} - \frac{1}{1 + \delta} & (1) \\ + (1 + \chi) \frac{1}{(1 + \psi)(\theta - 1)} \left[\psi \log(1 - \tau) + \log \left(\frac{1}{\eta \theta^\psi} \left(\frac{\theta}{\theta - 1} \right)^{\theta(1 + \psi)} \right) \right] & (2) \\ - \frac{\psi}{(1 + \psi)\theta} \left[(1 - \tau) - \frac{\delta(1 - \beta)}{1 - \beta\delta} \right] & (3) \\ + \log \left(1 - \frac{1 - \tau}{\theta} \right) + \frac{1}{\nu - 1} \log \left((\nu - 1) \frac{1}{\theta} (1 - \tau) + 1 \right) & (4) \\ - (1 - \tau)^2 \nu \frac{v_\varepsilon}{2} & (5) \\ - (1 - \tau)^2 \nu \left(\frac{1 - \beta}{1 - \beta\delta} \right) \left(\frac{\delta}{1 - \delta} \frac{v_\omega}{2} \right) & (6) \\ - (1 - \tau)^2 \frac{\beta(1 - \delta)}{1 - \beta\delta} \left(\frac{\delta}{1 - \delta} \frac{v_\omega}{2} \right) & (7) \\ + (1 + \chi) \left[\frac{1}{\delta} v_\varepsilon - \sigma \frac{1}{\delta^2} \frac{v_\varepsilon}{2} \right]. & (8) \end{cases}$$

Lines (1), (2), (3), and (8) are identical to the corresponding lines of our baseline welfare expression (equation 30) and capture (1) welfare for a representative agent, (2) aggregate productivity gains due to skill investment, (3) the average net cost of skill investment, and (8) the welfare gains associated with insurable risk.³⁸ The other components of social welfare reflect uninsurable heterogeneity and risk, and it is here that the planner's attitude toward innate inequality (indexed by ν) matters. Line (4) is the cost of consumption inequality reflecting differences in skills, which in turn are driven by heterogeneity in κ . Line (5) captures the welfare cost of inequality due to uninsurable preference heterogeneity, and line (6) is the cost of inequality due to uninsurable productivity heterogeneity reflecting past uninsurable ω shocks. Line (7) captures the cost of future increases in consumption inequality due to future ω shocks. Note that ν does not appear in this term, since the planner values insurance against future shocks just as agents themselves do.

³⁸On line (3), to simplify notation, we have omitted the term in τ_{-1} since Corollary 4.1 applies here as well.

TABLE III: Optimal Progressivity under Alternative Values for Inequality Aversion

Inequality Aversion ν	Optimal progressivity τ^*			
	Baseline	$(v_\varphi, v_\omega, v_\varepsilon) = 0$	$\chi = 0$	$(v_\varphi, v_\omega, v_\varepsilon) = 0, \chi = 0$
∞ (Rawlsian)	1.000	0.271	1.000	0.394
2	0.190	0.051	0.295	0.187
1 (Utilitarian)	0.084	-0.035	0.200	0.115
0 (Inequality-neutral)	-0.159	-0.392	-0.026	-0.129

We now consider the implications of alternative choices for the parameter ν . First, as $\nu \rightarrow 1$, the expression for $\mathcal{W}^\nu(g, \tau)$ collapses to our original social welfare function $\mathcal{W}(g, \tau)$ defined in equation (30). Thus, $\nu = 1$ corresponds to a utilitarian planner.

Second, as we increase ν , the planner becomes more averse to inequality in lifetime utility. A planner with $\nu = 2$ (i.e., a relative risk aversion of 2 over individual consumption equivalents) would choose $\tau = 0.19$, just above our estimate for the current U.S. tax/transfer system.

In the limit as $\nu \rightarrow \infty$, the planner is effectively Rawlsian. In this case, the welfare costs of inequality due to preference heterogeneity and past uninsurable shocks (lines 5 and 6) become arbitrarily large, which pushes the optimal value for τ toward 1 as long as either $v_\varphi > 0$ or $v_\omega > 0$. However, the cost of consumption inequality due to differences in skills always remains finite: the term in line (4) converges to $\log[1 - (1 - \tau)/\theta]$. This term does not explode because skill prices are bounded away from zero, and beyond a certain point, a marginal increase in progressivity will reduce the low-skilled pretax wage by more than it increases net transfers.

Third, as $\nu \rightarrow 0$, lines (4), (5), and (6) drop out, indicating that the planner is indifferent to consumption inequality due to initial heterogeneity in κ and φ and to lifetime inequality due to cumulated past ω shocks. Line (7) is still present, though, revealing that the planner still values insurance against future uninsurable shocks.

Table III reports optimal levels for τ for a range of values for the inequality aversion parameter ν and for four alternative parameterizations of the model: our baseline parameterization, a version in which κ is the only source of heterogeneity, and those two cases repeated but with $\chi = 0$.

VI.F. Political-Economic Determination of Progressivity

After our characterization of optimality, it is natural to ask the following question: if (g, τ) were determined through a political-economic mechanism, how would the voting outcome differ from the policy 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 skill decision is reversible, so the transition to a new steady state is immediate.

The challenge in analyzing a political-economic version of our model is twofold. First, voting has two dimensions, (g, τ) . 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 τ , agents agree on the amount of G to be provided, and this amount is a fraction $\chi/(1 + \chi)$ of aggregate output; (ii) notwithstanding multidimensional heterogeneity, the attitude of individual agents toward progressivity τ can be summarized by a single summary statistic, so voters differ effectively along only one dimension. As we shall see, these properties will ensure that the median voter theorem applies.

We begin by proving that all agents agree on the optimal size of government. This is because increasing g delivers identical marginal utility gains to all agents from greater public consumption ($\frac{\lambda}{g}$) and identical marginal losses via associated higher taxation ($\frac{1}{\lambda} \cdot \frac{\partial \lambda}{\partial g}$).

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

We now consider voting over τ and make an additional simplifying assumption: voting occurs before the realization of the i.i.d. insurable shock ε , so the individual state vector at the voting stage is $(\kappa, \varphi, \alpha)$. For an agent with characteristics $(\kappa, \varphi, \alpha)$, expected lifetime utility (ignoring, as usual, the separable term in τ_{-1}) is given by

$$\begin{aligned}
 U(\kappa, \varphi, \alpha; g, \tau) &= \mathcal{W}(g, \tau) - (1 - \tau) \frac{\beta \delta}{1 - \beta \delta} \frac{v_\omega}{2} \\
 &\quad + (1 - \tau) \left(\frac{\beta \delta}{\gamma - \beta \delta} \frac{v_\omega}{2} + \frac{v_\varphi}{2} - \frac{1}{\theta} + \frac{\psi}{1 + \psi} \frac{1}{\theta} \right) \\
 (45) \quad &\quad + (1 - \tau) \left(\alpha - \varphi + \frac{\eta}{\theta} \kappa - \frac{\psi}{1 + \psi} \frac{\eta}{\theta} \kappa \right).
 \end{aligned}$$

Proposition 9 [median voter]. *Under majority rule voting, the equilibrium τ is the value that*

maximizes equation (45) for the agent with the median value for the random variable $x = \alpha - \varphi + \frac{1}{1+\psi} \frac{\eta}{\theta} \kappa$.

The median voter theorem applies because preferences are single peaked in τ . This follows from the concavity of $\mathcal{W}(g, \tau)$ (Corollary 4.2). Concavity in $\mathcal{W}(g, \tau)$ translates into concavity in $U(\kappa, \varphi, \alpha; g, \tau)$ since the additional terms in equation (45) are linear in τ . From the last line of this equation it can be seen that, for the purposes of characterizing attitudes to progressivity, the three-dimensional vector $(\kappa, \varphi, \alpha)$ can be collapsed into the sufficient statistic x . Since x is a linear combination of normal (φ, α) and exponential (κ) variables, it is an EMG random variable.

We find that the median voter—the agent with the median value for x —would choose $\tau^{med} = 0.144$, a choice that is significantly higher than the utilitarian planner’s choice ($\tau^* = 0.084$). To understand why the median voter prefers more progressivity, note first that in the limiting case $\theta \rightarrow \infty$ and $v_\omega = 0$, $U(x^{med}; g, \tau) = \mathcal{W}(g, \tau)$, so she would choose exactly the same progressivity as the utilitarian planner. Thus, the reason the median voter prefers a higher value for τ has to do with the existence of permanent uninsurable shocks ($v_\omega > 0$) and heterogeneity in skill prices ($\theta < \infty$). With respect to the former, the median voter wants a higher τ because, in contrast to the utilitarian planner, she does not care about future cohorts whose initial uninsurable shock dispersion will be low. With respect to skill heterogeneity, the agent with median ability κ has less than average ability because κ is exponentially distributed. She therefore prefers more redistribution.

VII. SKILL INVESTMENT CONSTRAINTS

So far, our analysis has abstracted from potential constraints that might limit skill investments, such as inadequate parental resources or credit constraints. We now consider an extension to the model in which some workers do not get a chance to accumulate skills because their families are so poor that they cannot afford any skill investment. The welfare gains from progressive taxation are now potentially larger because redistribution can help lift families out of poverty and thereby broaden opportunities for skill acquisition. In the context of schooling, several authors have argued that redistribution could increase educational investment by relaxing borrowing constraints (e.g., Fernández and Rogerson 1998; Bénabou 2002; and Seshadri and Yuki 2004. For a survey, see Lochner and Monge-Naranjo 2012).

When a worker dies (which happens with probability $1 - \delta$), she is replaced by a child. There is no intergenerational altruism, so a child does not receive any bequests from the parent-worker she replaces. The fixed effects (φ, κ) are assumed independent across parents and children.

Workers can be either skilled or unskilled, and the fate of a newborn worker is sealed at birth. Independently of parental income, a share Λ of the newborn are unable to invest in skills and therefore become permanently unskilled (e.g., because of a high psychic cost of schooling). The opportunities of the remaining share $(1 - \Lambda)$ of the newborn hinge on the financial resources of their parents. In particular, a newborn child can make a skill investment only if her parent's consumption exceeds a threshold \underline{c} in the last period of the parent's life. Children who are lucky enough to get a skill investment opportunity are exactly like the workers analyzed in the previous sections: they draw an investment cost κ , choose an investment level s , and earn a wage per efficiency unit $p(s)$. All unskilled workers earn the same wage \underline{w} . Both \underline{w} and \underline{c} are exogenous parameters.

The law of motion for the share of unskilled workers ξ_t is

$$(46) \quad \xi_{t+1} = \delta \xi_t + (1 - \delta) \xi_t \cdot [\Lambda + (1 - \Lambda) \Pr(c < \underline{c} \mid \text{unskilled})] \\ + (1 - \delta) (1 - \xi_t) [\Lambda + (1 - \Lambda) \Pr(c < \underline{c} \mid \text{skilled})].$$

Note that the larger is Λ , the more children become unskilled for reasons unrelated to parental resources, and hence the smaller is the scope for influencing human capital through redistribution.

Equilibrium labor supply for all workers in this model remains as described in Proposition 2. Consumption for skilled workers is also as in Proposition 2, up to a different value for λ . Consumption for the unskilled has the same form, except that the skill price $p(s)$ is replaced by \underline{w} . Given the consumption rule, we can evaluate the probabilities that skilled and unskilled parents' consumption exceeds the threshold \underline{c} , which in turn determine intergenerational transition probabilities. Online Appendix B.14 contains all these expressions and additional details on this model extension.

As before, we focus on tax reforms in which the progressivity parameter τ is changed once and for all. This analysis must be done numerically because changes in τ induce transitional dynamics in the share of unskilled workers, ξ_t , according to equation (46).

We now describe how the new parameters \underline{w} , \underline{c} , and Λ are calibrated (all other parameter values are unchanged). We identify an unskilled worker as a high school dropout. In 2000, the

TABLE IV: Optimal Progressivity with Skill Investment Constraints

	Long-run ξ at τ^*	τ^*	Welfare gain: τ^{US} to τ^*
Baseline (exogenous $\xi = 0$)	0	0.084	0.62%
Extension with exogenous $\xi = 17.3\%$	17.3%	0.132	0.17%
Extension with endogenous ξ	17.2%	0.168	0.01%

median labor earnings of full-time high school dropouts was 54% of the median earnings of all full-time workers with more than a high school education.³⁹ This pins down the unskilled wage w . According to the National Longitudinal Survey of Youth 1997, the probability of a child becoming a high school dropout given that her parent is a high school dropout is 35.5%. The corresponding number when the parent is a high school graduate is 13.5%.⁴⁰ We use these two moments as targets for the two parameters Λ and \underline{c} . The implied parameter values are $\Lambda = 12.5\%$ and \underline{c} equal to 50% of median consumption. These parameter values imply that 26.3% (1.2%) of children of high school dropouts (graduates) live in families that fall below the consumption threshold. Thus, the calibrated model has the features that (i) most children of unskilled parents have a chance of becoming skilled because the consumption floor is quite low; (ii) very few skilled parents are so poor that their children cannot afford to become skilled, although some of those children become unskilled for exogenous reasons; and (iii) tax policy has little effect on the rate at which children of skilled parents become unskilled, whereas it can influence the rate at which children of unskilled parents become skilled.

This model extension increases the optimal degree of tax progressivity significantly compared to our baseline. Two forces are at work. First, the mere presence of unskilled workers induces the utilitarian planner to increase redistribution, both because these unskilled workers are relatively poor and because these workers' skill investments are necessarily insensitive to the tax system. If we hold the share of unskilled workers fixed at the initial steady-state level associated with τ^{US} , thereby abstracting from the effects of τ on the dynamics of ξ_t , this force alone would induce the planner to choose notably higher progressivity than in the baseline model: $\tau^* = 0.132$ (second row

³⁹Authors' computation based on the following Bureau of Labor Statistics data sources: Employment status of the civilian population 25 years and over by educational attainment, Labor Force Statistics Table A-4; Median usual weekly earnings of the employed population 25 years and over by educational attainment.

⁴⁰The calculations are based on the sample used in Abbott et al. (2013). These probabilities are obtained by averaging across the probabilities of children becoming dropouts conditional on their mother's and father's education. If these probabilities were constant over time, the steady-state unskilled share ξ would be 17.3%.

of Table IV), up from $\tau^* = 0.084$ in the absence of any unskilled workers.

The second force is that the planner now wants to use redistribution to influence the share of unskilled workers. The long-run steady-state share $\xi(\tau)$ turns out to be a U-shaped function of τ . This is a result of two opposing effects of tax progressivity on ξ . On the one hand, a larger τ expands redistribution to the poor. This, in turn, increases the number of children with parents above the consumption cutoff, which tends to make ξ a declining function of τ . On the other hand, a larger τ distorts incentives to work and invest in skills and lowers average earnings and consumption. This force tends to make ξ an increasing function of τ .

Around the empirical tax rate τ^{US} , higher progressivity lowers the share of unskilled workers. When taking into account this effect of τ on ξ , the optimal degree of progressivity increases further to $\tau^* = 0.167$. Since this is almost identical to our empirical estimate for the United States (Section II.), the welfare gain from tax reform is now negligible.

Suppose the government could devise a costless policy reform that would remove all investment constraints and effectively lower \underline{c} to zero. Such a policy would lead the economy to converge to a new steady state with only 12.5% of workers unskilled, 5 percentage points below the current U.S. level. The aggregate welfare gain of implementing such a policy would be 1.6%. The gain is relatively small because the transition to the new steady state takes an entire generation.

VIII. EMPIRICAL EVIDENCE

The aim of this section is to assess whether the key implications of our baseline model for the determinants of optimal progressivity find empirical support in a cross-country panel dataset. Proposition 4 (for the special case $\gamma = \beta$) establishes that optimal progressivity τ^* depends on seven model parameters: $(\sigma, \psi, \chi, \theta, v_\alpha, v_\varphi, v_\varepsilon)$. In what follows, we assume that preference parameters σ and ψ (defining, respectively, curvature over labor supply and over skill investment) are common across countries. We then use available data to construct country- and time-specific estimates of tax progressivity and empirical proxies for the other parameters, and ask whether observed progressivity covaries with these proxies in the same way that optimal theoretical progressivity varies with the corresponding parameters.

For cross-country estimates of progressivity, we use the World Tax Indicator (WTI) database

(Andrew Young School of Policy Studies 2010). This is the most comprehensive and comparable measure of tax progressivity available.⁴¹ For each country-year pair (c, t) , the database reports an index of marginal rate progression (MRP), which is the slope coefficient obtained from regressing marginal tax rates on logged gross income. In our model, the marginal tax rate schedule is given by $MTR(y) = 1 - \lambda(1 - \tau)y^{-\tau}$. Using the approximation $MTR(y) \simeq -\log(1 - MTR(y)) = -\log(\lambda(1 - \tau)) + \tau \log(y)$, it is clear that $MRP \simeq \tau$.

Our proxy for χ is the share of government purchases in output, $g = G/Y$, which equals $\chi/(1 + \chi)$ under the optimal model policy. Data on government's share of GDP come from the Penn World Tables (PWT) database (Feenstra, Inklaar, and Timmer 2015). The model predicts that τ should decline with g .

The remaining parameters $(\theta, v_\alpha, v_\varphi, v_\varepsilon)$ shape income inequality. The model's predictions for the relationship between optimal progressivity and inequality are subtle. Recall that income inequality in our model has two sources: endogenous skill investment and exogenous variation in labor productivity and the taste for work. The first source implies a hump-shaped relation between the optimal τ and θ , as depicted in Figure II. The second source dictates that the optimal τ is increasing in the uninsurable components of earnings inequality, v_α and v_φ , but declining in the insurable component, v_ε . The overall model income distribution is Pareto-lognormal (PLN), where θ is the Pareto-Lorenz coefficient, and $(v_\alpha, v_\varphi, v_\varepsilon)$ determine the variance of the lognormal component, which we denote \bar{v} hereafter.

The most comprehensive cross-country dataset on income inequality is the Standardized World Income Inequality Database (Solt 2016), which contains estimates of Gini coefficients for equalized household market income. The World Wealth and Income Database (Alvaredo et al. 2016) provides direct estimates of the Pareto coefficient θ for a smaller number of countries. Note that the Gini index of a PLN distribution can be expressed in closed form as a function of θ and \bar{v} . Thus, given country-year-specific values for the Gini and for θ , we can recover estimates for \bar{v} .

We choose 1990 as a starting year because many countries have missing data in earlier years and because this start date allows us to include the countries in the Eastern Bloc. The last year

⁴¹The main shortcoming of the WTI data, from the standpoint of estimating the progressivity of the overall tax and transfer system, is that it does not include government transfers. The estimated τ^{US} from the WTI database is around 0.06. This is very similar to the corresponding estimate based on IRS data provided by Guner, Kaygusuz, and Ventura (2014), which similarly does not include transfers (see their Table A5).

in the sample is 2005. Our final dataset, which combines the four databases described above, merged by country-year, comprises 1,585 country-year observations in its largest configuration. The number of countries grows from 65 in 1990 to over 100 in the most recent years. The version of the dataset that also includes estimates for Pareto coefficients has 351 country-year observations and covers 26 countries, most of which are present for all 16 years.

We run two sets of regressions, one on our larger sample of countries and one on the smaller. The general specification is

$$MRP_t^c = D_t + \beta_0 g_t^c + \beta_1' \mathbf{X}_t^c + \varepsilon_t^c,$$

where MRP_t^c (our proxy for τ) is the index of marginal rate progression in country c in year t , D_t are year dummies, g_t^c is government's share of output in country c in year t , and the vector \mathbf{X}_t^c contains the proxies for income inequality that are available in the particular sample. For the larger sample of countries, \mathbf{X}_t^c is simply the income Gini, whereas for the subsample in which estimates for the Pareto coefficient are available, it is the triple $(\theta_t^c, (\theta_t^c)^2, \bar{v}_t^c)$. This second specification is theoretically preferable because it allows us to test whether the distinct theoretical implications of endogenous versus exogenous income inequality for optimal progressivity are supported empirically. In all our regressions, we weight each country-year observation by the square root of GDP for that country-year.⁴² Table V reports our findings.

The regression in column (1) uses the larger sample. The coefficient on government's share of output is negative, as predicted by the model, and the one on inequality positive.⁴³ Both are significant at the 1% level. The regression in column (2) uses the smaller sample and constitutes the purest test of the theory. The coefficient on g_t^c remains negative and significant. The coefficients on θ and θ^2 are strongly significant and support the hump-shaped relation implied by the model. Finally, the coefficient on \bar{v}_t^c is positive and significant, again as predicted by theory (assuming variation in \bar{v}_t^c primarily reflects variation in v_α or v_φ).

In the remaining columns we introduce additional dummies. Columns (3) and (4) add seven regional dummies, and columns (5) and (6) add four additional dummies for the country's level of

⁴²We obtain similar results when weighting by country population or not weighting at all.

⁴³In contrast, Guvenen, Kuruscu, and Ozkan (2014) estimate a negative relation between gross income inequality and progressivity (both measured differently from the indicators we use). They use only a small subset of the countries in our dataset (United States, Denmark, Finland, France, Germany, Netherlands, Sweden, United Kingdom). When we run our regression on this subset of countries, we also find also a negative coefficient, but one that is not statistically significant.

TABLE V: Empirical Determinants of Progressivity across Countries

	(1)	(2)	(3)	(4)	(5)	(6)
	MRP	MRP	MRP	MRP	MRP	MRP
G/Y	-0.0658*** (-5.08)	-0.254*** (-5.82)	-0.0382** (-2.67)	-0.303*** (-6.01)	0.0431*** (3.52)	-0.185*** (-4.07)
Income Gini	0.0746*** (4.95)		0.0992*** (6.89)		0.0738*** (5.42)	
θ		0.139*** (7.89)		0.123*** (7.11)		0.00451 (0.22)
θ^2		-0.0266*** (-8.62)		-0.0225*** (-7.52)		-0.0000972 (-0.02)
\bar{v}		0.0248*** (3.83)		0.00585 (0.67)		0.0219** (3.30)
Regional Dummies	N	N	Y	Y	Y	Y
Development Dummies	N	N	N	N	Y	Y
N	1585	351	1585	351	1585	351
adj. R^2	0.035	0.346	0.280	0.478	0.451	0.629

t statistics in parentheses: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

development (both based on the World Bank classification).⁴⁴ Controlling for regional dummies increases the R^2 of the regressions and leaves all results qualitatively unchanged, except for the significance of the coefficient on \bar{v}_t^c . Adding development dummies changes the sign of the coefficient on g_t^c in column (5), but the sign is restored in the richer specification with the smaller sample in column (6). In this last regression, the coefficients on the Pareto tail lose significance, but the one on \bar{v}_t^c remains negative and significant.

Overall, this first pass at the data indicates that our theory identifies qualitatively relevant determinants of tax progressivity. We now ask whether the estimated coefficients from regressing observed progressivity on government purchases and income inequality are *quantitatively* consistent with the theory. The true theoretical relationship is nonlinear, but we can compute the local sensitivity of progressivity to these determinants for perturbations around our baseline parameterization to the United States.

To do so, we start with equation (30), take the first-order condition with respect to τ to derive

⁴⁴The case for adding these extra dummies is that they might capture actual determinants of progressivity that are not included in our theory. The case against including them is that if the theoretical determinants of progressivity are empirically relevant but noisily measured, then adding dummies that are correlated with these noisy variables could artificially reduce their estimated statistical significance.

an expression that implicitly defines the welfare-maximizing value for τ , and then use the implicit function theorem to compute the sensitivity of optimal progressivity to various structural parameters. We use the solution for the optimal size of government $g = \chi/(1+\chi)$ to translate progressivity sensitivity with respect to τ to predicted sensitivity with respect to g . For predicted sensitivity with respect to the variance of the normal component of income (\bar{v} in our empirical specification), we fix the ratios v_α/\bar{v} , v_φ/\bar{v} , and v_ε/\bar{v} at their estimated values for the United States. The theory-predicted value for $d\tau/dg$ is -0.690 compared to the regression coefficient of -0.254 reported in column (2) of Table V. The theory-predicted value for $d\tau/d\bar{v}$ is 0.175 compared to the empirical value of 0.025 . Thus, while the empirical regression coefficients have the right sign, they are smaller in magnitude than predicted by the model. Note, however, that to the extent that \bar{v} and g are measured with error in our datasets, the empirical coefficients will be biased toward zero.⁴⁵ Online Appendix C provides more details on the datasets used in our regressions, additional sensitivity analysis, and closed-form derivations for the theoretical counterparts to the empirical regression coefficients.

IX. CONCLUSIONS

This paper develops an equilibrium framework to study the optimal degree of progressivity of the tax system. The framework restricts the policy space to a particular functional form for the tax and transfer schedule with two salient features: (i) the model is fully tractable, and (ii) the functional form offers a good fit to the current U.S. system. Our main result is an expression for social welfare as a function of (i) policy parameters defining the degree of tax progressivity and the size of government and (ii) structural parameters defining preferences, technology, and households' access to private insurance. These parameters regulate the relative strengths of the economic forces pushing for and against progressivity. A utilitarian planner wants a progressive tax system to redistribute against inequality in initial conditions and to offer social insurance against life cycle productivity shocks that households cannot smooth privately. At the same time, the planner understands that higher progressivity distorts labor supply and skill investment.

When we parameterize the model and quantify the net impact of the various forces for and

⁴⁵It is difficult to summarize the theoretical sensitivity of τ with respect to the Pareto coefficient θ , because this relationship is nonmonotonic (see Figure II). We can note, however, that the regression coefficients in column (2) imply that progressivity in the data tends to be largest for countries with θ around 2.6, whereas given our baseline parameterization, the progressivity-maximizing value in the model is $\theta = 1.5$.

against progressivity, our baseline model suggests that a utilitarian planner could generate welfare gains by making the U.S. tax and transfer system less progressive. This result stands in contrast to most of the existing literature. For example, Saez (2001, Table 2) found optimal marginal tax rates to be around 50%, significantly higher than the current U.S. rates.

We emphasize two forces that limit the optimal degree of progressivity in our model. The first is that progressivity discourages skill investment and thereby reduces pretax wages as well as labor supply. In our baseline calibration, this effect is quantitatively important. Furthermore, even though wages do reflect skill investment choices, tax progressivity is not an effective way to compress inequality in pretax wages because greater relative scarcity of high-skill workers increases the skill premium in general equilibrium.

A second force limiting the optimal degree of progressivity is the need to provide public goods. We jointly analyzed the optimal size of government and the optimal tax/transfer system, and found an important interaction between the two: the more net revenue must be collected to finance public consumption, the less progressive the tax system should be. The logic is that individual households tend to underinvest in skills and work too little because they do not internalize the value of the public goods that additional tax payments can finance. Lower marginal tax rates (achieved through lower progressivity) help narrow the gap between private and social returns.

Although our baseline utilitarian model suggests a case for making the current U.S. tax and transfer system less progressive, several additional factors push in the opposite direction. First, a planner who is more inequality averse than our utilitarian baseline would prefer higher progressivity. Second, progressivity should also be higher if poverty is a barrier to skill investment. We explored an extension in which progressive taxation, by alleviating poverty, mitigates such investment constraints, inducing a larger optimal τ . This extension underlines the importance of the active research agenda on the role of credit constraints in skill investment.

Finally, our cross-country study of the correlates of variation in observed tax progressivity offers some empirical support for the theory's predictions about how progressivity should optimally vary with government purchases and income inequality.

REFERENCES

- Brant Abbott, Giovanni Gallipoli, Costas Meghir, and Giovanni L Violante. Education Policy and Intergenerational Transfers in Equilibrium. NBER Working Paper No. 18782, 2013.
- Facundo Alvaredo, Anthony B Atkinson, Thomas Piketty, Emmanuel Saez, and Gabriel Zucman. The World Wealth and Income Database, 2016. <http://www.wid.world> (accessed: August, 3, 2016).
- Andrew Young School of Policy Studies. Andrew Young School World Tax Indicators, 2010. Atlanta, GA: International Studies Program, <http://aysps.gsu.edu/isp/wti.html> (accessed: August, 11, 2016).
- Anthony B. Atkinson, Thomas Piketty, and Emmanuel Saez. Top Incomes in the Long Run of History. *Journal of Economic Literature*, 49(1):3–71, 2011.
- Orazio Attanasio and Steven J. Davis. Relative Wage Movements and the Distribution of Consumption. *Journal of Political Economy*, 104(6):1227–1262, 1996.
- Ozan Bakis, Baris Kaymak, and Markus Poschke. Transitional Dynamics and the Optimal Progressivity of Income Redistribution. *Review of Economic Dynamics*, 18(3):679–693, 2015.
- Robert J. Barro and Charles J. Redlick. Macroeconomic Effects from Government Purchases and Taxes. *Quarterly Journal of Economics*, 126:51–102, 2011.
- Roland Bénabou. Unequal Societies: Income Distribution and the Social Contract. *American Economic Review*, 90(1):96–129, 2000.
- Roland Bénabou. Tax and Education Policy in a Heterogeneous-Agent Economy: What Levels of Redistribution Maximize Growth and Efficiency? *Econometrica*, 70(2):481–517, 2002.
- Roland Bénabou. Inequality, Technology and the Social Contract. In Philippe Aghion and Steven Durlauf, editors, *Handbook of Economic Growth*, volume 1B, pages 1595–1638. Amsterdam: Elsevier Science, 2005.
- Richard Blundell, Luigi Pistaferri, and Ian Preston. Consumption Inequality and Partial Insurance. *American Economic Review*, 98(5):1887–1921, 2008. doi: DOI:10.1257/aer.98.5.1887. URL <http://www.stanford.edu/pista/aer2008.pdf>.
- Guillermo A. Calvo and Maurice Obstfeld. Optimal Time-Consistent Fiscal Policy with Finite Lifetimes. *Econometrica*, 56(2):411–432, 1988.
- Shu-Hua Chen and Jang-Ting Guo. Progressive Taxation and Macroeconomic (In) stability with Productive Government Spending. *Journal of Economic Dynamics and Control*, 37(5):951–963, 2013.
- John H. Cochrane. A Simple Test of Consumption Insurance. *Journal of Political Economy*, 99(5):957–976, 1991.
- Juan Carlos Conesa and Dirk Krueger. On the Optimal Progressivity of the Income Tax Code. *Journal of Monetary Economics*, 53(7):1425–1450, 2006.
- Jonathan Eaton and Harvey S. Rosen. Optimal Redistributive Taxation and Uncertainty. *Quarterly Journal of Economics*, 95(2):357–364, 1980.

- Emmanuel Farhi and Iván Werning. Insurance and Taxation over the Life Cycle. *Review of Economic Studies*, 80(2):596–635, 2013.
- Daniel Feenberg and Elisabeth Coutts. An Introduction to the TAXSIM Model. *Journal of Policy Analysis and Management*, 12(1):189–194, 1993.
- Robert C Feenstra, Robert Inklaar, and Marcel P Timmer. The Next Generation of the Penn World Table. *The American Economic Review*, 105(10):3150–3182, 2015.
- Martin S. Feldstein. The Effects of Taxation on Risk Taking. *Journal of Political Economy*, 77(5):755–764, 1969.
- Martin S. Feldstein. On the Optimal Progressivity of the Income Tax. *Journal of Public Economics*, 2(4):357–376, 1973.
- Raquel Fernández and Richard Rogerson. Public Education and Income Distribution: A Dynamic Quantitative Evaluation of Education-Finance Reform. *American Economic Review*, 88(4):813–833, 1998.
- Mikhail Golosov, Maxim Troskin, and Aleh Tsyvinski. Redistribution and Social Insurance. *American Economic Review*, 106(2):359–386, 2016.
- Aspen Gorry and Ezra Oberfield. Optimal Taxation Over the Life Cycle. *Review of Economic Dynamics*, 15(4):551–572, 2012.
- Nezih Guner, Remzi Kaygusuz, and Gustavo Ventura. Income Taxation of U.S. Households: Facts and Parametric Estimates. *Review of Economic Dynamics*, 17(4):559–581, 2014.
- Fatih Guvenen, Burhanettin Kuruscu, and Serdar Ozkan. Taxation of Human Capital and Wage Inequality: A Cross-Country Analysis. *Review of Economic Studies*, 81(2):818–850, 2014.
- Jonathan Heathcote and Hitoshi Tsujiyama. Optimal Income Taxation: Mirrlees Meets Ramsey. *Mimeo*, 2016.
- Jonathan Heathcote, Kjetil Storesletten, and Giovanni L. Violante. Insurance and Opportunities: A Welfare Analysis of Labor Market Risk. *Journal of Monetary Economics*, 55(3):501–525, 2008.
- Jonathan Heathcote, Fabrizio Perri, and Giovanni L. Violante. Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States, 1967-2006. *Review of Economic Dynamics*, 13(1):15–51, 2010.
- Jonathan Heathcote, Kjetil Storesletten, and Giovanni L. Violante. Consumption and Labor Supply with Partial Insurance: An Analytical Framework. *American Economic Review*, 104(7):2075–2126, 2014a.
- Jonathan Heathcote, Kjetil Storesletten, and Giovanni L. Violante. Optimal Tax Progressivity: An Analytical Framework. NBER Working Paper No. 19899, 2014b.
- James Heckman, Lance Lochner, and Christopher Taber. Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogeneous Agents. *Review of Economic Dynamics*, 1(1):1–58, 1998.
- Nanok C. Kakwani. Measurement of Tax Progressivity: An International Comparison. *Economic Journal*, 87(345):71–80, 1977.

- Michael P Keane. Labor Supply and Taxes: A Survey. *Journal of Economic Literature*, 49(4):961–1075, 2011.
- Gizem Kosar and Robert A Moffitt. Trends in Cumulative Marginal Tax Rates Facing Low-Income Families, 1997-2007. NBER Working Paper No. 22782, 2016.
- Dirk Krueger and Alexander Ludwig. Optimal Progressive Labor Income Taxation and Education Subsidies When Education Decisions and Intergenerational Transfers Are Endogenous. *American Economic Review*, 103(3):496–501, 2013.
- Lance Lochner and Alexander Monge-Naranjo. Credit Constraints in Education. *Annual Review of Economics*, 4(1):225–256, 2012.
- Costas Meghir and Luigi Pistaferri. Earnings, Consumption and LifeCycle Choices. In Orley C. Ashenfelter and David E. Card, editors, *Handbook of Labor Economics*, volume 4B of *Handbooks of Labor Economics*. Elsevier, 2010.
- James A. Mirrlees. An Exploration in the Theory of Optimum Income Taxation. *Review of Economic Studies*, 38(2):175–208, 1971.
- Richard A. Musgrave. *The Theory of Public Finance: A Study in Public Economy*. New York: McGraw-Hill, 1959.
- Mats Persson. The Distribution of Abilities and the Progressive Income Tax. *Journal of Public Economics*, 22(1):73–88, October 1983.
- Thomas Piketty and Emmanuel Saez. Income Inequality in the United States, 1913-1998. *The Quarterly Journal of Economics*, 118(1):1–39, 2003.
- Frank P. Ramsey. A Contribution to the Theory of Taxation. *Economic Journal*, 37(145):47–61, 1927.
- Casey Rothschild and Florian Scheuer. Redistributive Taxation in the Roy Model. *Quarterly Journal of Economics*, 128(2):623–668, 2013.
- Emmanuel Saez. Using Elasticities to Derive Optimal Income Tax Rates. *Review of Economic Studies*, 68(1):205–229, 2001.
- Paul A. Samuelson. The Pure Theory of Public Expenditure. *Review of Economics and Statistics*, 36(4):387–389, 1954.
- Florian Scheuer and Iván Werning. The Taxation of Superstars. NBER Working Paper No. 21323, 2015.
- Ananth Seshadri and Kazuhiro Yuki. Equity and Efficiency Effects Of Redistributive Policies. *Journal of Monetary Economics*, 51(7):14151447, 2004.
- Frederick Solt. The Standardized World Income Inequality Database. *Social Science Quarterly*, 97, 2016. SWIID Version 5.0, October 2014.
- Stefanie Stantcheva. Optimal Taxation and Human Capital Policies over the Life Cycle. *Mimeo*, 2015.
- Joseph E. Stiglitz. Equilibrium Wage Distribution. *Economic Journal*, 95(379):595–618, September 1985.

Alexis A Toda. A Note on the Size Distribution of Consumption: More Double Pareto than Lognormal. *Macroeconomic Dynamics*, Forthcoming, 2015.

Hal R. Varian. Redistributive Taxation as Social Insurance. *Journal of Public Economics*, 14(1):49–68, 1980.

RESEARCH DEPARTMENT, FEDERAL RESERVE BANK OF MINNEAPOLIS, AND CEPR

ECONOMICS DEPARTMENT, UNIVERSITY OF OSLO, AND CEPR

ECONOMICS DEPARTMENT, NEW YORK UNIVERSITY, AND CEPR AND NBER

ONLINE APPENDICES

Contents

AA. Empirical measurement of progressivity	2
B.1 Proof of Proposition 1 [hours and consumption]	7
B.2 Proof of Proposition 2 [skill price and skill choice]	8
B.3 Proof of Corollary 2.1 [distribution of skill prices]	9
B.4 Proof of Corollary 2.2 [aggregate quantities]	9
B.5 Proof of Proposition 3 [efficiency with $\chi = v_\omega = \tau = 0$]	10
B.6 Proof of Proposition 4 [closed-form social welfare]	14
B.7 Proofs of Corollaries 4.1, 4.2, 4.3, 4.4, 4.5	19
B.8 Proof of Corollary 4.6 [$\gamma = \beta$ case]	19
B.9 Proof of Proposition 5 [efficiency in RA model]	20
B.10 Derivation of welfare cost of between-skill consumption dispersion	21
B.11 Proof of Proposition 6 [condition for optimal progressivity]	22
B.12 Proof of Proposition 7 [inequality aversion]	22
B.13 Proof of Propositions 8-9 [median voter]	24
B.14 Skill investment constraints	26
B.14i. Consumption for skilled and unskilled	26
B.14ii. Calculating $\lambda(\xi)$	27
B.14iii. Social welfare in the presence of investment constraints	29
C.1 Cross-country dataset construction	32
C.2 Sensitivity	34
C.3 Theoretical counterparts to the empirical regression coefficients	34
D.1 CES skill aggregator in our model versus Benabou (2005)	37
D.2 Comparison of actual and optimal tax systems	39
D.3 Comparison between fixed and flexible skill investment models	40

APPENDIX A

AA. Empirical measurement of progressivity

Statutory progressivity. It is statutory marginal tax rates that are the relevant tax rates in determining household's labor supply and skill investment choices on the margin. This follows from a standard envelope argument: if an individual is already optimizing on the margin in terms of spending on tax-deductible items or in terms of tax avoidance activities, then the marginal value of working a little harder can be computed holding deductions fixed, i.e., using the statutory marginal rate. We therefore want to estimate the progressivity of statutory tax rates. We begin by proving our claim of Section II. that τ does measure the progressivity of statutory (as opposed to effective) tax rates when using taxable income (i.e., gross income y minus deductible expenses x) in the estimation. Tax liabilities are then given by the generalized formula:

$$(A1) \quad T(y, x) = (y - x) - \lambda(y - x)^{1-\tau},$$

where $y - x$ is taxable income, which implies the log-linear relation between post-government income and pre-government taxable income

$$(A2) \quad \log(y - x - T(y, x)) = \log \lambda + (1 - \tau) \log(y - x)$$

that we use to estimate the parameter τ , as discussed in detail in Section II. of the paper. The statutory marginal tax rate implied by (A1) is:

$$(A3) \quad MTR^s \equiv \frac{\partial T(y, x)}{\partial y} = 1 - \lambda(1 - \tau)(y - x)^{-\tau},$$

where note that, by definition, x is kept fixed as y varies marginally. Note, instead, that the effective marginal tax rate MTR^e requires computing the total derivative, or:

$$(A4) \quad MTR^e \equiv \frac{dT(y, x)}{dy} = \frac{\partial T(y, x)}{\partial y} + \frac{\partial T(y, x)}{\partial x} \frac{dx}{dy} = [1 - \lambda(1 - \tau)(y - x)^{-\tau}] \left(1 - \frac{dx}{dy}\right) \leq MTR^s,$$

which is less than or equal to the statutory marginal tax rate MTR^s , as long as x is a normal good. Intuitively, as income grows, so do non-taxable expenditures and thus taxable income rises less

than one for one. The average tax rate, expressed as a share of taxable income, is

$$(A5) \quad ATR \equiv \frac{T(y, x)}{y - x} = 1 - \lambda(y - x)^{-\tau}.$$

A natural measure of statutory progressivity is one minus the coefficient of residual income progression: $1 - (1 - MTR^s)/(1 - ATR)$. By combining (A3) and (A5) it is easy to see that this expression is exactly equal to τ , and thus our empirical approach to estimating progressivity captures the progressivity of statutory tax rates, as desired.

Average marginal tax rates. We now show that when assuming a balanced budget, the average income-weighted marginal tax rate is given by eq. (4). To see this, note that budget balance requires $gY = \int y_i - \lambda y_i^{1-\tau} di$. The income-weighted average marginal tax rate is then given by

$$\int [1 - \lambda(1 - \tau) y_i^{-\tau}] \left(\frac{y_i}{Y}\right) di = 1 - (1 - \tau) \int \lambda y_i^{1-\tau} (1/Y) di = 1 - (1 - \tau)(1 - g).$$

Measurement of tax deductions. As discussed in the main text, the source of our data used in Section II. is the sample from the Panel Study of Income Dynamics (PSID) for survey years 2000, 2002, 2004, and 2006 (when the PSID is biannual) described in detail in Heathcote, Perri and Violante (2012). From this data set, we are able to construct, for every household, measures of income and government transfers defined in the main text. Our aim is to obtain an accurate estimate of tax liabilities for household in the sample by running the NBER's TAXSIM program (Feenberg and Coutts, 1993) on every record in the data. Most of the fields required by TAXSIM (marital status, number of children, various types of income and government transfers, see the description of the software at <http://users.nber.org/~taxsim/taxsim9/>) are already available in our original sample. What requires more work is constructing, for each household, an estimate of their tax deductions in order to arrive at a reliable measure of taxable income. We begin by calculating the four major sources of itemized deductions in the US tax code: medical expenses, mortgage interest, state taxes paid, and charitable contributions. The PSID contains three comprehensive questions on out-of-pocket medical expenditures for, respectively, (i) nursing home and hospital bills, (ii) doctor, outpatient surgery, and dental bills, and (iii) prescriptions, in-home medical care, special facilities, and other medical services. Overall these three questions cover a very broad range of medical expenses.⁴⁶ Median (mean) household expenditures across ages 25-60,

⁴⁶More precisely for each of survey years 2001-03-05-07, the questions we extracted are: 2001: ER19842,

the age range of our sample, are \$750 (\$1,970), in line with existing estimates from other sources (see, e.g., Jung and Tran, 2013, for evidence based on the Medical Expenditure Panel Survey). We follow the instructions on the NBER TAXSIM website on how to calculate the deductible amount for medical expenses.⁴⁷ The PSID contains a question on the residual mortgage debt on the main residence for homeowners.⁴⁸ As per the US tax code, we cap this amount at \$1,000,000. To calculate interest payments, we multiply that amount by 7 percent, the average 30-year conventional annual mortgage rate over this period (FRED series MORTG). As for state taxes, our data set contains the household state of residence that TAXSIM uses to compute state taxes paid by households in calculating the itemized deduction (we used the Internal Revenue Service Statistics of Income (SOI) state codes, as indicated by the instructions for field 2). Finally, since the PSID has no information on charitable contributions, we use a simple imputation procedure. Based on the SOI data, in 2004 charitable contributions amounted to roughly 3% of income, for taxpayers with income greater than \$75,000, where the vast majority of tax returns claim itemized deductions (SOI Table 2.1).⁴⁹ List (2011, Table 2) reports, from survey data across all income levels, that over 2/3 of households donate to charity, and those who do contribute about 4 percent of their income. Thus a charitable contribution rate of 3 percent of household income, for all households in the sample, seems an appropriate rule of thumb for our imputation, and this is what we assume.⁵⁰ Adding up these four components we obtain a measure of itemized deduction. TAXSIM then calculates whether each household is better off taking an itemized deduction or the standard deduction.⁵¹ After running TAXSIM on the data, 49 percent of households in our sample would have taken the itemized deduction (96 percent of those with income above \$100,000). According to SOI Table 2.1, in 2004, around 40 percent of tax returns claimed the itemized deduction (97 percent of those with income above \$100,000), which confirms that our imputation procedure, combined with the

ER19848, ER19854. 2003: ER23279, ER23285, ER23291. 2005: ER27240, ER27246, ER27252. 2007: ER40415, ER40421, ER40447. Since the question asks to report expenditures incurred in the past two years, we divide the household's responses by a factor of two to arrive at annual estimates.

⁴⁷See http://users.nber.org/~taxsim/taxsim9/medical_deduction.html

⁴⁸For survey years 2001-03-05-07, the question is, respectively: ER17052, ER21051, ER25042, ER36042.

⁴⁹The exception is taxpayers with income above \$10M, whose charitable contributions exceed 6 percent of income. However, our PSID sample has no observation in that income range.

⁵⁰In particular, these survey data show no sizable differences across income levels. For example, List reports that for income levels between \$20,000-40,000, the fraction of households who donate is 0.58 with a contribution rate of 5 percent of household income. Thus, even for lower income levels, a 3 percent average contribution rate seems correct.

⁵¹The standard deduction varies by marital status. For example, during 2000-2006, for singles it grows steadily from \$4,400 to \$5,150, and for married couples from \$7,350 to \$10,300.

TAXSIM program, is quite accurate.

Imputation of social security benefits. As explained in the main text, to arrive at a comprehensive measure of gross income, we augment income reported in the PSID for each working household member with the employer’s share (50%) of the Federal Insurance Contribution Act (FICA) tax, which comprises the Old Age, Survivor and Disability Insurance (OASDI) tax (or Social Security tax), and the Hospital Insurance (HI) tax (Medicare tax). Incidentally, the Congressional Budget Office (CBO) makes this same adjustment when computing marginal tax rates. For the years 2000-2006, the OASDI tax rate was 12.4% (and the employer’s share half of that) applicable up to an earnings ceiling which varied by year. For details, see the tables at <https://www.ssa.gov/oact/COLA/cbb.html#Series>). The HI tax rate was 2.9% (again, the employer’s share being half) and there was no limit to the earnings subject to this tax. To derive our measure of post-government income, we subtract from gross income the entire FICA tax liability. For consistency, we need to make an imputation for the corresponding gain in social security benefits accruing to the household member because of the additional year of work (as explained in the text we make no analogous adjustment for Medicare benefits because Medicare eligibility is based on age rather than tied to lifetime earnings). We compute the present value Δ_{ij} of the extra social security benefits that individual i will receive by working at age j relative to the counterfactual where i does not work in that year, but where her past and the future earnings are unchanged. We then add Δ_{ij} to government transfers for that individual-year observation.

This calculation is implemented as follows. For every individual in the sample, we compute an age-earnings profile conditional on gender g and education e (less than high-school, high-school degree, and college degree) using a cubic polynomial in age j . Call these functions $h_{ge}(j)$. Since we observe income y_{ij} for every individual i only once at age j , our best estimate of age j^* earnings for this individual is

$$\hat{y}_{ij^*} = \frac{h_{ge}(j^*)}{h_{ge}(j)} y_{ij}.$$

Let \bar{Y}_i be the Average Index of Monthly Earnings (AIME) under the assumption the individual works from age 0 until retirement age $J^{ret} = 35$ (36 years, as in our sample), i.e.,

$$\bar{Y}_i = \frac{1}{12} \cdot \left(\frac{1}{J^{ret}} \sum_{m=0}^{J^{ret}} \hat{y}_{im} \right)$$

and let

$$\bar{Y}_{ij}^- = \bar{Y}_i - \frac{1}{12} \cdot \frac{y_{ij}}{J^{ret}}$$

be the counterfactual AIME in the absence of age j earnings. The implied annualized social security benefit gain from working at age j is:

$$\pi_{ij} = 12 \cdot [P(\bar{Y}_i) - P(\bar{Y}_{ij}^-)],$$

where P is the formula that determines monthly benefits as a function of AIME, see the explanation at <https://www.ssa.gov/oact/COLA/piaformula.html>.

The marginal gain to be imputed to post-government earnings is the present value of this pension gain discounted back to age j accounting for the fact that this additional pension income is paid in every year following retirement age, conditional on survival:

$$\Delta_{ij} = \left(\frac{1}{R^{J^{ret}-j}} \right) \pi_{ij} \sum_{m=J^{ret}}^J \frac{s_{j,m}}{R^{m-J^{ret}}},$$

where $s_{j,m}$ is the probability of surviving from age j to age m (i.e., the probability of collecting at age m), which is gender-specific and computed based on U.S. Life Tables). In this calculation, we assume that $J = 100$, so $s_{j,J} = 0$, and $R = 1.04$ annually.

APPENDIX B

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

B.1 Proof of Proposition 1 [hours and consumption]

We follow the proof in Heathcote, Storesletten, and Violante (2014), simplified by the absence of risk-free bonds in the economy. Since the only securities that are traded are insurance claims against ε shocks, without loss of generality we can think of our economy as an island economy where each island is populated by agents indexed by their fixed individual characteristic (φ, s) and their uninsurable wage component α . On each island, there are complete markets with respect to ε , so the competitive equilibrium allocation can be computed as the outcome of an island-specific planner problem. Since agents on an island are ex ante identical, the planner weights must be equal across agents. Moreover, since each island transfers zero net financial wealth between periods (by assumption) and preferences are time separable, the island-specific planner problem is static. The island social planner's problem, taking the aggregate fiscal variables (G, λ, τ) and the skill price $p(s)$ as given, is

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

subject to the resource constraint

$$(B1) \quad \int_E c(\varepsilon) dF_\varepsilon = \lambda \int_E \exp[(1-\tau)(p(s) + \alpha + \varepsilon)] h(\varepsilon)^{1-\tau} dF_\varepsilon.$$

It is immediate to see that all agents on each island consume the same amount. Exploiting this perfect risk-sharing outcome, and substituting the resource constraint into the intratemporal first-order condition (FOC), one obtains hours worked by individuals with shock ε on each island

$$(B2) \quad \log h(\varepsilon) = \frac{1}{(1+\sigma)} \log(1-\tau) - \varphi + \frac{1-\tau}{\sigma+\tau} \varepsilon - \frac{1}{\sigma+\tau} \left[\frac{(1-\tau)(1-2\tau-\sigma\tau)}{\sigma+\tau} \frac{v_\varepsilon}{2} \right],$$

where the first term is hours worked by the representative agent, and the term in the square bracket is the constant $\mathcal{M}(v_\varepsilon; \tau)$. Substituting (B2) into (B1) one obtains

$$(B3) \quad \log c = \log \lambda + \frac{1-\tau}{1+\sigma} \log(1-\tau) + (1-\tau) [\log p(s) + \alpha - \varphi] + \mathcal{M}(v_\varepsilon; \tau).$$

In the proof of Proposition 4, we solve for λ as a function of (τ, g) and other structural parameters using the government budget constraint. Note that in eq. 21,

$$\vartheta(\tau) = \log \lambda - \left(\log(1 - g) + \frac{\tau}{1 + \sigma} \log(1 - \tau) \right).$$

B.2 Proof of Proposition 2 [skill price and skill choice]

The education cost is given by $v(s) = \frac{\kappa^{-1/\psi}}{1+1/\psi} (s)^{1+1/\psi}$, where κ is exponentially distributed, $\kappa \sim \eta \exp(-\eta\kappa)$. Recall from eq. (14) in the main text that the optimality condition for skill investment is

$$(B4) \quad v'(s) = \left(\frac{s}{\kappa} \right)^{\frac{1}{\psi}} = (1 - \beta\delta) \mathbb{E}_0 \sum_{a=0}^{\infty} (\beta\delta)^a \frac{\partial u(c(\varphi, \alpha, s; g, \tau), h(\varphi, \varepsilon; \tau), g)}{\partial s}.$$

The skill level s affects only the consumption allocation (not the hours allocation) and only through the price $p(s; \tau)$, which is fixed over time. Hence, using (B3), (B4) can be simplified as

$$\left(\frac{s}{\kappa} \right)^{\frac{1}{\psi}} = (1 - \tau) \frac{\partial \log p(s; \tau)}{\partial s}.$$

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

$$(B5) \quad \log p(s; \tau) = \pi_0(\tau) + \pi_1(\tau) \cdot s,$$

which implies that the skill allocation has the form

$$(B6) \quad s(\kappa; \tau) = [(1 - \tau) \pi_1(\tau)]^{\psi} \cdot \kappa.$$

Since the exponential distribution is closed under scaling, skills inherit the exponential density shape from κ , with parameter $\zeta \equiv \eta [(1 - \tau) \pi_1(\tau)]^{-\psi}$, and its density is $m(s) = \zeta \exp(-\zeta s)$. We now turn to the production side of the economy. Effective hours worked N are independent of skill type s (see Proposition 1). Aggregate output is therefore

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

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

$$(B7) \quad \begin{aligned} \log p(s) &= \log \left[\frac{\partial Y}{\partial [N \cdot m(s)]} \right] = \frac{1}{\theta} \log Y - \frac{1}{\theta} \log [N \cdot m(s)] \\ &= \frac{1}{\theta} \log \left(\frac{Y}{N} \right) - \frac{1}{\theta} \log \zeta + \frac{\zeta}{\theta} s. \end{aligned}$$

Equating coefficients across equations (B5) and (B7) implies $\pi_1(\tau) = \frac{\zeta}{\theta} = \frac{1}{\theta} \frac{\eta}{[(1-\tau)\pi_1(\tau)]^\psi}$, which yields

$$(B8) \quad \pi_1(\tau) = \left(\frac{\eta}{\theta}\right)^{\frac{1}{1+\psi}} (1-\tau)^{-\frac{\psi}{1+\psi}}$$

and thus

$$(B9) \quad m(s) = (\eta)^{\frac{1}{1+\psi}} \left(\frac{\theta}{1-\tau}\right)^{\frac{\psi}{1+\psi}} \exp\left(-(\eta)^{\frac{1}{1+\psi}} \left(\frac{\theta}{1-\tau}\right)^{\frac{\psi}{1+\psi}} s\right).$$

Similarly, the base skill price is

$$(B10) \quad \pi_0(\tau) = \frac{1}{\theta} \log\left(\frac{Y}{N}\right) - \frac{\log\left(\frac{\eta}{\theta}\right)}{\theta(1+\psi)} + \frac{\psi}{\theta(1+\psi)} \log(1-\tau).$$

We derive a fully structural expression for $\pi_0(\tau)$ below in the proof of Corollary 2.2 when we solve for Y and N explicitly.

B.3 Proof of Corollary 2.1 [distribution of skill prices]

The log of the skill premium for an agent with ability κ is

$$\pi_1(\tau) \cdot s(\kappa; \tau) = \pi_1(\tau) \cdot [(1-\tau)\pi_1(\tau)]^\psi \cdot \kappa = \frac{\eta}{\theta} \cdot \kappa$$

where the first equality uses (B6), and the second equality follows from (B8). Thus, log skill premia are exponentially distributed with parameter θ . The variance of log skill prices is

$$var(\log p(s; \tau)) = var(\pi_0(\tau) + \pi_1(\tau) \cdot s(\kappa; \tau)) = \left(\frac{\eta}{\theta}\right)^2 var(\kappa) = \frac{1}{\theta^2}.$$

Since log skill premia are exponentially distributed, the distribution of skill prices in levels is Pareto. The scale (lower bound) parameter is $\exp(\pi_0(\tau))$ and the Pareto parameter is θ .

B.4 Proof of Corollary 2.2 [aggregate quantities]

Aggregate hours and aggregate effective hours are given, respectively, by

$$H(\tau) = \int \int h(\varphi, \varepsilon; \tau) dF_\varphi dF_\varepsilon,$$

$$N(s; \tau) = N(\tau) = \int \int \int \exp(\alpha + \varepsilon) h(\varphi, \varepsilon; \tau) dF_\varphi dF_\alpha dF_\varepsilon.$$

Using the expression for individual hours in Proposition 1 and integrating over the normal distributions for φ , α , and ε gives

$$\begin{aligned} H(\tau) &= (1 - \tau)^{\frac{1}{1+\sigma}} \exp\left(\frac{\tau(1 + \hat{\sigma}) - \hat{\sigma} v_\varepsilon}{\hat{\sigma}^2} \frac{1}{2}\right) \\ N(\tau) &= (1 - \tau)^{\frac{1}{1+\sigma}} \exp\left(\frac{\tau(1 + \hat{\sigma}) + \hat{\sigma} v_\varepsilon}{\hat{\sigma}^2} \frac{1}{2}\right) = H(\tau) \cdot \exp\left(\frac{1}{\hat{\sigma}} v_\varepsilon\right). \end{aligned}$$

Aggregate output is equal to aggregate labor earnings

$$\begin{aligned} \text{(B11)} \quad Y(\tau) &= \int \int \int \int p(s; \tau) \exp(\alpha + \varepsilon) h(\varphi, \varepsilon; \tau) dF_s dF_\varphi dF_\alpha dF_\varepsilon \\ &= \int p(s; \tau) dF_s \cdot N(\tau) \\ &= \frac{\theta}{\theta - 1} \cdot \exp(\pi_0(\tau)) \cdot N(\tau), \end{aligned}$$

where the last line follows from the fact that skill prices are Pareto distributed with scale $\exp(\pi_0(\tau))$ and Pareto parameter θ . Aggregate labor productivity is

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

Finally, one can solve for the base log skill price $\pi_0(\tau)$. From the production function, eq. (5) in the main text, we have that

$$\begin{aligned} Y &= \left\{ \int_0^\infty [N(\tau) \cdot \zeta \exp(-\zeta s)]^{\frac{\theta-1}{\theta}} ds \right\}^{\frac{\theta}{\theta-1}} \\ \text{(B12)} \quad &= N(\tau) \cdot \left(\frac{\theta}{\theta-1}\right)^{\frac{\theta}{\theta-1}} (\eta)^{-\frac{1}{(\theta-1)(1+\psi)}} \left(\frac{1-\tau}{\theta}\right)^{\frac{\psi}{(\theta-1)(1+\psi)}}. \end{aligned}$$

Comparing this equation to eq. (B11) it is immediate that

$$\pi_0(\tau) = \frac{1}{(\theta-1)(1+\psi)} \left[\psi \log\left(\frac{1-\tau}{\theta}\right) - \log(\eta) \right] + \frac{1}{\theta-1} \log\left(\frac{\theta}{\theta-1}\right),$$

which is the expression reported in Proposition 2.

B.5 Proof of Proposition 3 [efficiency with $\chi = v_\omega = \tau = 0$]

With $\chi = 0$ there is no desire for the publicly provided good, and thus the absence of a private market for this good is irrelevant. With $v_\omega = 0$ the absence of private markets for insuring shocks to α is similarly irrelevant: such shocks are simply assumed away. Recall that there are competitive

markets for consumption, for the labor supply of each skill type, and competitive insurance markets for shocks to ε . Thus, given $\chi = 0$ and $v_\omega = 0$ and absent government intervention (i.e., with $\tau = 0$ and $\lambda = 1$), the first welfare theorem applies and competitive equilibrium allocations are Pareto efficient and correspond to the solution to a planner's problem. We now derive the Pareto weights such that the solution to the planner's problem corresponds to the competitive equilibrium allocations. Here, we take as given a result that we'll formally prove in Corollary 4.6: when the weights the planner puts on future generations equals the agent's discount factor β , then social welfare is equal to average period utility in cross section. Moreover, absent uninsurable life-cycle shocks, average period utility is independent of age. Thus, the planner chooses allocations $c(\varphi, \kappa, \varepsilon)$, $h(\varphi, \kappa, \varepsilon)$, and $s(\varphi, \kappa, \varepsilon)$ to solve

$$\max \int \int \int \zeta(\varphi, \kappa, \varepsilon) \left\{ \log c(\varphi, \kappa, \varepsilon) - \frac{\exp[(1+\sigma)\varphi]}{1+\sigma} h(\varphi, \kappa, \varepsilon)^{1+\sigma} - \frac{(\kappa)^{-1/\psi}}{1+1/\psi} (s(\varphi, \kappa, \varepsilon))^{1+1/\psi} \right\} dF_\kappa dF_\varphi dF_\varepsilon$$

subject to

$$\int \int \int c(\varphi, \kappa, \varepsilon) F_\kappa dF_\varphi dF_\varepsilon = \left[\int_0^\infty \mathbb{E}(z)^{\frac{\theta-1}{\theta}} dz \right]^{\frac{\theta}{\theta-1}},$$

where effective hours by skill type z is given by

$$\mathbb{E}(z) = \int \int \int h(\varphi, \kappa, \varepsilon) \exp(\varepsilon) I_{\{s(\varphi, \kappa, \varepsilon)=z\}} dF_\kappa dF_\varphi dF_\varepsilon.$$

The first-order condition with respect to consumption is

$$\frac{\zeta(\varphi, \kappa, \varepsilon)}{c(\varphi, \kappa, \varepsilon)} = \mu,$$

where μ is the multiplier on the resource constraint. The competitive equilibrium consumption allocation is given by

$$\begin{aligned} \log c(\varphi, \kappa, \varepsilon) &= \log \lambda + \pi_0(\tau = 0) + \pi_1(\tau = 0) \cdot s(\kappa; \tau = 0) - \varphi + \frac{1}{\sigma} \frac{v_\varepsilon}{2} \\ &= \log \lambda + \pi_0(\tau = 0) + \frac{\eta}{\theta} \kappa - \varphi + \frac{1}{\sigma} \frac{v_\varepsilon}{2}. \end{aligned}$$

It follows immediately that the Pareto weights must take the form

$$\log \zeta(\varphi, \kappa) = \frac{\eta}{\theta} \kappa - \varphi + \varpi,$$

where ϖ is a constant. Now the average Pareto weight must equal one:

$$\int \int \zeta(\varphi, \kappa) dF_\varphi dF_\kappa = 1$$

Thus,

$$\begin{aligned} \int \int \exp\left(\frac{\eta}{\theta}\kappa - \varphi + \varpi\right) dF_\varphi dF_\kappa &= \exp(\varpi) \int \exp\left(\frac{\eta}{\theta}\kappa\right) dF_\kappa \\ &= \exp(\varpi) \frac{\theta}{(\theta - 1)} \\ &= 1, \end{aligned}$$

which implies $\exp(\varpi) = \frac{\theta-1}{\theta}$ and thus

$$\log \zeta(\varphi, \kappa) = \frac{\eta}{\theta}\kappa - \varphi - \log \frac{\theta}{\theta - 1}.$$

Therefore, we have shown that given the candidate Pareto weights, the planner's consumption allocation aligns with the competitive equilibrium allocation. We now verify that given the same Pareto weights, the equilibrium allocation for skill investment corresponds to the skill investment rule preferred by the planner. To simplify the analysis, we abstract from flexible labor supply and preference heterogeneity, so that agents are heterogeneous only with respect to κ , and the planner's skill investment rule must take the form $s(\kappa)$. Thus,

$$Y = N \cdot \left[\int f_s(z)^{\frac{\theta-1}{\theta}} dz \right]^{\frac{\theta}{\theta-1}}$$

where N is effective hours worked per capita and $f_s(z)$ is the density of skill type z . Let F_s denote the unknown CDF for skills. We know that

$$F_s(s(\kappa)) = F_\kappa(\kappa).$$

By the chain rule

$$f_s(s(\kappa))s'(\kappa) = f_\kappa(\kappa).$$

So

$$\begin{aligned} Y &= N \cdot \left[\int_0^\infty f_s(z)^{\frac{\theta-1}{\theta}} dz \right]^{\frac{\theta}{\theta-1}} = N \cdot \left[\int_0^\infty \left(\frac{f_\kappa(s^{-1}(z))}{s'(s^{-1}(z)))} \right)^{\frac{\theta-1}{\theta}} dz \right]^{\frac{\theta}{\theta-1}} \\ &= N \cdot \left[\int_0^\infty \left(\frac{f_\kappa(\kappa)}{s'(\kappa)} \right)^{\frac{\theta-1}{\theta}} s'(\kappa) d\kappa \right]^{\frac{\theta}{\theta-1}}, \end{aligned}$$

where the substitutions in the last line use $\kappa = s^{-1}(z)$ and $s'(\kappa)d\kappa = dz$ and also exploit the fact that the limits of integration do not change because $s(0) = 0$ and $s(\infty) = \infty$. Thus, the planner's problem can be formulated in Lagrangian form as follows:

$$\begin{aligned} \max_{\{c(\kappa), s(\kappa), s'(\kappa), \lambda, \xi(\kappa)\}} \Lambda &= \int \zeta(\kappa) \left\{ \log c(\kappa) - \frac{(\kappa)^{-1/\psi}}{1 + 1/\psi} s^{1+1/\psi} \right\} dF_\kappa \\ &+ \lambda \left\{ N \cdot \left[\int_0^\infty f_\kappa(\kappa)^{\frac{\theta-1}{\theta}} s'(\kappa)^{\frac{1}{\theta}} d\kappa \right]^{\frac{\theta}{\theta-1}} - \int c(\kappa) dF_\kappa \right\} \\ &+ \int_0^\infty \xi(\kappa) \left[s(\kappa) - \left(\int_0^\kappa s'(x) dx + s(0) \right) \right] d\kappa, \end{aligned}$$

where the first line is the objective, the second is the resource constraint, and the third is a set of constraints linking skill investment levels and derivatives. We know that $s(0) = 0$. The first-order conditions for $s(\kappa)$ and $s'(\kappa)$ are

$$- \left(\frac{s(\kappa)}{\kappa} \right)^{\frac{1}{\psi}} \zeta(\kappa) f(\kappa) + \xi(\kappa) = 0$$

and

$$\lambda N \frac{\theta}{\theta-1} \left[\int_0^\infty f_\kappa(\kappa)^{\frac{\theta-1}{\theta}} s'(\kappa)^{\frac{1}{\theta}} d\kappa \right]^{\frac{1}{\theta-1}} f(\kappa)^{\frac{\theta-1}{\theta}} \frac{1}{\theta} s'(\kappa)^{\frac{1-\theta}{\theta}} - \int_\kappa^\infty \xi(x) dx = 0.$$

The last first-order condition can be rewritten as

$$\lambda N^{1-\frac{1}{\theta}} \frac{\theta}{\theta-1} Y^{\frac{1}{\theta}} f_\kappa(\kappa)^{\frac{\theta-1}{\theta}} \frac{1}{\theta} s'(\kappa)^{\frac{1-\theta}{\theta}} = \int_\kappa^\infty \xi(x) dx.$$

Combining the two first-order conditions yields

$$(B13) \quad \frac{1}{\theta-1} \left(\frac{Y s'(\kappa)}{N f_\kappa(\kappa)} \right)^{\frac{1-\theta}{\theta}} = \int_\kappa^\infty \left(\frac{s(\kappa)}{\kappa} \right)^{\frac{1}{\psi}} \zeta(x) f_\kappa(x) dx.$$

Now the planner weights and competitive equilibrium skill investment rule with $\tau = 0$ are

$$\log \zeta(\kappa) = \frac{\eta}{\theta} \kappa - \log(\theta/(\theta-1))$$

$$s(\kappa) = \left(\frac{\eta}{\theta} \right)^{\frac{\psi}{1+\psi}} \cdot \kappa.$$

Substituting these into the first-order condition (B13) gives

$$\frac{1}{\theta-1} \left(\frac{Y \left(\frac{\eta}{\theta} \right)^{\frac{\psi}{1+\psi}}}{N f_\kappa(\kappa)} \right)^{\frac{1-\theta}{\theta}} = \int_\kappa^\infty \left\{ \left(\frac{\left(\frac{\eta}{\theta} \right)^{\frac{\psi}{1+\psi}} x}{x} \right)^{\frac{1}{\psi}} \exp \left(\frac{\eta}{\theta} x - \log \left(\frac{\theta}{\theta-1} \right) \right) \eta \exp(-\eta x) \right\} dx,$$

which simplifies to

$$\frac{Y}{N} = \theta (\theta - 1)^{\frac{\theta}{1-\theta}} \left(\frac{\eta}{\theta}\right)^{-\frac{1}{(\psi+1)(\theta-1)}},$$

which is exactly productivity per efficiency unit of labor supplied in the competitive equilibrium. This can be easily verified by substituting eq. (24) for π_0 into eq. (28) for aggregate output. Thus the planner's first-order condition is satisfied at the equilibrium allocation.

B.6 Proof of Proposition 4 [closed-form social welfare]

We prove this proposition in two steps. First, we show how to derive a closed-form solution for the residual fiscal instrument λ . Second, we substitute the allocations into the social welfare function and show how to obtain eq. (30).

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

$$(B14) \quad \lambda = \frac{(1-g)Y}{\tilde{Y}}.$$

To compute \tilde{Y} , it is useful to aggregate by age group. Let \tilde{Y}^a denote average per capita disposable income for agents of age a :

$$\begin{aligned} \tilde{Y}^a &= \int [y(s, \varphi, \varepsilon, \alpha)]^{1-\tau} m(s) ds dF_\alpha^a dF_\varphi dF_\varepsilon \\ &= \int [h(\varepsilon) \exp(p(s) + \alpha_a + \varepsilon)]^{1-\tau} m(s) ds dF_\alpha^a dF_\varphi dF_\varepsilon. \end{aligned}$$

Substituting in the hours allocation (20) the expression for the skill price (22), the density function $m(s)$ in eq. (B9), and integrating, we arrive at

$$\tilde{Y}^a = \mathcal{K} \times \exp\left(-\tau(1-\tau)\frac{v_\alpha^a}{2}\right),$$

where

$$\begin{aligned} \mathcal{K} &= (1-\tau)^{\frac{1-\tau}{1+\hat{\sigma}}} \exp\left(-\frac{1}{\hat{\sigma}}\mathcal{M}\right) \exp\left(\left(\frac{(1-\tau)(1+\hat{\sigma})}{\hat{\sigma}}\left(\frac{(1-\tau)(1+\hat{\sigma})}{\hat{\sigma}}-1\right)\right)\frac{v_\varepsilon}{2}\right) \\ &\quad \times \exp\left(-\tau(1-\tau)\frac{v_\varphi}{2}\right) \exp[(1-\tau)\pi_0(\tau)] \int \exp[(1-\tau)\pi_1(\tau)s] m(s) ds. \end{aligned}$$

Note that

$$\begin{aligned}
& \int_0^\infty \exp[(1-\tau)\pi_1(\tau)s] m(s) ds \\
&= \frac{\theta}{\theta-1+\tau} \int_0^\infty \frac{\theta-1+\tau}{\theta} \zeta \exp\left(-\frac{\theta-1+\tau}{\theta} \zeta s\right) ds \\
&= \frac{\theta}{\theta-1+\tau},
\end{aligned}$$

and recall that

$$\begin{aligned}
\pi_0(\tau) &= \frac{1}{\theta-1} \left\{ \frac{1}{1+\psi} [\psi \log(1-\tau) - \log \eta - \psi \log \theta] + \log\left(\frac{\theta}{\theta-1}\right) \right\} \\
\mathcal{M} &= \frac{(1-\tau)[1-\tau(1+\hat{\sigma})] v_\varepsilon}{\hat{\sigma} \cdot 2}.
\end{aligned}$$

Now sum across age groups to obtain

$$\begin{aligned}
\tilde{Y} &= (1-\delta) \sum_{a=0}^\infty \delta^a \tilde{Y}^a = \mathcal{K} \times (1-\delta) \sum_{a=0}^\infty \delta^a \exp\left(-\tau(1-\tau) \frac{v_\alpha^a}{2}\right) \\
\text{(B15)} \quad &= \mathcal{K} \times \frac{(1-\delta) \exp\left(-\tau(1-\tau) \frac{v_\alpha^0}{2}\right)}{1-\delta \exp\left(\frac{-\tau(1-\tau)}{2} v_\omega\right)}.
\end{aligned}$$

Substituting (B12) and (B15) into (B14) and simplifying, we arrive at a solution for the equilibrium value of λ which, in logs, is

$$\begin{aligned}
\text{(B16)} \log \lambda &= \log(1-g) + \frac{\tau(1-\tau)}{\sigma+\tau} \left(\frac{1+\sigma}{\sigma+\tau} + 2 + \sigma \right) \frac{v_\varepsilon}{2} + \frac{\tau}{1+\sigma} \log(1-\tau) \\
&\quad + \tau(1-\tau) \frac{v_\varphi}{2} \\
&\quad - \log(1-\delta) + \tau(1-\tau) \frac{v_\alpha^0}{2} + \log \left[1 - \delta \exp\left(\frac{-\tau(1-\tau)}{2} v_\omega\right) \right] \\
&\quad + \left(\frac{\psi}{1+\psi} \right) \frac{\tau}{\theta-1} \log\left(\frac{1-\tau}{\theta}\right) - \left(\frac{1}{1+\psi} \right) \frac{\tau}{\theta-1} \log \eta \\
&\quad + \left(\frac{\theta+\tau-1}{\theta-1} \right) \log\left(\frac{\theta}{\theta-1}\right) + \log\left(\frac{\theta-1+\tau}{\theta}\right).
\end{aligned}$$

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

$$\begin{aligned}
u(c_a, h, G) &= \log \lambda + \frac{1-\tau}{1+\sigma} \log(1-\tau) - (1-\tau) \varphi \\
\text{(B17)} \quad &+ \frac{1-\tau}{(\theta-1)(1+\psi)} \left[\psi \log(1-\tau) + \log\left(\frac{1}{\eta(\theta-1)^{(1+\psi)}}\right) \right] + \mathcal{M} \\
&\quad - \exp\left(-\frac{1+\sigma}{\hat{\sigma}(1-\tau)} \mathcal{M}\right) \exp\left(\frac{1+\sigma}{\hat{\sigma}} \varepsilon\right) \left(\frac{1-\tau}{1+\sigma}\right) + (1-\tau) \kappa \frac{\eta}{\theta} + \chi \log G + (1-\tau) \alpha_a
\end{aligned}$$

The disutility cost from investing in education is

$$(B18) \quad v(s(\kappa)) = -\frac{\kappa^{-1/\psi}}{1+1/\psi} \cdot s(\kappa; \tau)^{1+1/\psi} = -(1-\tau) \kappa \frac{\eta}{(1+1/\psi)\theta}.$$

Average cross-sectional utility (excluding skill investment costs) at age a , which we denote by \bar{u}_a , is

$$\begin{aligned} \bar{u}_a &= \int \int \int \int u(c_a, h, G) dF_\kappa dF_\varepsilon dF_\varphi dF_\alpha^a \\ &= \bar{u} - (1-\tau) \frac{v_\alpha^a}{2} \\ &= \bar{u} - (1-\tau) \frac{av_\omega}{2}, \end{aligned}$$

where

$$\begin{aligned} \bar{u} &= \log \lambda + \frac{1-\tau}{1+\sigma} \log(1-\tau) - (1-\tau) \frac{v_\varphi}{2} + \frac{1-\tau}{(\theta-1)(\psi+1)} \left[\psi \log(1-\tau) + \log \left(\frac{1}{\eta} \frac{\theta}{(\theta-1)^{(1+\psi)}} \right) \right] \\ &\quad \mathcal{M} - \frac{1-\tau}{1+\sigma} + \frac{1-\tau}{\theta} + \chi \log(gY) \end{aligned} \quad (B19)$$

and where the derivation of the expression for \bar{u} exploits the facts that

$$\begin{aligned} \int \exp \left(-\frac{1+\sigma}{\hat{\sigma}(1-\tau)} \mathcal{M} \right) \exp \left(\frac{1+\sigma}{\hat{\sigma}} \varepsilon \right) dF_\varepsilon &= 1, \\ \int (1-\tau) \kappa \frac{\eta}{\theta} dF_\kappa &= \frac{1-\tau}{\theta}. \end{aligned}$$

Substituting in eq. (B19) expression (B16) for λ and (B12) for Y gives

$$\begin{aligned} (B20) \quad \bar{u} &= \log(1-g) + \chi \log g + (1+\chi) \frac{\log(1-\tau)}{(1+\hat{\sigma})(1-\tau)} - \frac{1}{(1+\hat{\sigma})} \\ &\quad + (1+\chi) \left[\frac{1}{\theta-1} \frac{\psi}{1+\psi} \log(1-\tau) + \frac{1}{\theta-1} \frac{1}{1+\psi} \log \left(\frac{1}{\eta \theta^\psi} \left(\frac{\theta}{\theta-1} \right)^{\theta(1+\psi)} \right) \right] \\ &\quad - \left[-\log \left(1 - \left(\frac{1-\tau}{\theta} \right) \right) - \left(\frac{1-\tau}{\theta} \right) \right] - (1-\tau)^2 \frac{v_\varphi}{2} \\ &\quad + \log \left(\frac{1-\delta \exp \left(\frac{-\tau(1-\tau)}{2} v_\omega \right)}{1-\delta} \right) + (1+\chi) \left[\frac{1}{\hat{\sigma}} v_\varepsilon - \sigma \frac{1}{\hat{\sigma}^2} \frac{v_\varepsilon}{2} \right]. \end{aligned}$$

Average skill investment costs for agents born after the tax reform are

$$\bar{v}_Y = \mathbb{E} \left[-(1-\tau) \kappa \frac{\eta}{(1+1/\psi)\theta} \right] = - \left(\frac{\psi}{1+\psi} \right) \frac{1}{\theta} (1-\tau),$$

whereas average net costs for those born prior to the reform are

$$\bar{v}_O = \frac{\psi}{1+\psi} \frac{1-\tau}{\theta} - \frac{\psi}{1+\psi} \frac{1-\tau_{-1}}{\theta} = - \left(\frac{\psi}{1+\psi} \right) \frac{1}{\theta} (\tau - \tau_{-1}).$$

Now we are in a position to add up across cohorts to compute social welfare defined as

$$\begin{aligned} \mathcal{W}_0(g, \tau; \tau_{-1}) &= (1-\gamma) \frac{\gamma - \beta\delta}{\gamma(1-\beta\delta)} \sum_{j=-\infty}^{\infty} \gamma^j U_{j,0}(g, \tau; \tau_{-1}) \\ &= (1-\gamma) \frac{\gamma - \beta\delta}{\gamma(1-\beta\delta)} \sum_{j=-\infty}^{-1} \gamma^j U_j + (1-\gamma) \frac{\gamma - \beta\delta}{\gamma(1-\beta\delta)} \sum_{j=0}^{\infty} \gamma^j U_j, \end{aligned}$$

where the second line partitions the population into cohorts born before and after the tax reform.

Starting with the agents born after the tax reform,

$$\begin{aligned} (1-\gamma) \frac{\gamma - \beta\delta}{\gamma(1-\beta\delta)} \sum_{j=0}^{\infty} \gamma^j U_j &= (1-\gamma) \frac{\gamma - \beta\delta}{\gamma(1-\beta\delta)} \sum_{j=0}^{\infty} \gamma^j \left((1-\beta\delta) \sum_{a=0}^{\infty} (\beta\delta)^a \bar{u}_a - \bar{v}_Y \right) \\ &= \frac{\gamma - \beta\delta}{\gamma} \sum_{a=0}^{\infty} (\beta\delta)^a \bar{u}_a - \frac{\gamma - \beta\delta}{\gamma(1-\beta\delta)} \bar{v}_Y. \end{aligned}$$

Now

$$\bar{u}_a = \bar{u} - (1-\tau) a \frac{v_\omega}{2},$$

so

$$\begin{aligned} \sum_{a=0}^{\infty} (\beta\delta)^a \bar{u}_a &= \frac{\bar{u}}{1-\beta\delta} - (1-\tau) \frac{v_\omega}{2} \{ \beta\delta + 2(\beta\delta)^2 + \dots \} \\ &= \frac{\bar{u}}{1-\beta\delta} - \frac{\beta\delta}{(1-\beta\delta)^2} (1-\tau) \frac{v_\omega}{2} \end{aligned}$$

and thus

$$(1-\gamma) \frac{\gamma - \beta\delta}{\gamma(1-\beta\delta)} \sum_{j=0}^{\infty} \gamma^j U_j = \frac{\gamma - \beta\delta}{\gamma(1-\beta\delta)} \bar{u} - \frac{\gamma - \beta\delta}{\gamma(1-\beta\delta)^2} \beta\delta (1-\tau) \frac{v_\omega}{2} - \frac{\gamma - \beta\delta}{\gamma(1-\beta\delta)} \bar{v}_Y.$$

Now consider agents born before the reform (the youngest of which are age 1 at the time of reform):

$$\begin{aligned} (1-\gamma) \frac{\gamma - \beta\delta}{\gamma(1-\beta\delta)} \sum_{j=-\infty}^{-1} \gamma^j U_j &= (1-\gamma) \frac{\gamma - \beta\delta}{\gamma(1-\beta\delta)} \times \\ &\quad \gamma^{-1} (\beta\delta)^1 \{ (1-\beta\delta) (\bar{u}_1 + (\beta\delta)\bar{u}_2 + (\beta\delta)^2\bar{u}_3 + \dots) - \bar{v}_o \} \\ &\quad + \gamma^{-2} (\beta\delta)^2 \{ (1-\beta\delta) (\bar{u}_2 + (\beta\delta)\bar{u}_3 + \dots) - \bar{v}_o \} + \dots \end{aligned}$$

Adding the pieces here involving \bar{u} and \bar{v}_o gives

$$(1 - \gamma) \frac{\beta\delta}{\gamma(1 - \beta\delta)} (\bar{u} - \bar{v}_o).$$

The term in v_ω is

$$\begin{aligned} & (1 - \gamma) \frac{\gamma - \beta\delta}{\gamma(1 - \beta\delta)} \times \left(- (1 - \tau) \frac{v_\omega}{2} \right) \times (1 - \beta\delta) \times \\ & \left\{ \left(\frac{\beta\delta}{\gamma} \right) (1 + 2(\beta\delta) + 3(\beta\delta)^2 + \dots) + \left(\frac{\beta\delta}{\gamma} \right)^2 (2 + 3(\beta\delta) + 4(\beta\delta)^2 + \dots) + \dots \right\} \\ = & (1 - \gamma) \frac{\gamma - \beta\delta}{\gamma} \times \left(- (1 - \tau) \frac{v_\omega}{2} \right) \times \left\{ \frac{\beta\delta}{\gamma} \frac{1}{(1 - \beta\delta)^2} + \left(\frac{\beta\delta}{\gamma} \right)^2 \left(\frac{1}{1 - \beta\delta} + \frac{1}{(1 - \beta\delta)^2} \right) + \dots \right\} \\ = & (1 - \gamma) \frac{\gamma - \beta\delta}{\gamma} \times \left(- (1 - \tau) \frac{v_\omega}{2} \right) \times \left\{ \frac{\frac{\beta\delta}{\gamma}}{(1 - \beta\delta)(1 - \frac{\beta\delta}{\gamma})} \left[\frac{1}{(1 - \beta\delta)} + \frac{\left(\frac{\beta\delta}{\gamma} \right)}{\left(1 - \frac{\beta\delta}{\gamma} \right)} \right] \right\}. \end{aligned}$$

Now we can add together the contributions to social welfare from agents born before and after the reform. The two terms involving \bar{u} add up exactly to \bar{u} . The terms in \bar{v}_Y and \bar{v}_0 simplify to give $-\left(\frac{\psi}{1+\psi}\right) \frac{1}{\theta} \left[(1 - \tau) + \frac{\beta\delta(1-\gamma)}{\gamma(1-\beta\delta)} (1 - \tau_{-1}) \right]$. The term in v_ω is $-\frac{\beta\delta}{\gamma - \beta\delta} (1 - \tau) \frac{v_\omega}{2}$. Collecting all these terms gives the expression for social welfare in Proposition 4. In particular, collecting the terms in v_ω , we obtain

$$- \left[(1 - \tau) \frac{\beta\delta}{\gamma - \beta\delta} \frac{v_\omega}{2} - \log \left(\frac{1 - \delta \exp \left(\frac{-\tau(1-\tau)}{2} v_\omega \right)}{1 - \delta} \right) \right].$$

When $\gamma = \beta$, the first term in square brackets simplifies to $(1 - \tau) \frac{\delta}{1 - \delta} \frac{v_\omega}{2}$. The second term can be approximated as follows:

$$\log \left(\frac{1 - \delta \exp \left(\frac{-\tau(1-\tau)}{2} v_\omega \right)}{1 - \delta} \right) \approx \log \left(1 + \frac{\delta}{1 - \delta} \tau(1 - \tau) \frac{v_\omega}{2} \right) \approx \tau(1 - \tau) \frac{\delta}{1 - \delta} \frac{v_\omega}{2}.$$

Adding the two pieces, we have

$$(1 - \tau) \frac{\delta}{1 - \delta} \frac{v_\omega}{2} - \tau(1 - \tau) \frac{\delta}{1 - \delta} \frac{v_\omega}{2} = (1 - \tau)^2 \frac{\delta}{1 - \delta} \frac{v_\omega}{2} = (1 - \tau)^2 \frac{v_\alpha}{2},$$

where the last equality reflects the fact that v_α is the cross-sectional variance of the cumulated innovations ω , and the δ in the numerator reflects our assumption that wage shocks start realizing at age $a = 1$. We use this approximate result when we interpret the various components of social

welfare in Section V.B.. Note that this approximation is extremely accurate for plausible parameter values. For example, when evaluated at $\beta = \gamma$, the empirical $\tau^{US} = 0.181$, and the calibrated values $\delta = 0.971$ and $v_\omega = 0.003$, the accurate and the approximated values for the welfare cost if uninsurable life-cycle risk are 0.03371 and 0.03368, respectively.

B.7 Proofs of Corollaries 4.1, 4.2, 4.3, 4.4, 4.5

4.1: In eq. (30), the term in τ_{-1} is additively separable from all the others containing g and τ .

4.2: Differentiating the expression for social welfare twice with respect to τ , it is straightforward to show that each term except the last one involving insurable risk is strictly concave in τ . The term in insurable risk has a second derivative equal to

$$-(1 + \chi) \frac{\sigma - 2\tau}{(\sigma + \tau)^4} (1 + \sigma)^2 v_\varepsilon,$$

which is less than or equal to zero if $\sigma \geq 2$. Thus, $\sigma \geq 2$ is a (weak) sufficient condition for global concavity of social welfare with respect to τ . It is straightforward to verify that the social welfare expression is concave in g . **4.3:** Differentiating eq. (30) with respect to τ , the first-order condition has no terms involving g . Thus, the optimal choice for τ is independent of g . **4.4:** Differentiating eq. (30) with respect to g , the first-order condition is

$$\frac{-1}{1-g} + \frac{\chi}{g} = 0,$$

which immediately gives the expression for g^* in eq. (31). **4.5:** The parameter η only appears in an additively separable constant in eq. (30). Thus, this parameter does not appear in the first-order conditions defining the optimal choices for g and τ .

B.8 Proof of Corollary 4.6 [$\gamma = \beta$ case]

In eq. (29) when $\gamma = \beta$, the constant term Γ simplifies to $\frac{1-\delta}{1-\beta\delta}$. Let $\mathbb{E}[u_0]$ denote expected period utility for newborn agents from consumption and leisure. The contribution to social welfare from newborn agents is then

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

where $(1 + \beta + \beta^2 + \dots)$ reflects the weights the planner puts on current and future cohorts of age zero. Note that $(1 - \delta)$ is the size of the population at age zero. Similarly, the age 1 component is

given by

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

where the term $(1 - \delta) \delta$ is the size of the population at age 1. And so on. Now we need to compute how skill acquisition costs factor into social welfare. Education costs for the newborn and future cohorts are

$$(1 - \beta) \frac{1 - \delta}{1 - \beta\delta} \cdot (1 + \beta + \beta^2 + \dots) \cdot \mathbb{E}[v(s(\kappa, \tau), \kappa)] = \frac{1 - \delta}{1 - \beta\delta} \cdot \mathbb{E}[v(s(\kappa, \tau), \kappa)]$$

If skill accumulation decisions are fully reversible, the net skill investment cost for an agent of type κ given a new progressivity value τ and a past progressivity value τ_{-1} is $v(s(\kappa; \tau), \kappa) - v(s(\kappa; \tau_{-1}), \kappa)$. Thus, with $\gamma = \beta$, the contribution to social welfare from net skill investments from cohorts who entered the economy in the past (ages $a = 1, 2, \dots$) is

$$\begin{aligned} & (1 - \beta) \frac{1 - \delta}{1 - \beta\delta} \cdot \sum_{a=1}^{\infty} \left(\frac{\beta\delta}{\beta} \right)^a \mathbb{E}[v(s(\kappa; \tau), \kappa) - v(s(\kappa; \tau_{-1}), \kappa)] \\ &= \frac{\delta(1 - \beta)}{1 - \beta\delta} \cdot \mathbb{E}[v(s(\kappa; \tau), \kappa) - v(s(\kappa; \tau_{-1}), \kappa)] \end{aligned}$$

Adding the two pieces, and ignoring the term in τ_{-1} , since it is separable from τ and thus irrelevant for optimization, gives

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

Adding up these various welfare components gives the expression for social welfare in eq. (32).

B.9 Proof of Proposition 5 [efficiency in RA model]

We will prove this proposition for general CRRA utility with risk aversion coefficient $\gamma > 0$. The baseline model corresponds to the special case $\gamma = 1$. Absent household heterogeneity, the planner's problem (for a planner with access to lump-sum taxes) is

$$\max_{H, C, G} \left\{ \frac{C^{1-\gamma}}{1-\gamma} - \frac{H^{1+\sigma}}{1+\sigma} + \chi^\gamma \frac{G^{1-\gamma}}{1-\gamma} \right\}$$

such that

$$C + G = H.$$

The first-order conditions give the following solution

$$\begin{aligned} H^* &= (1 + \chi)^{\frac{\gamma}{\gamma + \sigma}} \\ g^* &= \frac{G^*}{H^*} = \frac{\chi}{1 + \chi}. \end{aligned}$$

We now show that a planner with access only to the tax schedule in (1) can replicate the same allocations by setting $\tau = -\chi$ and $\lambda = (1 + \chi)^{-1 - \frac{\chi\gamma}{\gamma + \sigma}}$. To verify this, consider the representative household's problem:

$$\begin{aligned} \max_{C, H} & \left\{ \frac{C^{1-\gamma}}{1-\gamma} - \frac{H^{1+\sigma}}{1+\sigma} \right\} \\ \text{s.t.} & \\ C &= \lambda H^{1-\tau}. \end{aligned}$$

with FOC

$$(\lambda H^{1-\tau})^{-\gamma} \lambda (1 - \tau) H^{-\tau} = H^\sigma.$$

Substituting in the candidate expressions for τ and λ gives

$$H = (1 + \chi)^{\frac{\gamma}{\sigma + \gamma}} = H^*.$$

Substituting this expression for hours plus the candidate expressions for τ and λ into the government budget constraint gives

$$G = H - \lambda H^{1-\tau} = \frac{\chi}{1 + \chi} H^*$$

so that $G/H = G^*/H^*$.

B.10 Derivation of welfare cost of between-skill consumption dispersion

The skill-related component of consumption, $p(s)^{1-\tau}$, is Pareto distributed with parameter $P = \theta/(1 - \tau)$. If consumption is Pareto distributed with Pareto parameter P , the expected value for consumption is $P/(P - 1)$. Log consumption is then exponentially distributed, with exponential parameter P and the expected value for log consumption is $1/P$. Let F_c denote the Pareto cumulative distribution function (CDF) for this consumption, and let F_z denote the Exponential CDF for $z = \log c$. The welfare cost ϖ of consumption dispersion (assuming logarithmic preferences)

can then be calculated as the fraction ϖ by which safe consumption must be reduced to deliver the same expected utility as risky consumption:

$$\begin{aligned}\log\left(\frac{P}{P-1}(1-\varpi)\right) &= \int \log c \, dF_c \\ \log\left(\frac{P}{P-1}(1-\varpi)\right) &= \int z \, dF_z \\ \log\left(\frac{P}{P-1}\right) + \log(1-\varpi) &= \frac{1}{P} \\ \varpi \approx -\log(1-\varpi) &= \log\left(\frac{P}{P-1} - \frac{1}{P}\right)\end{aligned}$$

Since in our economy $P = \theta/(1-\tau)$ and ϖ is small, ϖ is approximately equal to the expression in eq. (36).

B.11 Proof of Proposition 6 [condition for optimal progressivity]

Assume $\gamma = \beta$ and approximate the sixth line of the social welfare expression (eq. 30) by $-(1-\tau)^2 \frac{v_\alpha}{2}$. Then the derivative of the social welfare expression with respect to τ is

$$\frac{\partial \mathcal{W}(g, \tau; \tau_{-1})}{\partial \tau} \Big|_{\tau=0} = - \left(\frac{1}{1+\sigma} + \frac{\psi}{1+\psi} \frac{1}{\theta-1} \right) (1+\chi) - \frac{1}{1+\psi} \frac{1}{\theta} + \frac{1}{1+\sigma} + \frac{1}{\theta-1} + v_\varphi + v_\alpha.$$

It is immediate that this derivative is positive if and only if the condition in Proposition 6 is satisfied.

B.12 Proof of Proposition 7 [inequality aversion]

We begin by computing expected utility (excluding skill investment costs) for an agent with states $(\kappa, \varphi, \alpha)$ prior to ε being drawn. Substituting (B16) into (B17) and taking the expected value of this prior to ε being drawn yields, after some simplifications:

$$\tilde{u}(\kappa, \varphi, \alpha; g, \tau) = \mathcal{C}(g, \tau) + (1-\tau)\alpha - (1-\tau)\varphi + (1-\tau)\frac{\eta}{\theta}\kappa - \frac{1-\tau}{1+\sigma} + \chi G(g, \tau)$$

where $\mathcal{C}(g, \tau)$ captures the components showing up in the log consumption allocation that are common across all individuals

$$\begin{aligned}
\mathcal{C}(g, \tau) &= \log(1 - g) + \frac{(1 - \tau)}{(\sigma + \tau)^2} (\sigma + 2\tau + \sigma\tau) \frac{v_\varepsilon}{2} \\
&\quad + \left(\frac{1}{1 + \sigma} + \frac{\psi}{1 + \psi} \frac{1}{(\theta - 1)} \right) \log(1 - \tau) \\
&\quad + \tau(1 - \tau) \frac{v_\varphi}{2} + \tau(1 - \tau) \frac{v_\alpha^0}{2} + \log \left(\frac{1 - \delta \exp \left(\frac{-\tau(1 - \tau)}{2} v_\omega \right)}{1 - \delta} \right) \\
&\quad - \frac{1}{(\theta - 1)(\psi + 1)} \log(\eta) + \frac{\theta}{(\theta - 1)} \log \left(\frac{\theta}{\theta - 1} \right) - \frac{\psi}{(1 + \psi)(\theta - 1)} \log(\theta) \\
&\quad + \log \left(\frac{\theta - 1 + \tau}{\theta} \right)
\end{aligned}$$

and where

$$\begin{aligned}
G(g, \tau) &= \log g + \log Y(\tau) \\
&= \log g + \left(\frac{\tau(1 + \hat{\sigma}) + \hat{\sigma} v_\varepsilon}{\hat{\sigma}^2} \right) + \log \left(\left(\frac{\theta}{\theta - 1} \right)^{\frac{\theta}{\theta - 1}} \right) + \log \left(\eta^{\frac{1}{1 + \psi} \frac{1}{1 - \theta}} \theta^{\frac{\psi}{\psi + 1} \frac{1}{1 - \theta}} \right) \\
&\quad + \log \left((1 - \tau)^{\frac{1}{1 + \sigma} - \frac{\psi}{\psi + 1} \frac{1}{1 - \theta}} \right).
\end{aligned}$$

Now, let's add the net cost of skill investment:

$$v(\kappa; \tau) = \begin{cases} [(1 - \tau) - (1 - \tau_{-1})] \frac{\psi}{(1 + \psi)} \frac{\eta}{\theta} \kappa & \text{if } a > 0 \\ (1 - \tau) \frac{\psi}{(1 + \psi)} \frac{\eta}{\theta} \kappa & \text{if } a = 0. \end{cases}$$

Next, we compute the value for constant consumption forever that, assuming equilibrium hours and skill investment, gives an agent lifetime utility equal to what would accrue, in expectation, to an agent with type $(\kappa, \varphi, \alpha; g, \tau)$. Define the answer as

$$\bar{c}(\kappa, \varphi, \alpha; g, \tau) = \exp \left(\mathcal{C}(g, \tau) + (1 - \tau) \alpha - (1 - \tau) \varphi + (1 - \tau) \frac{\eta}{\theta} \kappa - (1 - \tau) \frac{v_\omega}{2} \frac{\beta \delta}{1 - \beta \delta} \right).$$

Now suppose that the contribution to social welfare from the cohort of age a at the time of the tax reform is an age-dependent weight times

$$\begin{aligned}
\mathcal{V}_a(g, \tau) &= \log \left(\int \int \int \bar{c}(\kappa, \varphi, \alpha; g, \tau)^{1 - \nu} dF_\kappa dF_\varphi dF_\alpha^a \right)^{\frac{1}{1 - \nu}} - \frac{1 - \tau}{1 + \sigma} + \chi G(g, \tau) \\
&\quad - \frac{\psi}{(1 + \psi) \theta} [(1 - \tau) - \mathbb{I}_{\{a > 0\}} (1 - \tau_{-1})].
\end{aligned}$$

Integrating the first term and rearranging yields:

$$\begin{aligned}\mathcal{V}_a(g, \tau) &= \mathcal{C}(g, \tau) - (1 - \tau) \frac{v_\omega}{2} \frac{\beta\delta}{1 - \beta\delta} + (1 - \tau) ((1 - \tau)(1 - \nu) - 1) \frac{(v_\varphi + av_\omega)}{2} \\ &\quad + \frac{1}{\nu - 1} \log \left((\nu - 1) \frac{1}{\theta} (1 - \tau) + 1 \right) \\ &\quad - \frac{1 - \tau}{1 + \sigma} + \chi G(g, \tau) - \frac{\psi}{(1 + \psi)\theta} [(1 - \tau) - \mathbb{I}_{\{a > 0\}} (1 - \tau_{-1})].\end{aligned}$$

The planner's objective is

$$(B21) \quad \mathcal{W}_0(g, \tau) = (1 - \gamma) \frac{\gamma - \beta\delta}{\gamma(1 - \beta\delta)} \sum_{a=1}^{\infty} \gamma^{-a} (\beta\delta)^a \mathcal{V}_a(g, \tau) + (1 - \gamma) \frac{\gamma - \beta\delta}{\gamma(1 - \beta\delta)} \sum_{j=0}^{\infty} \gamma^j \mathcal{V}_0(g, \tau).$$

Aggregating across cohorts using (B21) with $\gamma = \beta$ yields, after some manipulations, the expression in Proposition 7. To conclude, it is useful to explain why lines (4)-(6) in Proposition 7 measure the cost of consumption inequality reflecting skill (κ) differentials, preference heterogeneity, and past uninsurable shocks. Define ω_κ as the solution to

$$\log \left(\int \exp \left((1 - \tau) \frac{\eta}{\theta} \kappa \right)^{1-\nu} dF_\kappa \right)^{1/(1-\nu)} = \log[(1 + \omega_\kappa)\theta/(\theta - (1 - \tau))]$$

where $\exp[(1 - \tau)\eta/\theta \cdot \kappa]$ is the factor in consumption attributable to κ and $\theta/(\theta - (1 - \tau))$ is the expected value for this factor. Given our exponential distribution for κ , $\omega_\kappa \approx \log(1 - (1 - \tau)/\theta) + 1/(\nu - 1) \cdot \log(1 + (\nu - 1) \cdot (1 - \tau)/\theta)$. Similarly, for the welfare cost of consumption inequality due to heterogeneity in the taste for leisure, line (5), define ω_φ as the solution to

$$\log \left(\int \exp(-(1 - \tau)\varphi)^{1-\nu} dF_\varphi \right)^{1/(1-\nu)} = \log((1 + \omega) \exp((1 - \tau)((1 - \tau) - 1)v_\varphi/2))$$

where $\exp(-(1 - \tau)\varphi)$ is the factor in consumption attributable to φ , and $\exp((1 - \tau)((1 - \tau) - 1)v_\varphi/2)$ is the expected value for this factor. Given our Normal distribution for φ , $\omega_\varphi = -(1 - \tau)^2 \nu \cdot v_\varphi/2$.

Line (6) in the welfare expression in Proposition 7 is obtained analogously.

B.13 Proof of Propositions 8-9 [median voter]

To understand Proposition 8, note that the preferred value for g for agent i obeys the first-order condition

$$(B22) \quad \chi \frac{1}{g} = \frac{c_i}{1 - g} \cdot \frac{1}{c_i}.$$

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 with a marginal increase in g . Since c_i can be expressed as $c_i(g, \tau) = \lambda(g) \bar{c}_i(\tau) = (1 - g) \bar{\Lambda} \bar{c}_i(\tau)$, where the terms $\bar{\Lambda}$ and $\bar{c}_i(\tau)$ are independent of g , the derivative of individual consumption with respect to g is (minus) the first term on the right-hand side of (B22). The second term is the marginal utility of private consumption.

We now move to the proof of Proposition 9. Substituting the allocations $s(\kappa; \tau)$, $h(\varphi, \varepsilon; \tau)$, and $c(\varphi, \alpha, s; g, \tau)$ into expected utility (7) yields

$$\begin{aligned} U = & -\frac{(1-\tau)\eta}{(1+1/\psi)\theta}\kappa + (1-\beta\delta)\mathbb{E}\sum_{j=0}^{\infty}(\beta\delta)^j\left\{\log\lambda(g,\tau) + \frac{\log(1-\tau)}{(1+\hat{\sigma})} + \mathcal{M} + (1-\tau)\left[\alpha_j - \varphi + \kappa\frac{\eta}{\theta}\right]\right. \\ & \left. + \frac{1-\tau}{(\theta-1)(\psi+1)}\left(\psi\log(1-\tau) + \log\left(\frac{1}{\eta(\theta-1)^{(1+\psi)}}\right)\right) - \frac{1-\tau}{1+\sigma}\exp\left(-\frac{1+\sigma}{\hat{\sigma}(1-\tau)}\mathcal{M} + \frac{1+\sigma}{\hat{\sigma}}\varepsilon\right)\right\} \\ & + \chi\log G. \end{aligned}$$

Now, suppose that the choice of τ is made before observing ε . Then, the term in ε becomes $-\frac{1-\tau}{1+\sigma}$ (see the proof of Corollary 6.1). In addition,

$$(1-\beta\delta)\mathbb{E}\sum_{j=0}^{\infty}(\beta\delta)^j(1-\tau)\alpha_j = (1-\tau)\left(\alpha - \frac{\beta\delta}{1-\beta\delta}\frac{v_\omega}{2}\right).$$

Thus,

$$\begin{aligned} \int U(\kappa, \varphi, \alpha, \varepsilon; g, \tau) dF_\varepsilon = & \log\lambda(g, \tau) + \frac{\log(1-\tau)}{(1+\hat{\sigma})} + \mathcal{M} \\ & + \frac{1-\tau}{(\theta-1)(1+\psi)}\left(\psi\log(1-\tau) + \log\left(\frac{1}{\eta(\theta-1)^{(1+\psi)}}\right)\right) \\ & - (1-\tau)\varphi + \frac{1-\tau\eta}{1+\psi}\kappa - \frac{1-\tau}{1+\sigma} \\ & + (1-\tau)\left(\alpha - \frac{\beta\delta}{1-\beta\delta}\frac{v_\omega}{2}\right) + \chi\log G. \end{aligned}$$

Recall that the baseline social welfare function is

$$\begin{aligned} \mathcal{W}(g, \tau) = & \log\lambda(g, \tau) + \frac{\log(1-\tau)}{(1+\hat{\sigma})} + \mathcal{M} + \frac{1-\tau}{(\theta-1)(1+\psi)}\left(\psi\log(1-\tau) + \log\left(\frac{1}{\eta(\theta-1)^{(1+\psi)}}\right)\right) \\ & - (1-\tau)\frac{v_\varphi}{2} + \frac{(1-\tau)}{(1+\psi)\theta} - \frac{1-\tau}{1+\sigma} \\ & - (1-\tau)\left(\frac{\beta\delta}{\gamma-\beta\delta}\frac{v_\omega}{2}\right) + \chi\log G. \end{aligned}$$

Therefore, we can express expected utility, conditional on the state $(\kappa, \varphi, \alpha)$, as

$$\begin{aligned} U(\kappa, \varphi, \alpha; g, \tau) &= \mathcal{W}(g, \tau) + (1 - \tau) \frac{v_\varphi}{2} - \frac{(1 - \tau)}{(1 + \psi)\theta} \\ &\quad + (1 - \tau) \left(\frac{\beta\delta}{\gamma - \beta\delta} \frac{v_\omega}{2} - \frac{\beta\delta}{1 - \beta\delta} \frac{v_\omega}{2} \right) \\ &\quad + (1 - \tau) \left(\alpha - \varphi + \frac{1}{1 + \psi} \frac{\eta}{\theta} \kappa \right). \end{aligned}$$

Note that $U(\kappa, \varphi, \alpha; g, \tau)$ is strictly concave in τ , since $\mathcal{W}(g, \tau)$ is concave in τ and the additional terms in $U(\kappa, \varphi, \alpha; g, \tau)$ are linear in τ . We need to determine the median voter. A useful property is that the three individual states $(\kappa, \alpha, \varphi)$ enter as a linear combination. Let

$$x = \alpha - \varphi + \frac{1}{1 + \psi} \frac{\eta}{\theta} \kappa.$$

The median voter is the agent with the median value for x . Since α and φ are normally distributed and κ is exponentially distributed, the random variable x follows an Exponentially Modified Gaussian distribution.

B.14 Skill investment constraints

B.14i. Consumption for skilled and unskilled

We begin by calculating the consumption of skilled and unskilled workers. This is needed to evaluate the probabilities that consumption of skilled and unskilled parents exceed the threshold \underline{c} . These probabilities are in turn inputs to the transition of ξ , see equation (46). Consumption of a skilled person is

$$\begin{aligned} \log c &= \log \lambda(\xi) + \frac{1 - \tau}{1 + \sigma} \log(1 - \tau) + \frac{(1 - \tau) [1 - 2\tau - \sigma\tau] v_\varepsilon}{(\sigma + \tau) 2} + (1 - \tau) [\log p(s) + \alpha - \varphi] \\ &= \log \lambda(\xi) + \frac{1 - \tau}{1 + \sigma} \log(1 - \tau) + \frac{(1 - \tau) [1 - 2\tau - \sigma\tau] v_\varepsilon}{(\sigma + \tau) 2} + (1 - \tau) [\alpha - \varphi + \pi_0 + \pi_1 s] \\ &= C(\tau, \xi) + (1 - \tau) \left[\alpha - \varphi + \frac{\eta}{\theta} \kappa + \pi_0 \right], \end{aligned}$$

where $\lambda(\xi)$ is defined below and the constant $C(\tau, \xi)$ is defined as

$$C(\tau, \xi) = \log \lambda(\xi) + \frac{1 - \tau}{1 + \sigma} \log(1 - \tau) + \frac{(1 - \tau) [1 - 2\tau - \sigma\tau] v_\varepsilon}{(\sigma + \tau) 2}$$

Moreover, the random variable $\alpha - \varphi + \frac{\eta}{\theta} \kappa$ is distributed according to an Exponentially Modified Gaussian distribution with an exponential parameter θ and Gaussian parameters $(-\frac{1}{2}(v_\alpha + v_\varphi), v_\alpha + v_\varphi)$.

The probability that the child of a skilled parent will become unskilled is therefore

$$\begin{aligned}\Pr(c < \underline{c} | \text{skilled}) &= \Pr\left(\alpha - \varphi + p(s) < \frac{\ln \underline{c} - C(\tau, \xi)}{1 - \tau} - \pi_0\right) \\ &= \Pr\left(\alpha - \varphi + \frac{\eta}{\theta} \kappa < \frac{\ln \underline{c} - C(\tau, \xi)}{1 - \tau} - \pi_0\right).\end{aligned}$$

Consumption of an unskilled person is

$$\begin{aligned}\ln \underline{c}(\alpha, \varphi, \xi) &= \log \lambda(\xi) + \frac{1 - \tau}{1 + \sigma} \log(1 - \tau) + \frac{(1 - \tau)[1 - 2\tau - \sigma\tau] v_\varepsilon}{(\sigma + \tau)} \frac{1}{2} + (1 - \tau)[\alpha - \varphi + \ln \underline{w}] \\ &= C(\tau, \xi) + (1 - \tau)[\alpha - \varphi + \ln \underline{w}]\end{aligned}$$

The probability that the child of an unskilled parent will be unskilled is

$$\begin{aligned}\Pr(c < \underline{c} | \text{unskilled}, \xi) &= \Pr(C(\tau, \xi) + (1 - \tau)[\alpha - \varphi + \ln \underline{w}] < \ln \underline{c}) \\ &= \Phi\left(\frac{1}{\sqrt{v_\alpha + v_\varphi}} \left(\frac{\ln \underline{c} - C(\tau, \xi)}{1 - \tau} - \ln \underline{w} + \frac{v_\alpha + v_\varphi}{2}\right)\right)\end{aligned}$$

where Φ is a c.d.f. of a standard Normal distribution.

B.14ii. Calculating $\lambda(\xi)$

Recall that effective hours worked N is independent of the skill level. Aggregate output per effective hour worked is therefore

$$\begin{aligned}\frac{Y(\xi)}{N} &= \xi \frac{Y_U}{N} + (1 - \xi) \frac{Y_S}{N} \\ &= \xi \underline{w} + (1 - \xi) \frac{\theta}{\theta - 1} \cdot \exp(\pi_0).\end{aligned}$$

Define \tilde{Y} as $\tilde{Y} = \int y_i^{1-\tau} di$. Budget balance and $G = gY$ then implies $\lambda = (1 - g)Y/\tilde{Y}$. To compute \tilde{Y} , it is useful to aggregate by age group. Let \tilde{Y}^a denote average per capita disposable income for agents of age a :

$$\begin{aligned}\tilde{Y}^a(\xi) &= \int [y(s, \varphi, \varepsilon, \alpha)]^{1-\tau} m(s) ds dF_\alpha^a dF_\varphi dF_\varepsilon \\ &= \xi \int [h(\varepsilon, \varphi) \exp(\alpha_a + \varepsilon) \underline{w}]^{1-\tau} dF_\alpha^a dF_\varphi dF_\varepsilon \\ &\quad + (1 - \xi) \int [h(\varepsilon) \exp(p(s) + \alpha_a + \varepsilon)]^{1-\tau} m(s) ds dF_\alpha^a dF_\varphi dF_\varepsilon \\ &= \left(\xi (\underline{w})^{1-\tau} + (1 - \xi) \int [\exp((1 - \tau)(\pi_0 + \pi_1 s))]^{1-\tau} m(s) ds\right) \int [h(\varepsilon, \varphi) \exp(\alpha_a + \varepsilon)]^{1-\tau} dF_\alpha^a dF_\varphi dF_\varepsilon.\end{aligned}$$

Substituting in the hours allocation (20), the expression for the skill price (22), the density function $m(s)$ (B9), and integrating, we arrive at

$$\ln \tilde{Y}^a(\xi) = \mathcal{L} - \tau(1-\tau) \frac{v_\alpha^a}{2} + \ln \left[\xi (\underline{w})^{1-\tau} + (1-\xi) \exp((1-\tau)\pi_0) \frac{\theta}{\theta-1+\tau} \right],$$

where the constant \mathcal{L} , common for all cohorts, is defined as

$$\mathcal{L} = \frac{1-\tau}{1+\sigma} \ln(1-\tau) - \tau(1-\tau) \frac{v_\varphi}{2} + \mathcal{M},$$

where \mathcal{M} is defined in Proposition 1. Now sum across age groups to obtain

$$\begin{aligned} \tilde{Y}(\xi) &= (1-\delta) \sum_{a=0}^{\infty} \delta^a \tilde{Y}^a = \exp(\mathcal{L}) \times \left[\xi (\underline{w})^{1-\tau} + (1-\xi) \exp((1-\tau)\pi_0) \frac{\theta}{\theta-1+\tau} \right] \times \\ &\quad \frac{(1-\delta) \exp\left(-\tau(1-\tau) \frac{v_\alpha^0}{2}\right)}{1-\delta \exp\left(\frac{-\tau(1-\tau)}{2} v_\omega\right)}. \end{aligned}$$

Substituting (B12) and (B15) into (B14) and simplifying, we arrive at a solution for the equilibrium value of λ which, in logs, is

$$\begin{aligned} \ln \lambda(\xi) &= \ln(1-g) + \ln Y - \ln \tilde{Y} \\ &= \ln(1-g) + \ln \left[\xi \underline{w} + (1-\xi) \frac{\theta}{\theta-1} \cdot \exp(\pi_0) \right] + \frac{\ln(1-\tau)}{1+\sigma} + \frac{\tau(1+\hat{\sigma}) + \hat{\sigma} v_\varepsilon}{\hat{\sigma}^2} \frac{1}{2} \\ &\quad - \left(\frac{1-\tau}{1+\sigma} \ln(1-\tau) - \tau(1-\tau) \frac{v_\varphi}{2} + \frac{(1-\tau)(1-\tau(1+\hat{\sigma})) v_\varepsilon}{\hat{\sigma}} \frac{1}{2} \right) \\ &\quad - \ln \left[\xi (\underline{w})^{1-\tau} + (1-\xi) \exp((1-\tau)\pi_0) \frac{\theta}{\theta-1+\tau} \right] \\ &\quad - \ln(1-\delta) + \tau(1-\tau) \frac{v_\alpha^0}{2} + \ln \left(1 - \delta \exp\left(\frac{-\tau(1-\tau)}{2} v_\omega\right) \right) \end{aligned}$$

It follows that the model with unskilled workers and investment constraints changes λ relative to its equivalent in the benchmark economy analyzed in Sections III.–VI., i.e., λ_{BM} , as follows,

$$\ln \lambda(\xi) = \ln \lambda_{BM} + \ln \left(\frac{\theta-1}{\theta-1+\tau} \right) - \tau\pi_0 + \ln \left(\frac{\xi \underline{w} + (1-\xi) \frac{\theta}{\theta-1} \cdot \exp(\pi_0)}{\xi (\underline{w})^{1-\tau} + (1-\xi) \frac{\theta}{\theta-1+\tau} \exp((1-\tau)\pi_0)} \right).$$

Note that $\lambda(0) = \lambda_{BM}$.

B.14iii. Social welfare in the presence of investment constraints

Consider now the expression for social welfare when the share of unskilled, ξ_t , is changing over time. It is convenient to express social welfare by adding up current utilities, appropriately discounted, at each point in time;

$$\begin{aligned} \mathcal{W}(g, \tau; \{\xi_t\}_{t=0}^{\infty}) &= \frac{(1-\gamma)(\gamma-\beta\delta)}{\gamma(1-\beta\delta)} \left[\sum_{t=0}^{\infty} (\gamma^t(\beta\delta)^0 \bar{u}_{0,t} + \gamma^{t-1}(\beta\delta)^1 \bar{u}_{1,t} + \gamma^{t-2}(\beta\delta)^2 \bar{u}_{2,t} + \dots) \right. \\ &\quad \left. - \sum_{t=0}^{\infty} \gamma^t (1-\xi_t) \bar{v} - (1-\xi_0) \sum_{j=-\infty}^{-1} \gamma^j (\beta\delta)^{-j} \bar{v} \right], \end{aligned}$$

where $\bar{u}_{a,t}$ denotes the expected current utility (excluding the education costs) of an individual who is a years old in period t . The last term captures the education costs of the individuals alive at the time of the reform. Recall that education costs are fully reversible, so we can ignore the investments these individuals made before the reform. Note that the age component of $\bar{u}_{a,t}$ does not depend on ξ_t and is therefore time invariant. We can therefore express it as $\bar{u}_{a,t} = \bar{u}_t + \bar{u}^a$, where $\bar{u}^a = (1-\beta\delta)(1-\tau)jv_{\omega}/2$ (see the proof of Proposition 1). It follows that \mathcal{W} can be expressed, after some algebra, as

$$\begin{aligned} \mathcal{W}(g, \tau; \{\xi_t\}_{t=0}^{\infty}) &= (1-\gamma) \sum_{t=0}^{\infty} \gamma^t \frac{\bar{u}_t}{1-\beta\delta} - (1-\tau) \frac{v_{\omega}}{2} \frac{\beta\delta}{\gamma-\beta\delta} \\ &\quad - \left(\frac{\beta\delta}{\gamma} (1-\xi_0) + \frac{\gamma-\beta\delta}{\gamma} \sum_{t=0}^{\infty} \gamma^t (1-\xi_t) \right) \frac{(1-\gamma)}{(1-\beta\delta)} \bar{v}. \end{aligned}$$

We must now calculate expected period utility excluding education costs, \bar{u}_t . Recall that the current utility – excluding the investment cost – for a skilled agent with state $(\kappa, a, \alpha, \varphi, \varepsilon, \xi)$ is Recall that the current utility – excluding the investment cost – for a skilled agent with state $(\kappa, a, \alpha, \varphi, \varepsilon, \xi)$ is

$$\begin{aligned} u(c_a, h, G) &= \chi \left(\ln g + \ln \left(\frac{Y}{N} \right) + \ln N \right) + \log \lambda + \frac{1-\tau}{1+\sigma} \log(1-\tau) \\ &\quad - (1-\tau) \varphi + (1-\tau) \alpha_a - \exp \left(-\frac{1+\sigma}{\hat{\sigma}(1-\tau)} \mathcal{M} \right) \exp \left(\frac{1+\sigma}{\hat{\sigma}} \varepsilon \right) \frac{(1-\tau)}{(1+\sigma)} + \mathcal{M} \\ &\quad + (1-\tau) \kappa \frac{\eta}{\theta} + (1-\tau) \pi_0, \end{aligned}$$

The utility for an unskilled person is the same, except that the last line is replaced by $(1-\tau) \ln \underline{w}$. Take the cross-sectional expectation over $u(c_a, h, G)$ w.r.t. ε, φ , and α_a in equation (B23). The

first two lines of the welfare then become

$$\begin{aligned} & \chi \ln g + \chi \ln \left(\xi \underline{w} + (1 - \xi) \frac{\theta}{\theta - 1} \cdot \exp(\pi_0) \right) + \chi \left(\frac{1}{1 + \sigma} \ln(1 - \tau) + \frac{\tau(1 + \hat{\sigma}) + \hat{\sigma} v_\varepsilon}{\hat{\sigma}^2} \frac{1}{2} \right) + \log \lambda \\ & + \frac{1 - \tau}{1 + \sigma} \log(1 - \tau) - \frac{1 - \tau}{1 + \sigma} + \mathcal{M} - (1 - \tau) \frac{v_\varphi + v_{\alpha_a}}{2} \end{aligned}$$

Now take the expectation over κ in equation (B23), which implies that the third line yields

$$E \left\{ (1 - \tau) \kappa \frac{\eta}{\theta} + (1 - \tau) \pi_0 \right\} = (1 - \tau) \pi_0 + \frac{1 - \tau}{\theta}$$

Calculate now the cross-sectional utility cost of skilled investment:

$$E \{ -v(s(\kappa; \tau)) \} = -\frac{\psi}{1 + \psi} \frac{1}{\theta} (1 - \tau)$$

When adding up current utility across generations, the term $-(1 - \tau)^2 v_{\alpha_a}/2$ becomes, as in Proposition 3,

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

It follows that social welfare can be expressed as

$$\begin{aligned} \mathcal{W}(g, \tau; \{\xi_t\}_{t=0}^\infty) &= (1 - \gamma) \sum_{t=0}^\infty \gamma^t \frac{\bar{u}_t}{1 - \beta \delta} - (1 - \tau) \frac{v_\omega}{2} \frac{\beta \delta}{\gamma - \beta \delta} \\ &\quad - \left(\frac{\beta \delta}{\gamma} (1 - \xi_0) + \frac{\gamma - \beta \delta}{\gamma} \sum_{t=0}^\infty \gamma^t (1 - \xi_t) \right) \frac{(1 - \gamma)}{(1 - \beta \delta)} \frac{\psi}{1 + \psi} \frac{1 - \tau}{\theta}. \end{aligned}$$

where

$$\begin{aligned} \frac{\bar{u}_t}{1 - \beta \delta} &= \ln(1 - g) + \chi \ln g + (1 + \chi) \frac{\tau}{1 + \sigma} \ln(1 - \tau) - \frac{1 - \tau}{1 + \sigma} \\ &\quad + (1 + \chi) \frac{\tau(1 + \hat{\sigma}) + \hat{\sigma} v_\varepsilon}{\hat{\sigma}^2} \frac{1}{2} \\ &\quad - (1 - \tau)^2 \frac{v_\varphi + v_\alpha^0}{2} - \ln(1 - \delta) + \ln \left(1 - \delta \exp \left(\frac{-\tau(1 - \tau) v_\omega}{2} \right) \right) \\ &\quad + (1 + \chi) \ln \left[\xi_t \underline{w} + (1 - \xi_t) \frac{\theta}{\theta - 1} \cdot \exp(\pi_0) \right] \\ &\quad - \ln \left[\xi_t (\underline{w})^{1 - \tau} + (1 - \xi_t) \exp \left((1 - \tau) \pi_0 \right) \frac{\theta}{\theta - 1 + \tau} \right] \\ &\quad + (1 - \xi_t) (1 - \tau) \left(\pi_0 + \frac{1}{\theta} \right) + \xi_t (1 - \tau) \ln \underline{w}, \end{aligned}$$

Note that, as in the benchmark case, the public good provision g does not interact with neither τ nor ξ , so it will not matter for the optimal τ . With $\gamma = \beta$ the welfare expression simplifies to

$$\begin{aligned}\mathcal{W} &= \frac{1-\beta}{1-\beta\delta} \sum_{t=0}^{\infty} \beta^t \bar{u}(\xi_t) - (1-\tau) \frac{\delta}{1-\delta} \frac{v_\omega}{2} - \frac{\psi}{1+\psi} \frac{1-\tau}{\theta} \frac{(1-\beta)}{(1-\beta\delta)} \left(1 - \xi_0 + (1-\delta) \sum_{t=1}^{\infty} \gamma^t (1 - \xi_t) \right) \\ &= \frac{(1-\beta)}{(1-\beta\delta)} \left[\sum_{t=0}^{\infty} \beta^t \bar{u}(\xi_t) - \left(1 - \xi_0 + (1-\delta) \sum_{t=1}^{\infty} \gamma^t (1 - \xi_t) \right) \frac{\psi}{1+\psi} \frac{1-\tau}{\theta} \right]\end{aligned}$$

For each possible τ , we solve numerically for the sequence $\{\xi_t\}_{t=0}^{\infty}$ and the associated social welfare \mathcal{W} .

APPENDIX C

This Appendix provides additional detail to Section VIII. in the main paper. We begin by describing the construction of the data set, and next we present a sensitivity analysis for our regressions.

C.1 Cross-country dataset construction

Tax progressivity. The Andrew Young School World Tax Indicator (WTI) database (Andrew Young School of Policy Studies, 2010), contains measures of personal income tax rates and tax progressivity for 189 countries during the period 1981-2005 (for many countries data are missing in the first part of the sample). The enclosed documentation states that tax rates adjust for allowances/deductions, tax credits, significant local taxes and other main rules of the tax code. For each country-year pair (c, t) , the dataset reports an index of marginal rate progression (MRP_t^c), i.e. the slope coefficient from regressing marginal tax rates on the log of gross income. Recall that in our model, $MTR(y) = 1 - \lambda(1 - \tau)y^{-\tau}$, thus using the approximation $\log(1 - MTR(y)) \simeq -MTR(y)$ reveals that this slope coefficient approximately equals τ .

The data also contain average and marginal tax rates at $y, 2y, 3y,$ and y where y is average per-capita income. Thus, we can also estimate exactly the parameter τ in our tax function for country c at time t as

$$(C1) \quad \tau_t^c(y) = 1 - \frac{1 - MTR_t^c(y)}{1 - ATR_t^c(y)},$$

and averaging over the four levels of y available, we obtain an estimate of τ_t^c . Since information on marginal and average tax rates below average income is used to calculate the MRP_t^c index, but it is not available in the public data, and thus we cannot use it for this alternative strategy, the MRP_t^c index is a more comprehensive measure of progressivity and is the one we use in the baseline regressions below. The correlation between the two measures is 0.92 and significant at 1 pct level, and our results are robust to using the alternative estimate of τ_t^c as we show below.

The crucial feature of these data is that estimates of progressivity are comparable across country since they are computed with the exact same method and similar data sources.

Proxy for χ . The parameter χ directly determines the government expenditure share of output, $g = G/Y$, equal to $\chi/(1 + \chi)$ in the model, so use this variable as a proxy for χ . Data on

government consumption as a fraction of GDP come from the Penn World Tables version 8.1 (Feenstra, Inklaar, and Timmer, 2015). We use the share of government consumption at current PPPs (variable name *cash_g*). From the PWT, we also obtain country GDP and population, which we use to construct our weights in the regressions.

Proxy for θ . From the model, we know that θ is the Pareto-Lorenz coefficient of the income distribution. The World Wealth and Income Database (Alvaredo et al, 2016) provides estimates of this parameter for a number of countries.

Proxy for variances of earnings. In our model, the variance of labor income risk (v_α, v_ε) and of preference heterogeneity v_φ determine income inequality, over and above the role played by θ . Measures of cross-sectional gross income inequality are therefore good proxies for the size of these variances. The most comprehensive cross-country dataset on income inequality is the Standardized World Income Inequality Database (SWIID) (Solt, 2016) which contains estimates of Gini coefficients for equivalized household market income for 173 countries going back, for some countries like the United States, all the way to 1960. The data are available at <http://fsolt.org/swiid/>. Each country/year observation has 100 different estimates for the Gini that reflect the degree of uncertainty in estimates and imputation procedures: from these we construct an average and use the average as our estimate of Gini for that country/year pair.

The income distribution in our model is Pareto-LogNormal, i.e. it is a lognormal distribution with a Pareto upper tail. Griffiths and Hajargasht (2013) provide a closed form expression for the Gini coefficient of such distribution (G^{P-LN}) as a function of only two parameters, the Pareto coefficient (θ in the model) and the variance of the lognormal distribution ($\bar{v} = v_\alpha + v_\varphi + \left(\frac{1+\hat{\sigma}}{\hat{\sigma}}\right)^2 v_\varepsilon$ in the model):

$$G^{P-LN} = \frac{2 \exp(\theta(\theta - 1)\bar{v})}{2\theta - 1} \Phi\left(\left(1 - 2\theta\right)\sqrt{\frac{\bar{v}}{2}}\right) + 2\Phi\left(\sqrt{\frac{\bar{v}}{2}}\right) - 1$$

where Φ is the cdf of the lognormal distribution. Thus, given our country-specific estimates for the Gini and for θ , we can recover \bar{v} .

In our main analysis, we start the sample in 1990 for two reasons: (i) in the earlier years many countries have missing data, and (ii) we can include the countries in the Eastern block in all our analysis. The last year of the sample is 2005. Once we merge all the information together, our final

TABLE C1: Empirical determinants of progressivity across countries

	(1)	(2)	(3)	(4)	(5)	(6)
	τ	τ	τ	τ	τ	τ
G/Y	-0.0978*** (-4.19)	-0.326*** (-4.24)	-0.0970*** (-4.01)	-0.443*** (-5.18)	0.0703*** (3.39)	-0.214** (-2.83)
Income Gini	0.0861** (3.30)		0.132*** (5.26)		0.0900*** (3.68)	
θ		0.285*** (8.62)		0.251*** (8.18)		0.0274 (0.80)
θ^2		-0.0551*** (-8.79)		-0.0455*** (-8.37)		-0.00340 (-0.53)
\bar{v}		0.0328** (2.61)		-0.0117 (-0.71)		0.0195 (1.45)
Regional Dummies	N	N	Y	Y	Y	Y
Development Dummies	N	N	N	N	Y	Y
N	1585	351	1585	351	1585	351
adj. R^2	0.018	0.285	0.377	0.500	0.574	0.644

t statistics in parentheses: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

dataset comprises of 1585 country-years observations in its largest configuration that includes data on progressivity, inequality and government consumption share of output. The number of countries grows from 65 in 1990 to over 103 in the latest years, and 135 countries are present for at least one year. The median number of observations per country is 14. The dataset includes 351 country-year observations in its smallest configuration that also includes data on Pareto coefficients, and thus exogenous variances of income \bar{v} , i.e. 26 countries per year, most of which are present for all 16 years.

C.2 Sensitivity

In Table C1 we report the counterpart of Table V where, instead of the marginal rate of progressivity, we use the measure of τ in equation (C1) as our dependent variable. As clear from a comparison of the two tables, results are very robust.

C.3 Theoretical counterparts to the empirical regression coefficients

We now describe how we compute the theoretical analogues of the regression coefficients reported in Table V. We start with our baseline expression for social welfare, eq. 30. Recall that for

each country-year in the data we observe government purchases as a share of output, g_t^c , the variance of the normal component of earnings, \bar{v}_t^c , and the Pareto coefficient, θ_t^c . The theoretical expression for social welfare instead involves the structural parameters $(\chi, v_\alpha, v_\varphi, v_\varepsilon, \theta)$. We use the optimality condition for government purchases, $g = \chi/(1 + \chi)$ to translate the sensitivity of optimal progressivity with respect to χ to the implied sensitivity with respect to g_t^c . Since we do not have country-specific empirical counterparts for v_α , v_φ or v_ε we will assume that the ratios of these variances to \bar{v}_t^c are common across countries, and equal to the ratios implied by our calibration to the United States. Let δ_α , δ_φ and δ_ε denote these ratios, where, for example, $\delta_\alpha = v_\alpha/(v_\alpha + v_\varphi + ((1 + \hat{\sigma})^2/\hat{\sigma}^2)v_\varepsilon)$. Given $(v_\alpha)_t^c = \delta_\alpha \bar{v}_t^c$, $(v_\varphi)_t^c = \delta_\varphi \bar{v}_t^c$, $(v_\varepsilon)_t^c = \delta_\varepsilon \bar{v}_t^c$, we can compute the sensitivity of optimal progressivity with respect to \bar{v}_t^c .

Differentiating eq. 30 with respect to τ gives the condition that implicitly defines optimal progressivity, as a function of structural parameters:

$$F((1 - \tau), \chi, \theta, \bar{v}, \sigma, \psi) = 0$$

where

$$\begin{aligned} F((1 - \tau), \chi, \theta, \bar{v}, \sigma, \psi) &= \frac{-(1 + \chi)}{(1 + \sigma)(1 - \tau)} + \frac{1}{1 + \sigma} - (1 + \chi) \frac{\psi}{(1 + \psi)(\theta - 1)} \frac{1}{(1 - \tau)} + \frac{\psi}{(1 + \psi)\theta} \\ &\quad + \frac{\frac{1}{\theta}}{1 - (\frac{1 - \tau}{\theta})} - \frac{1}{\theta} \\ &\quad + \left((1 - \tau) (\delta_\alpha + \delta_\varphi) - \frac{(1 - (1 - \tau)) (\sigma + 1)^2 (1 + \chi)}{((\sigma + 1) - (1 - \tau))^3} \delta_\varepsilon \right) \bar{v} \end{aligned}$$

Using the implicit function theorem, we can compute the theoretical sensitivity of progressivity τ with respect to the variance of the normally distributed component of earnings \bar{v} :

$$\begin{aligned} \frac{d\tau}{d\bar{v}} &= -\frac{d(1 - \tau)}{d\bar{v}} \\ &= \frac{(1 - \tau) (\delta_\alpha + \delta_\varphi) - \tau (\sigma + 1)^2 \frac{(1 + \chi)}{(\sigma + \tau)^3} \delta_\varepsilon}{\frac{(1 + \chi)}{(1 + \sigma)(1 - \tau)^2} + (1 + \chi) \frac{\psi}{(1 + \psi)(\theta - 1)} \frac{1}{(1 - \tau)^2} + \frac{1}{(\theta - (1 - \tau))^2} + \left((\delta_\alpha + \delta_\varphi) + \frac{(-2\tau + \sigma)}{(\sigma + \tau)^4} (1 + \sigma)^2 (1 + \chi) \delta_\varepsilon \right) \bar{v}} \end{aligned}$$

Evaluated at our calibration to the United States, and at the value for τ that is optimal given that calibration (0.084), this derivative is equal to 0.175.

Now consider the sensitivity of progressivity to government purchases. By the Chain Rule,

$$\frac{d\tau}{dg} = \frac{d\tau}{d\chi} \frac{d\chi}{dg} = \frac{d\tau}{d\chi} \frac{1}{(1 - g)^2}$$

Using the implicit function theorem to compute $\frac{d(1-\tau)}{d\chi}$ gives

$$\begin{aligned} \frac{d\tau}{dg} &= -\frac{d(1-\tau)}{d\chi} \frac{1}{(1-g)^2} \\ &= \frac{1}{(1-g)^2} \frac{-\frac{1}{(1+\sigma)(1-\tau)} - \frac{\psi}{(1+\psi)(\theta-1)} \frac{1}{(1-\tau)} - \tau \frac{(\sigma+1)^2}{(\sigma+\tau)^3} v \delta_\varepsilon}{\frac{1+\chi}{(1+\sigma)(1-\tau)^2} + \frac{(1+\chi)\psi}{(1+\psi)(\theta-1)} \frac{1}{(1-\tau)^2} + \frac{1}{(\theta-(1-\tau))^2} + \left((\delta_\alpha + \delta_\varphi) + \frac{(-2\tau+\sigma)}{(\sigma+\tau)^4} (1+\sigma)^2 (1+\chi) \delta_\varepsilon \right) \bar{v}} \end{aligned}$$

Evaluating this expression at our baseline parameter values and at $g = 0.189$ and $\tau = 0.084$, we find $d\tau/dg = -0.690$.

APPENDIX D: ADDITIONAL MATERIAL

D.1 CES skill aggregator in our model versus Benabou (2005)

Our paper studies optimal taxation within the class of log-linear tax functions in eq. (1) in the main text, assuming a production technology with a constant elasticity of substitution between different skill inputs. As discussed in the main text, Benabou (2002, 2005) studies optimal taxation within the same class of tax functions and he too assumes a CES production function over various skill inputs. However, as it turns out, the link between the optimal tax progressivity parameter τ and the elasticity of substitution between output from different skill types, θ , is quite different in the two models. In our model, the optimal progressivity τ is hump-shaped in θ : rising in θ for low levels of θ and falling in θ for large θ (see Figure 2 of the main text). In contrast, in Benabou (2002, 2005), the optimal progressivity falls monotonically with the elasticity of substitution θ . We now explain why the link between θ and τ in Benabou (2002, 2005) is so different from the link in our paper.

Consider the formulation in Benabou (2005). Section 3.2 of his paper develops an extension of the baseline model to a case in which skills are aggregated in a CES fashion to produce a final good. Equation (30) in his paper describes this technology (in equilibrium, with constant labor supply). Note that the distribution of agents across what he labels “skills”, $g(i)$, is uniform by assumption. Re-expressed in our notation, his technology is

$$\frac{Y}{N} = \left(\int_0^1 [k(i) \cdot g(i)]^{\frac{\sigma-1}{\sigma}} di \right)^{\frac{\sigma}{\sigma-1}},$$

where $g(i) = 1$ and $k(i)$ is human capital of individual i .

Contrast this to our baseline production technology (our eq.5), which is

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

where $m(s)$ is (in equilibrium) an exponential distribution. In our model skill is a choice and due to the cost of choosing a higher skill, the equilibrium distribution of skills $m(s)$ has density falling with s .

These two technologies may appear to be very similar, but the way the two models work is quite different. In Benabou’s model, the distribution $g(i)$ is exogenous and uniform, and progressivity

effects the distribution of $k(i)$. Given a uniform distribution $g(i)$, productivity is maximized when human capital is evenly distributed, which calls for progressive taxation in his model. Moreover, this force is larger the smaller is θ .

In our model, in contrast, effective hours $N(s)$ (our formal analogue to Benabou's human capital) are, in equilibrium, equal across skill types. The way in which the tax system impacts the distribution of effective hours by different skill types is by affecting the shape of the distribution $m(s)$. Now the more progressive is the tax system, the more clustered towards zero is the distribution $m(s)$, since agents do less skill investment, and high skill types are more scarce. Because a uniform distribution for $m(s)$ maximizes productivity, progressivity is therefore productivity-reducing, and this force is larger the smaller is θ .

D.2 Comparison of actual and optimal tax systems

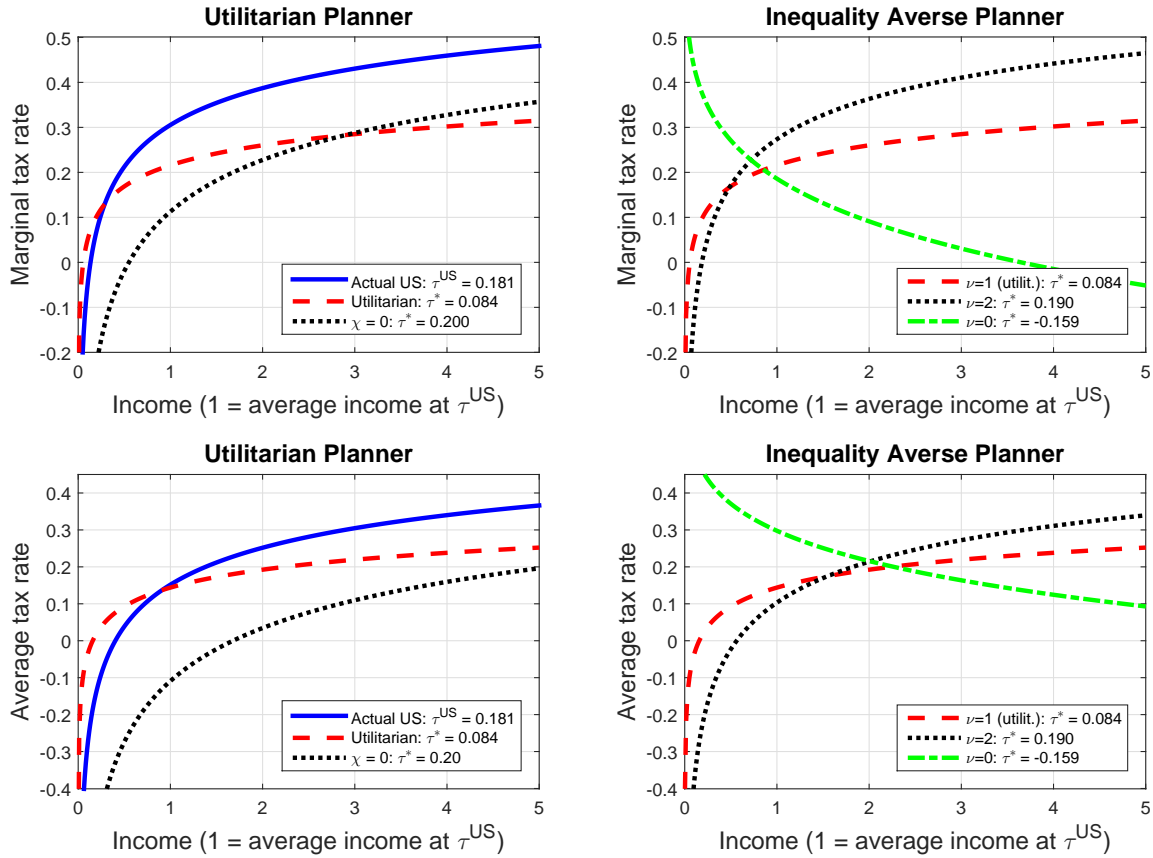


FIGURE D2: Left panels: the US tax system compared to the utilitarian optimum with valued government expenditures, and to the utilitarian optimum when government expenditures are not valued ($\chi = 0$). Right panels: Utilitarian optimum compared to the optimum for a planner with more inequality aversion than the utilitarian planner ($\nu = 2$) and for an inequality-neutral planner ($\nu = 0$).

D.3 Comparison between fixed and flexible skill investment models

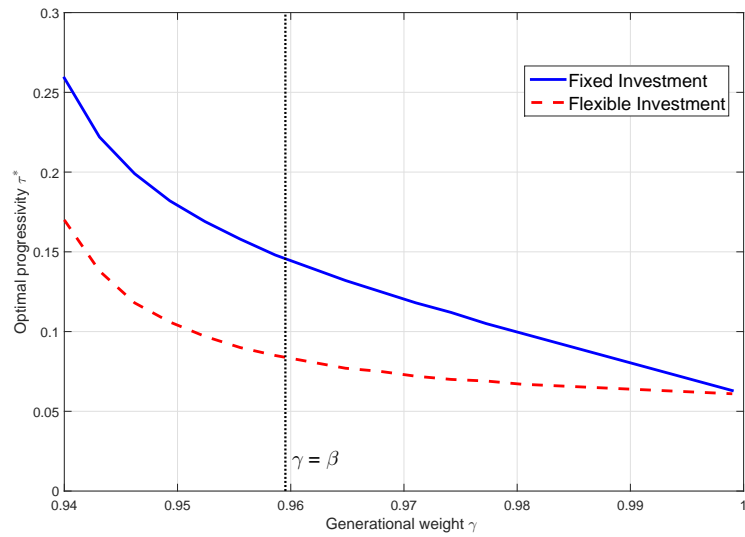


FIGURE D3: Optimal degree of progressivity τ^* as a function of the generational weight γ in the fixed investment model and the baseline flexible investment model.

References

- [1] Benabou, R (2002) “Tax and Education Policy in a Heterogeneous-Agent Economy: What Levels of Redistribution Maximize Growth and Efficiency?,” *Econometrica*, 70(2), 481–517.
- [2] Benabou, R (2005) *Inequality, Technology, and the Social Contract*, in Handbook of Economic Growth, P. Aghion and S. Durlauf eds., North-Holland: Vol 1B, Chapter 25, 1595-1638.
- [3] Feenberg, D., and E. Coutts (1993): “An Introduction to the TAXSIM Model,” *Journal of Policy Analysis and Management*, 12(1), 189-194.
- [4] Heathcote, J., F. Perri, and G. L. Violante (2010): “Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States, 1967-2006,” *Review of Economic Dynamics*, 13(1), 15-51.
- [5] Heathcote, J., K. Storesletten, and G. L. Violante (2014): “Consumption and Labor Supply with Partial Insurance: An Analytical Framework,” *American Economic Review*, 104(7), 2075-2126.
- [6] Jung, J., and C. Tran (2013): “Medical Consumption over the Life Cycle: Facts from a US Medical Expenditure Panel Survey,” UNSW Australian School of Business Research Paper.
- [7] List, J. A. (2011): “The Market for Charitable Giving,” *Journal of Economic Perspectives*, 25(2), 157-180.
- [8] Hajargasht, G., and W.E. Griffiths (2013): “Pareto–Lognormal distributions: Inequality, Poverty, and Estimation from Grouped Income Data,” *Economic Modelling*, 33, 593–604.