ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Heathcote, Jonathan; Storesletten, Kjetil; Violante, Giovanni L.

Working Paper

Optimal tax progressivity: An analytical framework

IFS Working Papers, No. W14/27

Provided in Cooperation with:
Institute for Fiscal Studies (IFS), London

Suggested Citation: Heathcote, Jonathan; Storesletten, Kjetil; Violante, Giovanni L. (2014) : Optimal tax progressivity: An analytical framework, IFS Working Papers, No. W14/27, Institute for Fiscal Studies (IFS), London, https://doi.org/10.1920/wp.ifs.2014.1427

This Version is available at:
https://hdl.handle.net/10419/119560

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

Institute for Fiscal Studies

Optimal tax progressivity: an analytical framework

IFS Working Paper W14/27

Jonathan Heathcote
Kjetil Storesletten
Giovanni L. Violante

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

The Institute for Fiscal Studies (IFS) is an independent research institute whose remit is to carry out rigorous economic research into public policy and to disseminate the findings of this research. IFS receives generous support from the Economic and Social Research Council, in particular via the ESRC Centre for the Microeconomic Analysis of Public Policy (CPP). The content of our working papers is the work of their authors and does not necessarily represent the views of IFS research staff or affiliates.

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

THIS VERSION: JANUARY 30, 2014

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. At the same time, progressivity reduces incentives to work and to invest in skills, and aggravates the externality associated with valued public expenditures. 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 preferences, 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 externality linked to valued government purchases play quantitatively similar roles in limiting desired progressivity.

Keywords: Progressivity, Income Distribution, Skill Investment, Labor Supply, Partial Insurance, Valued Government Expenditures, Welfare

JEL Codes: 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.

1 Introduction

In deciding 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 private markets to hedge against poor "initial conditions" that translate into low expected 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 the associated 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; 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 create more inequality in pre-tax wages, thereby undermining the original redistributive intent (e.g., Feldstein, 1973; Stiglitz, 1982).

An additional factor impacting on the optimal degree of progressivity comes into play when the government provides goods and services that are valued by households but cannot be purchased privately. Individuals do not then internalize that, by working more hours or acquiring more skills, the associated additional output allows the government to supply more public goods. This 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 (2014). The economy is populated by a continuum of infinitely lived households that choose how much to consume and how much to work, and which face idiosyncratic labor market shocks of two types. Some shocks are privately insurable and do not transmit to consumption, whereas others remain uninsurable in equilibrium and induce consumption volatility. Individuals are heterogeneous ex ante with respect to two characteristics: learning ability and disutility of work effort. Those en-

dowed with higher learning ability invest more in skills prior to entering the labor market, and more diligent individuals work and earn more at every skill level. An aggregate production technology with imperfect substitutability across skill types determines the marginal product and equilibrium price of each skill type. The resulting equilibrium income distribution features a Pareto tail whose coefficient is exactly the elasticity of substitution across skill types in production.

The government uses a nonlinear income tax and transfer system to provide social insurance and to finance publicly provided goods and services. According to this tax system, 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 (we discuss this class of tax and transfer systems in detail in Section 2). In addition to τ, the planner also chooses λ, which determines net tax revenue and thus the share of output g devoted to public goods.

Because the model is tractable and parsimonious, we can derive a closed-form expression for social welfare as a function of τ and g and the (six) structural parameters of the model describing preferences, technology, and households' access to private consumption insurance. Each term in this welfare expression has an economic interpretation and embodies one of the channels shaping the optimal progressivity trade-off discussed above. With this expression in hand, we ask what degree of progressivity τ would be chosen by a benevolent planner.

The planner's desire to provide social insurance with respect to privately uninsurable idiosyncratic productivity shocks calls for a positive value for τ and thus marginal tax rates that rise with earnings. Similarly, initial heterogeneity in innate learning ability and preference for leisure translates into consumption dispersion that a utilitarian planner would like to counteract via a progressive tax and transfer system. However, the planner understands that the more progressive taxes are, the lower labor supply and skill investment will be, where the respective elasticities with respect to τ are governed by the Frisch elasticity of labor supply and the degree of complementarity between skill types in production. In addition, the presence of valued government expenditure constitutes a force toward regressive taxation ($\tau < 0$).

After qualitatively inspecting these channels, we investigate their relative quantitative impacts on optimal net progressivity. The model yields closed-form solutions for the cross-sectional (co-)variances of wages, hours, and consumption. Exploiting the empirical counterparts of these moments from the Panel Study of Income Dynamics and the Consumer Expenditure Survey for 2000-

2006, we estimate the structural parameters determining the relative magnitude of the forces at play and perform a quantitative analysis. Our findings indicate that a utilitarian government would choose less progressivity than is currently embedded in the U.S. tax/transfer system.[1] The optimal value for τ^* is 0.062, which implies an average (income-weighted) marginal tax/transfer rate of 24% compared with the current 31%. Switching to the optimal τ^* yields welfare gains on the order of half a percent of lifetime consumption. Endogenous labor supply and endogenous skill investment play quantitatively similar roles in limiting progressivity, and in the absence of either one of these channels, optimal progressivity would be substantially higher.

We consider a range of sensitivity analyses and extensions that further illuminate the economic forces determining optimal progressivity. When we mute the desire for redistribution in the social welfare objective function to isolate the insurance motive, the optimal tax/transfer system is close to a flat tax set at 19% of income. The logic is that although progressivity does act as a substitute for missing insurance against life-cycle productivity shocks, it also depresses labor supply and skill investment, which are already inefficiently low in the presence of publicly provided goods. These forces almost exactly offset each other and lead to a proportional tax. If government expenditures are not valued by households, one of the key forces towards regressivity vanishes, and the optimal degree of progressivity becomes similar to that in the actual U.S. system. Progressive *consumption* taxation offers more efficient insurance with respect to lifetime productivity shocks than progressive earnings taxation because consumption is independent of the insurable component of earnings fluctuations that, ideally, the planner wants to leave undistorted. Finally, if existing cohorts cannot modify their skill levels after labor market entry, the planner prefers more progressivity than in our baseline (reversible investment) model, since the planner can then redistribute without reducing skill investment in the short run.

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 Benabou (2002). Common to both models is the absence of trade in non-contingent bonds (an assumption in Benabou's model, an equilibrium outcome in ours), which helps deliver analytical tractability. We also adopt the same specification for the tax/transfer function. Key elements that

[1] By "current" system, we mean the one that was in place until the mid-2000s. Recent fiscal measures (e.g., extensions of UI benefits and the sunsetting of the Bush tax cuts) have increased progressivity further.

differentiate our framework are our multiskill production technology, the partial insurance structure, heterogeneity in the taste for work, and the presence of valued government-provided goods. These elements allow us to make closer contact to micro-data and to analyze new forces shaping optimal progressivity that turn out to be quantitatively important. Benabou also postulates a different model for human capital investment, in which goods are an input, which allows him to explore how education subsidies relax credit constraints.

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, valued government expenditures, policy effects on skill prices) and more stylized in others (notably, the fact that wealth is in zero net supply). Relative to these papers, the key advantage of our framework is that it is tractable, and thus 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 class of tax/transfer schemes. The Mirrlees (1971) approach to optimal taxation is built on a different foundation. Rather than postulating an exogenously restricted set of instruments, the goal is to characterize the fully optimal tax system in the context of an informational friction that prevents the planner from directly observing individual productivity and thus rules out productivity-type-specific lump-sum taxes. Classic examples of this approach, with quantitative applications to the U.S. economy, are Saez (2001) and Diamond and Saez (2011).

Although the Mirrlees approach allows more flexibility in the design of the tax system, the problem of solving for constrained-efficient allocations becomes quite difficult outside simple static environments. Researchers have only recently incorporated persistent labor productivity shocks (Farhi and Werning, 2012; Golosov, Troshkin, and Tsyvinski, 2012; Gorry and Oberfield, 2012), human capital accumulation (Stantcheva, 2013), and imperfect substitutability across worker types (Rothschild and Scheuer, 2013). Our model embeds all of these ingredients, yet remains tractable. The cost we pay is that we exogenously restrict the set of tax instruments available to the planner. However, we will argue that our parametric specification is sufficiently flexible that the potential welfare gains from moving to a fully nonparametric tax schedule are likely to be small.

The rest of the paper is organized as follows. Section 2 presents our tax function and discusses

its properties. Section 3 describes the economic environment. Section 4 contains a characterization of the equilibrium allocations in closed form. Section 5 solves analytically for social welfare as a function of the fiscal policy chosen by the government (progressivity τ and public spending g) and as a function of all other structural parameters of the model. Section 6 calibrates the model and explores the quantitative implications of the theory for the optimal degree of progressivity. Section 7 contains four extensions: a politico-economic analysis, progressive consumption taxation, transitional dynamics, and the introduction of skill bias in the production technology. Section 8 concludes. All proofs are in the Appendix.

2 Tax function

We study the optimal degree of progressivity within the class of tax and transfer policies defined by

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

This class has a long tradition in public finance, starting from Feldstein (1969). More recently, Persson (1983) and Benabou (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.

We can see why τ is a natural index of progressivity in two ways. First, eq. (1) implies the following mapping between disposable (post-government) earnings $\tilde{y}_i$ and pre-government earnings y_i :

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

Thus, $(1-\tau)$ measures the elasticity of post-tax to pre-tax income.[2] Second, a tax scheme is commonly labeled progressive (regressive) if the ratio of marginal to average tax rates is larger (smaller) than one for every level of income y_i. Within our class, we have

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

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

The case $\tau = 0$ implies a ratio of one and yields a flat tax rate of $1 - \lambda$. When $\tau > 0$, the ratio in eq. (3) is larger than one and the tax system is therefore progressive. Conversely, when $\tau < 0$, the tax system is regressive.

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 government expenditure. Assuming a balanced budget, the average income-weighted marginal tax rate is then simply

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

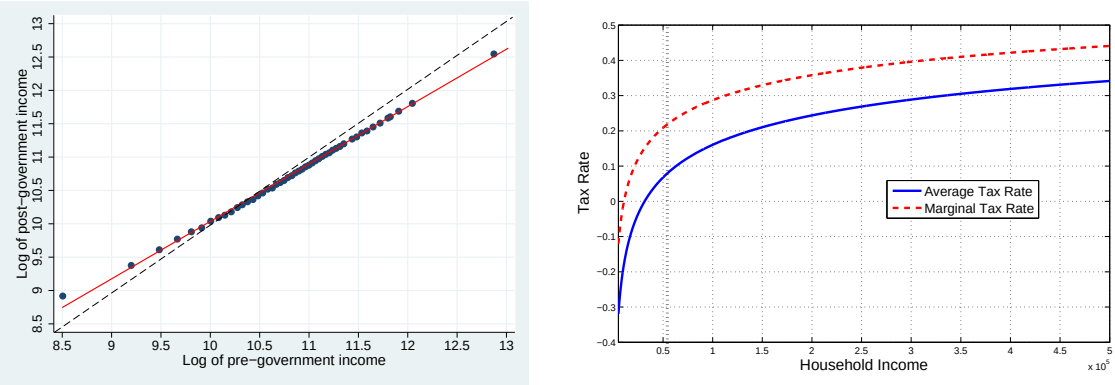
From eq. (4) it is immediate that when $g = 0$, the average income-weighted marginal tax rate is exactly τ.[3] Holding fixed g, the average marginal rate is increasing in τ. Holding fixed τ, the average marginal rate is increasing in g, since increasing net tax revenue while maintaining progressivity necessitates higher tax rates across the income distribution.

Empirical fit: We now demonstrate that this functional form offers a remarkably good representation of the actual tax/transfer scheme in the United States. We use data from the Panel Study of Income Dynamics (PSID) for survey years 2000, 2002, 2004, and 2006. We restrict attention to households aged 25-60 because we focus on labor income, and because we want to abstract from the intergenerational dimension of redistribution between the working-age population and retirees.[4] Pre-government household income includes labor earnings, private transfers

[3]Budget balance requires $gY = \int y_i - \lambda y_i^{1-\tau} di$. The income-weighted average marginal tax rate is then

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

[4]The rest of the sample selection criteria are the same as in Heathcote, Perri, and Violante (2010). In particular, we require a positive lower bound on annual hours worked (i.e., that either the head or the spouse works at least 260 hours per year or one quarter part-time) because we will estimate eq. (2) in log form. 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.



(a) Statistical fit on U.S. data (PSID 2000-2006)

(b) Implied average and marginal tax

Figure 1: Representation of the actual U.S. tax/transfer system through our tax/transfer function. The estimated value of τ^{US} is 0.151.

(transfers include alimony, child support, help from relatives, miscellaneous transfers, private retirement income, annuities, and other retirement income), and income from interests, dividends, and rents. Post-government income equals pre-government income minus federal and state income taxes computed using the NBER's TAXSIM program (Feenberg and Coutts, 1993), plus public transfers (AFDC/TANF, SSI and other welfare receipts, social security benefits, unemployment benefits, worker's compensation, and veterans' pensions).[5]

We estimate τ^{US} by least squares using eq. (2) in log form. The point estimate is $\tau^{US} = 0.151$ ($S.E. = 0.003$). The simple model fits the empirical relationship between pre- and post-government earnings distributions remarkably well: $R^2 = 0.96$. In Figure 1(a) we collapse our $13,721$ observations into 50 quantiles (each containing 2% of total observations).[6] Figure 1(b) plots the average and marginal tax rates implied by our tax/transfer scheme evaluated at τ^{US} (mean income is normalized to 1).[7] The implied income-weighted marginal tax rate is 0.33.

[5] In some instances, asset income is taxed differently from labor earnings. Because we cannot split observed taxes paid into taxes on earnings versus taxes on asset income, we estimate progressivity using total income as the tax base and total taxes as the tax take. The presence of asset income has minimal impact on our empirical estimates of progressivity, since asset income is very small in our sample. In part that is because we focus on households of working age, and in part it reflects the facts that the PSID undersamples the very rich, and –even conditional being interviewed– households grossly underreport asset income.

[6] The coordinates of each circle in the figure are the mean of the corresponding quantile of the pre-government income distribution (*x* axis), and the mean post-government income across the observations in that same quantile (*y* axis).

[7] Bakis, Kaymak, and Poschke (2013) combine CPS data with TAXSIM and obtain a value of $\tau^{US} = 0.17$ for a longer period, 1979-2009. Guner, Kaygusuz, and Ventura (2012a) estimate this same function on a large cross-

The PSID data have three potential limitations for the purposes of estimating progressivity: (i) the PSID undersamples the very rich, (ii) 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 (including all the top earners) of before-tax income.[8] The CBO measures of taxes and transfers are more comprehensive than those reported in the PSID. Their measure of taxes includes both employee- and employer-paid social insurance taxes, and their measure of transfers includes the value of Food Stamps vouchers, school lunches, housing and energy assistance, and benefits provided by Medicare and Medicaid. Moreover, the CBO adds to its measure of pre-government income employer-paid health insurance premiums, and the employer's share of social security and payroll taxes.

From the CBO tables we construct before and after government income for the first, second, third, and fourth quintiles of the before-government income distribution, and for the 81st-90th percentiles, the 91st-95th percentiles, the 96th-99th percentiles, and the top 1%. We used these moments to estimate the progressivity parameter τ^{US} for the 2000-2006 period and obtained $\tau^{US} = 0.155$, which is nearly identical to our PSID estimate for the same years.[9] Interestingly, the CBO data show an increase in progressivity during the Great Recession, with τ averaging 0.185 over the period $2008 - 2010$. Since the PSID is the data source we use to estimate other model parameters in Section 6, we will use the PSID-based estimate ($\tau^{US} = 0.151$) in our baseline analysis.

Discussion: One way to think about our exercise is as follows. We ask, within the tax system class that is currently in place, how much more or less progressive should taxes be, and what

sectional data set from U.S. Internal Revenue Service (the "Public Use Tax File"). They estimate a smaller value for progressivity because these data do not include any government transfers. The same caveat applies to the estimate in Chen and Guo (2011).

[8]The 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 Internal Revenue Service (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).

[9]The CBO reports statistics for households of all ages. To avoid conflating forced retirement saving and genuine intragenerational redistribution in our estimate for tax progressivity, we excluded social insurance taxes and Social Security and Medicare transfers from the CBO measures of taxes and transfers. If we do not exclude those items, we obtain a higher estimate for progressivity, $\tau^{US} = 0.232$. The key reason is that the income before taxes and transfers of retirees is low, but they receive large amounts of Social Security and Medicare transfers which makes the system look more progressive. However, in our view this higher estimate exaggerates true progressivity because a large portion of retirement transfers reflects taxes paid earlier in life and simply substitutes for private saving.

would the associated welfare gains be? Of course, although this functional form (eq. 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, if $\tau > 0$, or globally concave, if $\tau < 0$. As a result, marginal tax rates are monotonic in income. The same restriction applies to the average tax rate. Second, it does not allow for lump-sum transfers in cash, since $T(0) = 0$.[10]

Analyses of optimal tax design in the Mirrlees tradition often emphasize the importance of allowing for lump-sum transfers. Heathcote and Tsujiyama (2013) consider the welfare gains of moving from tax systems of the type described by eq. (1) to affine systems and to systems that do not impose any parametric restrictions on the shape of the tax schedule. Their environment is a stripped-down version of the model developed here. Under their baseline social welfare function, Heathcote and Tsujiyama find that the welfare gains of moving from the tax system described above with $\tau = 0.151$ to the constrained-efficient and fully nonparametric Mirrlees system are very small, on the order of 0.1 percent of consumption. The welfare gains of moving to the optimal system in the affine class are typically negative, indicating that allowing for a lump-sum transfer component in the tax system is less important than allowing for marginal tax rates to increase with income.

These findings suggest that restrictions implicit in the system described by eq. (1) may not be particularly important from a normative standpoint. Moreover, as will become clear, an important advantage of the functional form we use is that when we embed it in our structural equilibrium model, the model remains tractable, and the trade-offs from increasing or reducing progressivity are transparent. In addition, restricting attention to this functional form allows us to incorporate a range of model features that turn out to be quantitatively important in shaping optimal progressivity, including skill investment choices, persistent life-cycle uninsurable shocks, and preference heterogeneity. Conducting a Mirrlees-style optimal taxation exercise in this rich environment would be an extremely challenging numerical exercise.

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

3 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 into the next period with constant probability $\delta < 1$, and a cohort of newborn agents of size $(1-\delta)$ enters the economy. 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 her labor productivity z_i. Every period she supplies hours of work $h_i \geq 0$ to the market and consumes a private good c_i and a publicly provided good G.[11]

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

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

where $\theta > 1$ is the elasticity of substitution across skill types, $N(s)$ denotes average effective hours worked by skill type s, and $m(s)$ is the density of individuals with skill level s. In this baseline specification, all skill levels enter symmetrically in the production technology, and thus any equilibrium differences in skill prices will reflect relative scarcity in the context of a model in which different skill types are imperfect substitutes. In Section 7.4 we will consider an extension where the technology features different relative weights on different skill types, which introduces an additional (exogenous) driver for skill price differences.

The rate of transformation between private and public consumption is one, and thus the aggregate resource constraint for the economy is

$$Y = \int_0^1 c_i \, di + G. \tag{6}$$

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

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

[11] G has two possible interpretations. The first is that it is a pure public good, like national defense or the judicial system. The second is that it is an excludable good produced by the government and distributed uniformly across households, such as public health care or public transportation.

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 quadratic form

$$v_i(s_i) = -\frac{1}{\kappa_i}\frac{s_i^2}{2\mu}, \tag{8}$$

where $\kappa_i \geq 0$ is a parameter, heterogeneous across individuals, which 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 $\kappa_i \sim Exp(\eta)$, an exponential distribution with parameter η. The parameter μ is a scaling constant. As we demonstrate below, the combination of quadratic skill investment costs and exponentially distributed ability yields Pareto right tails in the wage and earnings distributions.

The period utility function u_i is specified as

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

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. 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

$$\frac{1}{\hat{\sigma}} = \frac{1-\tau}{\sigma+\tau}, \tag{10}$$

which measures the after-tax elasticity of hours worked to a transitory wage shock.[12] Finally, $\chi \geq 0$ measures the taste for the publicly provided good G relative to private consumption.[13]

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

$$\log z_{ia} = \alpha_{ia} + \varepsilon_{ia}. \tag{11}$$

[12]We abstract from the extensive margin of labor supply decisions, especially relevant for the second earner in the household. See Guner, Kaygusuz, and Ventura (2012b) for a recent analysis of the effects of tax reforms on the joint labor supply decisions of married households.

[13]Note that the model is essentially unchanged if a fixed fraction of public expenditure is wasted, so that only a fraction is delivered to consumers. Given logarithmic utility from G, this amounts to adding an irrelevant constant to preferences.

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 $\alpha_{i0} = 0, \forall i$.[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).[15] A law of large numbers (e.g., Uhlig, 1996) implies that individual-level shocks induce no aggregate uncertainty in the economy as a whole.

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

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

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; the third component is the number of hours worked by the individual. Eq. (12) shows the determinants of individual earnings: (i) skills accumulated before labor market entry, in turn reflecting innate idiosyncratic learning ability κ; (ii) fortune in labor market outcomes determined by the realization of idiosyncratic efficiency shocks; and (iii) work effort, reflecting, in part, innate idiosyncratic diligence measured by (the inverse of) φ.

Because idiosyncratic labor productivity is exogenous, the two channels via which taxation will impact the equilibrium pre-tax earnings distribution are by changing skill investment choices, and thus skill prices, and by changing labor supply decisions.

Financial assets: We adopt the partial-insurance structure developed in Heathcote et al. (2014) and assume that there are only two types of financial assets in the economy. The first is a non-state-contingent bond b with price q. The second is a full set of insurance claims against the ε shock. Thus, by assumption the ε shocks are fully insurable, whereas the α shocks can potentially be smoothed only by borrowing and lending via the risk-free bond. 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

[14] Setting the dispersion in initial conditions α_{i0} to zero does not mean that there is no initial inequality in productivity. Recall that newborn agents enter the economy with heterogeneous skill levels s_i (also a fixed individual effect), reflecting the dispersion in innate learning ability κ_i.

[15] The empirical autocovariance function for individual wages displays a sharp decline at the first lag, indicating the presence of a transitory component in wages. At the same time, within-cohort wage dispersion increases approximately linearly with age, suggesting the presence of permanent shocks.

if and only if $\varepsilon \in \mathbf{E} \subseteq E$. Our model nests several market structures. First, when $v_\omega = 0$, the economy displays full insurance. When $v_\varepsilon = 0$, it is a bond economy, as in Huggett (1993). In general, when $v_\omega > 0$ and $v_\varepsilon > 0$, ours is a *partial insurance* economy, i.e., an economy that offers more insurance opportunities than a bond economy but less insurance than complete markets.[16] In our framework, greater progressivity reduces the equilibrium demand for insurance, for two reasons. First, public redistribution directly substitutes for private insurance. Second, progressivity dampens the response of hours, and thus earnings, to insurable shocks, and hence reduces pre-tax earnings inequality.[17]

Finally, for convenience, we assume that there exist actuarially fair annuities against survival risk. All assets in the economy are in zero net supply, and newborn agents start with zero initial wealth. There are no intergenerational links in our model.[18]

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

Government: The government runs the tax/transfer scheme described in Section 2 and provides each household with an amount of goods or services equal to G. Without loss of generality, we assume that government expenditures are a fraction g of aggregate output, i.e., $G = gY$. Since we abstract from public debt, the government budget constraint holds period by period and reads as

$$g \int_0^1 y_i \, di = \int_0^1 \left(y_i - \lambda y_i^{1-\tau} \right) \, di. \tag{13}$$

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

[16] The 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 et al. (2014), 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 i.i.d.

[17] Note that because the extent of risk-sharing is exogenous in the model, making the tax system more progressive does not affect the *supply* of private insurance. In contrast, public insurance can crowd out private insurance in environments featuring moral hazard or limited commitment such as Chetty and Saez (2010) or Krueger and Perri (2010).

[18] Bakis et al. (2013) argue that private bequests provide a form of insurance against a bad draw of initial conditions (κ, φ), which diminishes the redistributive role of taxation and reduces optimal progressivity.

3.1 Agent's problem

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

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

Thus, the marginal disutility of skill investment for an individual with learning ability κ_i must equal the discounted present value of expected benefits from the skill investment.

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 c_{ia} and bond holdings $b_{i,a+1}$ for next period.

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

$$\int_E Q(\varepsilon)\, B(\varepsilon)\, d\varepsilon = b_{ia}, \tag{15}$$

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

$$c_{ia} + \delta q b_{i,a+1} = \lambda\left[p(s_i)\exp\left(\alpha_{ia}+\varepsilon_{ia}\right)h_{ia}\right]^{1-\tau} + B(\varepsilon_{ia}), \tag{16}$$

where the δ pre-multiplying the bond price reflects the return on the annuity for survivors.

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 eq. (11). In addition, agents face limits on borrowing that rule out Ponzi schemes and non negativity constraints on consumption and hours worked.

3.1.1 A special case: the representative agent problem

It is useful to solve for a special case of the agent's problem. When $v_\varphi = v_\omega = v_\varepsilon = 0$ and $\theta = \infty$, there is no 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:

$$\max_{C,H} \left\{ \log C - \frac{H^{1+\sigma}}{1+\sigma} + \chi \log G \right\} \tag{17}$$
$$s.t.$$
$$C = \lambda H^{1-\tau},$$

and the production technology simplifies to $Y = H$, implying $G = gH$. Taking the fiscal variables (λ, g, τ) as given, the optimal choices for the representative agent are

$$\log H^{RA}(\tau) = \frac{1}{1+\sigma} \log(1-\tau), \tag{18}$$
$$\log C^{RA}(g,\tau) = \log \lambda(g,\tau) + \frac{1-\tau}{1+\sigma} \log(1-\tau). \tag{19}$$

And substituting $\lambda(g,\tau)$ from the government budget $G = H - \lambda H^{1-\tau}$ into eq. (19) gives

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

These allocations show that a more progressive tax system (a higher value for τ) reduces labor supply and therefore reduces equilibrium consumption. The intuition is that higher progressivity raises the marginthe al tax rate faced by representative agent. In the limit, as $\tau \to 1$, $H^{RA}(\tau) \to 0$. Note that, with logarithmic utility, the average level of taxation (λ) has no impact on labor supply, which explains why hours worked (and output) are independent of the level of expenditures g.

4 Equilibrium

We now adopt a recursive formulation to define a stationary competitive equilibrium for our economy. The state vector for the beginning-of-the-period decision when insurance claims are purchased is (φ, α, s, b). The individual state vector for the end-of-period consumption/saving and labor supply decisions is $\left(\varphi, \alpha, \varepsilon, s, \bar{B}\right)$, where $\bar{B} = B(\varepsilon; \varphi, \alpha, s, b)$. Finally, since initial wealth is zero, the state vector for the skill accumulation decision at age $a = 0$ reduces to the pair of fixed individual effects (κ, φ).

Given (g, τ), a *stationary recursive competitive equilibrium* for our economy is a tax parameter λ, asset prices $Q(\cdot)$ and q, skill prices $p(s)$, decision rules $s(\kappa, \varphi)$, $c\left(\varphi, \alpha, \varepsilon, s, \bar{B}\right)$, $h\left(\varphi, \alpha, \varepsilon, s, \bar{B}\right)$, $b'\left(\varphi, \alpha, \varepsilon, s, \bar{B}\right)$, and $B(\cdot; \varphi, \alpha, s, b)$, and aggregate quantities $N(s)$ such that

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

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

$$p(s) = \left(\frac{Y}{N(s) \cdot m(s)}\right)^{\frac{1}{\theta}}.$$

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

4. The government budget is balanced: λ satisfies eq. (13).

Proposition 1 [competitive equilibrium]. *There exists a competitive equilibrium characterized by no bond trading across individuals, i.e.,* $b'(\varphi, \alpha, \varepsilon, s, B(\varepsilon; \varphi, \alpha, s, 0)) = 0$ *for all* $(\varphi, \alpha, \varepsilon, s)$. *The interest rate* $r^* \equiv -\log q$ *that supports this equilibrium satisfies*

$$\rho - r^* = (1-\tau)\left[(1-\tau)+1\right]\frac{v_\omega}{2}, \tag{20}$$

where $\rho \equiv -\log \beta$ *is the agents' discount rate.*

The proof for Proposition 1 in the Appendix is based on a guess and verify strategy. We first guess that the bond is not traded and solve for the equilibrium consumption allocation. Next, we use the consumption allocation to construct the expected marginal rate of substitution and show that it is independent of any individual state. Thus, at the interest rate that clears the bond market, all agents are indifferent between borrowing and lending on the margin and are thus content to maintain a zero bond position.[19]

In equilibrium, the intertemporal dissaving motive (the left-hand side of equation 20) determined by the gap between ρ and r^* exactly equals the precautionary saving motive (the right-hand side), which is increasing in the size of the uninsurable wage risk v_ω and decreasing in the progressivity parameter τ. The logic is that as τ rises, the government provides more social insurance and the private precautionary demand for savings falls.

[19] As discussed in Heathcote et al. (2014) Section 2.3.2, this result is a generalization of the insight in Constantinides and Duffie (1996). Here, we further generalize by endogenizing the wage through the skill investment decision and a technology featuring imperfect substitutability across skill types.

Proposition 1 has two implications that are instrumental for analytical tractability. First, individual wealth is a redundant state variable: individuals start their life with zero wealth and remain with zero wealth forever. All remaining individual states are exogenous variables.[20] Second, there is no self-insurance via noncontingent borrowing and lending against α shocks. In contrast, there is perfect insurance, by assumption, against ε shocks. Thus, in equilibrium, there is a dichotomy between one type of risk that is uninsured and another that is fully insured. We use the label "uninsurable" to denote the α shock (and its innovations ω) and the label "insurable" to denote the ε shock.

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

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

$$\log h\left(\varphi, \varepsilon; \tau\right) = \log H^{RA}\left(\tau\right) - \varphi + \frac{1}{\widehat{\sigma}}\varepsilon - \frac{1}{\widehat{\sigma}(1-\tau)}\mathcal{M}\left(v_{\varepsilon}; \tau\right), \tag{21}$$

where H^{RA} are hours worked by the "representative agent" in eq. (18) and $\mathcal{M}\left(v_{\varepsilon}; \tau\right) = \frac{(1-\tau)(1-\tau(1+\widehat{\sigma}))}{\widehat{\sigma}}\frac{v_{\varepsilon}}{2}$.

The consumption allocation is given by

$$\log c\left(\varphi, \alpha, s; g, \tau\right) = \log\left[C^{RA}\left(g, \tau\right)\vartheta(\tau)\right] + (1-\tau)\left[\log p\left(s; \tau\right) + \alpha - \varphi\right] + \mathcal{M}\left(v_{\varepsilon}; \tau\right), \tag{22}$$

where C^{RA} is consumption of the "representative agent" in eq. (19) and $\vartheta(\tau)$ is a constant.

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

[20]The skill level is endogenous but fixed after age zero, and is hence pre-determined with respect to consumption and labor supply decisions.

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 illustrated 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 response of hours to insurable wage shocks.

The consumption allocation is additive in five separate components. The first component is (rescaled) consumption of the representative agent, described in Section 3.1.1.[21] Consumption is increasing in 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 initial conditions s and φ and from ex post 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 3 [skill price and skill choice]. *In equilibrium, skill prices are given by*

$$\log p\left(s;\tau\right)=\pi_0\left(\tau\right)+\pi_1\left(\tau\right)\cdot s\left(\kappa;\tau\right), \tag{23}$$

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

The skill investment allocation is given by

$$s\left(\kappa;\tau\right)=\mu\left(1-\tau\right)\pi_1\left(\tau\right)\cdot\kappa=\sqrt{\frac{\eta\mu\left(1-\tau\right)}{\theta}}\cdot\kappa \tag{24}$$

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

This proposition has a number of important implications. First, the log of the equilibrium skill price has a "Mincerian" shape, i.e., it is an affine function of s. The constant $\pi_0(\tau)$ is the base

[21] The 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. If we had specified (λ,τ) as the fiscal policy instruments, and g as the residual variable, this constant would drop out.

log-price of the lowest skill level $(s = 0)$ and $\pi_1(\tau)$ is the marginal return to skill. As is evident from (24), a higher value for τ (more progressivity) depresses skill investment and compresses the skill distribution toward zero. In the limit as $\tau \to 1$, $s(\kappa) \to 0$: there is no incentive to invest in higher wage skills if all the excess returns will be taxed away. Because of imperfect substitution in production, a rise in the relative scarcity of high skill types increases the marginal return $\pi_1(\tau)$ and reduces the base price $\pi_0(\tau)$. Thus, our model features a "Stiglitz effect" (Stiglitz, 1985): progressivity increases the equilibrium marginal return to skill investment.

With the solution for the skill price and the consumption allocation in hand, the expression for the skill choice s is easy to understand. Substituting the period-utility specification in (9), the consumption allocation in (22), and the skill price in (23) into the first-order condition (14), expression (24) follows immediately. Note that, holding fixed the skill premium, $\pi_1(\tau)$, the partial equilibrium elasticity of skill investment with respect to $(1 - \tau)$ is exactly unity. Taking into account the equilibrium response of the skill premium, the general equilibrium elasticity of skill investment with respect to $(1 - \tau)$ is only one-half.

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 implied log change in wages, which is independent of the *level* of wages or hours.

Corollary 3.1 [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*

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

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 23) and skills retain the exponential shape of the distribution of learning ability κ (equation 24). 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 and the variance of log skill prices is independent of τ.

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).

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

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

$$H(\tau) = \mathbb{E}[h(\varphi,\varepsilon;\tau)] = (1-\tau)^{\frac{1}{1+\sigma}} \cdot \exp\left[\left(\frac{\tau(1+\hat{\sigma})}{\hat{\sigma}^2} - \frac{1}{\hat{\sigma}}\right)\frac{v_\varepsilon}{2}\right], \tag{25}$$

$$N(\tau) = \mathbb{E}[\exp(\alpha+\varepsilon)h(\varphi,\varepsilon;\tau)] = H(\tau)\cdot\exp\left(\frac{1}{\hat{\sigma}}v_\varepsilon\right). \tag{26}$$

Output is given by

$$Y(\tau) = \mathbb{E}[p(s;\tau)\exp(\alpha+\varepsilon)h(\varphi,\varepsilon;\tau)] = \mathbb{E}[p(s;\tau)]\,N(\tau), \tag{27}$$

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

Aggregate labor productivity is

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

Note that progressivity affects aggregate output through two channels: τ affects the average skill price (wage) via its impact on skill investment choices, and τ affects average hours worked. From eq. (26) the elasticity of aggregate hours $N(\tau)$ with respect to τ at $\tau = 0$ is $\frac{-1}{1+\sigma}$. From eq. (23) the elasticity of the average skill price $\pi_0(\tau)$ is $\frac{-1}{2(\theta-1)}$. We will return to these two elasticities in Section 5.3 when characterizing the conditions under which the optimal tax system is progressive.

Before turning to the characterization of the optimal degree of progressivity, we briefly discuss the efficiency of the competitive equilibrium. The competitive 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 there is an externality in the individual labor supply decision when the publicly provided good is valued: agents do not internalize that, by working more hours, the quantity of public good will increase. Proposition 6 in Section 6.3 states that when the economy features complete markets with respect to wage shocks ($v_\omega = 0$) and does not feature any public-good externality ($\chi = 0$), the equilibrium with $\tau = 0$ is efficient.

5 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 7.3 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 initial and final steady state, and the motive to tax the existing stock of skills affects the optimal choice of progressivity.

5.1 Social welfare function

We will evaluate alternative policies using a social welfare function according to which the government puts equal weight on all agents within a given cohort. Given a law of large numbers, it follows that the contribution to social welfare from any given cohort is remaining expected lifetime utility for that cohort, where eq. (7) defines expected lifetime utility at age zero. The overlapping

generations structure of the model also 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 lifetime utility of future generations at rate γ. Social welfare evaluated as of date 0 is then given by

$$\mathcal{W}\left(g,\tau;\tau_{-1}\right)\equiv(1-\gamma)\Gamma\sum_{j=-\infty}^{\infty}\gamma^{j}U_{j,0}\left(g,\tau;\tau_{-1}\right), \tag{28}$$

where $U_{j,0}\left(g,\tau;\tau_{-1}\right)$ is remaining expected lifetime utility (discounted back to date of birth) for the cohort that entered the economy at date j as of date 0.[22] The constant $\Gamma=\frac{(\gamma-\beta\delta)}{\gamma(1-\beta\delta)}$ pre-multiplying the summation is a convenient normalization.[23]

The next proposition expresses social welfare as an explicit function of the two policy instruments (g,τ) for the baseline model with fully reversible investment.

Proposition 4 [closed-form social welfare]. *In the model with fully reversible investment, when the social welfare function is given by eq. (28), social welfare from implementing policy (g,τ) at date 0 is*

$$\begin{aligned}\mathcal{W}(g,\tau;\tau_{-1}) &= \log(1-g)+\chi\log g+(1+\chi)\frac{\log(1-\tau)}{(1+\hat{\sigma})(1-\tau)}-\frac{1}{(1+\hat{\sigma})} \\ &+(1+\chi)\left[\frac{\log\left(1-\tau\right)}{2\left(\theta-1\right)}+\frac{1}{2\left(\theta-1\right)}\log\left(\frac{\mu}{\eta}\left(\theta-1\right)^{-2\theta}\theta^{(2\theta-1)}\right)\right] \\ &-\frac{1}{2\theta}(1-\tau)+\frac{\beta\delta}{\gamma}\frac{(1-\gamma)}{(1-\beta\delta)}\frac{1}{2\theta}(1-\tau_{-1}) \\ &-\left[-\log\left(1-\left(\frac{1-\tau}{\theta}\right)\right)-\left(\frac{1-\tau}{\theta}\right)\right] \\ &-\left(1-\tau\right)^{2}\frac{v_{\varphi}}{2} \\ &-\left[\left(1-\tau\right)\frac{\beta\delta}{\gamma-\beta\delta}\frac{v_{\omega}}{2}-\log\left(\frac{1-\delta\exp\left(\frac{-\tau(1-\tau)}{2}v_{\omega}\right)}{1-\delta}\right)\right] \\ &+(1+\chi)\left[\frac{1}{\hat{\sigma}}v_{\varepsilon}-\sigma\frac{1}{\hat{\sigma}^{2}}\frac{v_{\varepsilon}}{2}\right].\end{aligned} \tag{29}$$

[22] Remaining lifetime utility is independent of the previous value for public good provision g_{-1}. It 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.

[23] Following 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 given date on the felicity of one agent who is s years older than another is $(\beta/\gamma)^{s}$.

In order to obtain the expression in eq. (29), one has to first solve for the value $\lambda(g,\tau)$ that balances the government budget. Next, plugging the consumption, hours, and skill allocations into (31), one obtains the expression for social welfare as a function of primitive preference, technology, and policy parameters in eq. (29).

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

In eq. (29) 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 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 $\sigma \geq 2$, it is also globally concave in τ.*

As we show in the Appendix, aside from the term multiplying v_ε in the last row of (29), the social welfare expression 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.[24] Establishing concavity is useful since it means that a first-order approach can be useful in computing the welfare-maximizing values for τ and g.

Corollary 4.3: [Samuelson condition] *The welfare-maximizing value for g is given by*

$$g^* = \frac{\chi}{1+\chi}. \tag{30}$$

Because g only appears in the first line of eq. (29), the optimal choice for public expenditures in the economy can easily be derived by differentiating (29) with respect to g. In our economy, the optimal fraction of output to devote to public expenditure is independent of how much inequality there is in the economy and independent of how progressive the tax system is. It only depends on households' relative taste for the public good χ. We label this the "Samuelson condition" because, as we show in Section 5.2.1, it is reminiscent of Samuelson's (1954) dictum that, in a representative agent model, a government choosing public spending should equate the marginal

[24] See the Appendix for a more thorough interpretation of sufficient conditions.

rate of substitution between private and public consumption to the marginal rate of transformation between the two goods, which is equal to one.

Corollary 4.4: [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 eq. (29). Thus, the welfare-maximizing choice for τ will always be independent of the value for g and in particular independent 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 4.5: [irrelevance of μ and η] *The welfare-maximizing value for progressivity τ is independent of the parameters μ and η.*

The welfare-maximizing value for progressivity τ is independent of the weight on the disutility of skill investment μ and the exponential parameter in the distribution for skill investment costs η since these two parameters only appear in an additively separable constant in the second line of the social welfare expression (29). Thus, the only parameter that will matter in determining how skill investment shapes optimal progressivity is θ, the elasticity of substitution between different skill types in production.

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 utility in the cross section*

$$\mathcal{W}(g,\tau) = (1-\delta)\sum_{j=0}^{\infty}\delta^j \mathbb{E}\left[u\left(c\left(\varphi,\alpha_j,s(\kappa;\tau);g,\tau\right),h\left(\varphi,\varepsilon;\tau\right),G\left(g,\tau\right)\right)\right] - \mathbb{E}\left[v\left(s(\kappa;\tau),\kappa\right)\right], \tag{31}$$

where the first expectation is taken with respect to the equilibrium cross-sectional distribution of $(\varphi, \alpha_j, s, \varepsilon)$ and the second expectation with respect to the cross-sectional distribution of (s, κ).[25]

Thus, when the planner values cohorts entering at successive dates at the same rate that agents discount over time, the optimal policy (g, τ) is simply the policy that maximizes average period

[25]The uninsurable shock α_j is indexed by age j because the conditional variance of the unit root process α depends on age j.

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

5.2 Decomposition of the social welfare function

We now demonstrate that every term in (29) has an economic interpretation and captures one of the forces that determine 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 case $\gamma = \beta$.

5.2.1 Welfare of the representative agent

Substituting allocations (18) and (19) into the objective function of the representative agent problem of Section 3.1.1 (after solving for λ), one obtains welfare for the representative agent,

$$\mathcal{W}^{RA}(g,\tau) = \log(1-g) + \chi \log g + (1+\chi)\frac{\log(1-\tau)}{(1+\hat{\sigma})(1-\tau)} - \frac{1}{(1+\hat{\sigma})}, \tag{32}$$

which is precisely the first line of the social welfare expression in (29).

What does this welfare expression imply for optimal policy? Differentiation of (32) with respect to g yields the Samuelson condition, eq. (30). This value for g equates the marginal rate of substitution between private and public consumption for the representative agent (which is equal to $\chi g/(1-g)$) to the technological rate of transformation between the two goods (which is equal to one).

Differentiation of (32) with respect to τ yields

$$\tau^*_{RA} = -\chi. \tag{33}$$

Eq. (33) states that a benevolent government in the representative agent economy would choose *regressive* taxes, with the extent of regressivity proportional to the relative taste for the public good. The logic is that because there are no private markets for the publicly produced good, there is an externality. Each individual agent does not internalize that, by working more and producing more, output will increase and the government will be able to provide more G. A regressive tax increases labor supply, as is clear from the hours allocation (18), and *entirely* corrects this externality, a result we restate in the following corollary.[26]

[26] If there was a private market for G, and G was a nonrival nonexcludable good, then the welfare theorems would apply and the first best could be implemented with $\lambda = 1$ and $\tau = 0$.

Corollary 4.7 [efficiency in the RA model]. *If $v_\varphi = v_\omega = v_\varepsilon = 0$ and $\theta = \infty$ (i.e., in a representative agent economy), then $g^* = \frac{\chi}{1+\chi}$ and $\tau^* = -\chi$ implement the first best.*

To gain some intuition, note that one could alternatively implement the first best with lump-sum taxes that do not distort labor supply and finance the desired amount of G. Given $\tau = -\chi$, the marginal tax rate at the equilibrium level of income Y is exactly equal to zero (just as it would be with lump-sum taxes), whereas given $g = \chi/(1+\chi)$, the average tax rate is exactly sufficient to finance the optimal level of expenditure. Thus, the system perfectly replicates a lump-sum tax.

5.2.2 Welfare from skill investment

The second, third, and fourth lines in eq. (29) are all related to the skill investment choice. To begin with, from equations (23), (26), and (27), the term in square brackets in the second line is the log of aggregate productivity (output per efficiency unit of labor) in the economy, also equal to $\log \mathbb{E}\left[p\left(s;\tau\right)\right]$:

$$\log\left(\frac{Y(\tau)}{N(\tau)}\right) = \frac{\log\left(1-\tau\right)}{2\left(\theta-1\right)} + \frac{1}{2\left(\theta-1\right)}\log\left(\frac{\mu}{\eta}\left(\theta-1\right)^{-2\theta}\theta^{(2\theta-1)}\right). \tag{34}$$

The pattern of skill investments determines aggregate productivity through the constant elasticity of substitution (CES) technology: the more evenly distributed are skills, the higher is productivity. Eq. (34) indicates that higher progressivity reduces productivity because it reduces skill investment and thereby compresses the skill distribution toward zero. Note that, in terms of its impact on social welfare, productivity is multiplied by $(1+\chi)$ because the higher the desire for public good, the more valuable is an additional unit of output.

Skill investment is not costless. The third line of (29) is the contribution to welfare from skill investment costs. The average skill investment cost for all the future cohorts and for the past cohorts who readjust their skill level is

$$\mathbb{E}\left[v\left(s(\kappa;\tau),\kappa\right)\right] = \frac{1}{2\theta}(1-\tau), \tag{35}$$

which is the first term on the third line.[27] Skill investment costs are decreasing in τ because more progressivity reduces skill acquisition. Combining (34) and (35), it is easy to show that the pro-

[27] If older cohorts already have some skills at the time of the tax reform, they only need to pay net new investment costs of $\frac{1}{2\kappa_i\mu}\left[s\left(\kappa_i;\tau\right)^2 - s\left(\kappa_i;\tau_{-1}\right)^2\right]$, which accounts for the second seperable term in τ_{-1} in the welfare expression. See the Appendix for details.

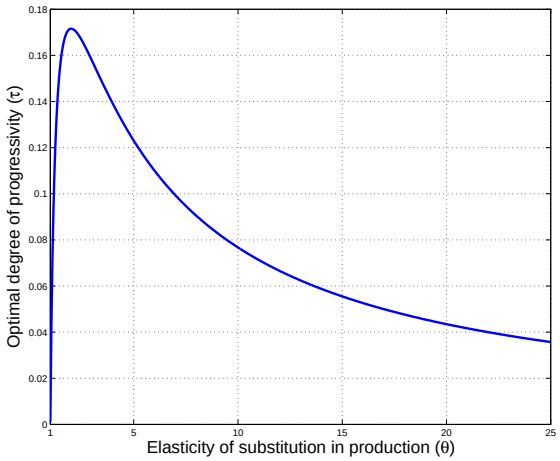


Figure 2: Optimal τ as a function of the elasticity of substitution across skill levels in production (θ)

ductivity gain from skill investment net of education costs is maximized at $\tau = -1/(\theta - 1) < 0$: a regressive system invites larger skill investments and the stronger is complementarity in production, the stronger is this force.

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

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

or the term in the fourth line of (29).[28] This cost is decreasing in τ because higher progressivity reduces after-tax earnings and consumption dispersion.

We have learned that offsetting forces determine the optimal level of progressivity with respect to skill acquisition: more progressivity diminishes aggregate productivity, but it also decreases consumption dispersion across skill types. Which force dominates? In Figure 2 we set $\chi = 0$,

[28]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. Let F_c denote the Pareto cumulative distribution function (CDF) for 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 percentage by which safe consumption must be reduced to deliver the same expected utility as risky consumption: $\log((1-\varpi)P/(P-1)) = \int \log c \, dF_c = 1/P$. Now, since in our example $P = \theta/(1-\tau)$ and ϖ is small, $\varpi \approx -\frac{1-\tau}{\theta} - \log\left(\frac{\theta-(1-\tau)}{\theta}\right)$ as in eq. (36).

$\sigma = \infty$, and $v_\varphi = v_\omega = v_\varepsilon = 0$ to isolate the skill investment channel and show how τ^* varies with θ. The figure reveals that the skill investment component of social welfare calls for progressivity ($\tau^* > 0$), a result we prove formally in Section 5.3. At the same time, the optimal degree of progressivity is nonmonotone in θ. As $\theta \to \infty$, the economy converges to a representative agent economy, so $\tau^* = 0$. For lower values for θ, the utilitarian government chooses $\tau^* > 0$ to reduce the welfare loss from consumption dispersion. However, as $\theta \to 1$ and skill complementarity increases in production, the distortion to aggregate productivity from progressivity becomes more important – because of the term $\frac{1}{2(\theta-1)} \log(1-\tau)$ in (34) – and τ^* falls. At the same time, a regressive tax scheme, which would raise productivity, would make consumption inequality explode: eq. (36) shows that, as $\theta \to 1$, the term $-\log(\theta - 1 + \tau)$ goes to $-\infty$ as τ approaches zero. Overall, these two forces balance out, and as $\theta \to 1$, $\tau^* \to 0$ – a flat tax system. Thus, the optimal degree of progressivity is largest for intermediate values for θ.

5.2.3 Welfare from preference heterogeneity and uninsurable wage risk

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

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

This term is the familiar Lucas representation of the welfare cost of consumption dispersion when the underlying shocks are lognormal: one-half of the variance of log consumption times the coefficient of risk aversion, which is equal to one 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 eq. (29):

$$\begin{aligned} \text{welf. cost of } var_\alpha(\log c) &= \left[(1-\tau)\frac{\beta\delta}{\gamma-\beta\delta}\frac{v_\omega}{2} - \log\left(\frac{1-\delta\exp\left(\frac{-\tau(1-\tau)}{2}v_\omega\right)}{1-\delta}\right)\right] \\ &\simeq \frac{1}{2}\cdot var_\alpha(\log c) \text{ when } \gamma = \beta, \end{aligned} \tag{38}$$

where $var_\alpha(\log c)$ denotes within-skill group variance of log consumption in cross section. The approximation in the second line of eq. (38) is extremely accurate for plausible parameter values, as

we show in the proof of Proposition 4 in the Appendix. As can be seen from (37) and (38), a higher τ reduces consumption dispersion stemming from both preference heterogeneity and uninsurable risk. Since consumption inequality lowers welfare, these two forces push the optimal τ towards one, the value at which there would be zero consumption dispersion.

5.2.4 Welfare from insurable wage risk

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

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

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

The first term is the log productivity gain from insurable wage variation. As explained when discussing the equilibrium allocations, more insurable wage dispersion is good news for welfare, because individual hours worked become more positively correlated with individual productivity and aggregate output increases. Hours dispersion is, however, costly in welfare terms because of the convexity in the disutility of hours. This cost is captured by the last term in the welfare expression, which is the cross-sectional variance of log hours due to insurable shocks multiplied by σ, which measures the aversion to hours fluctuations.[29]

The sum of these two terms (productivity gain net of the disutility costs of hours fluctuations) is maximized at $\tau = 0$ because hours respond efficiently to insurable shocks with $\tau = 0$, whereas $\tau \neq 0$ induces misallocation in hours worked. Thus, larger insurable wage risk will push the optimal τ toward zero.[30]

5.3 When should taxes be progressive?

By differentiating the expression for social welfare in eq. (29) with respect to τ, one can obtain a necessary and sufficient parametric condition for the optimal tax system to be progressive.

[29] As 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.

[30] This finding is different from the result in Chetty and Saez (2010) that the standard Mirrlees-style formulas for the optimal tax schedule are unaffected by the presence of private insurance that does not generate moral hazard. The main reason for this difference is that Chetty and Saez (2010) assume that the government can tax earnings net of private insurance payments whereas we assume that taxes are levied on labor income only.

Proposition 5 [condition for progressivity]. *The optimal value for τ is strictly positive if and only if*

$$\frac{1}{2}\left(\frac{1}{\theta-1}-\frac{1}{\theta}\right)+(v_\varphi+v_\alpha)>\chi\left(\frac{1}{2(\theta-1)}+\frac{1}{1+\sigma}\right). \tag{41}$$

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.[31] The first term on the left-hand side is the gain from reducing consumption inequality across skill types net of the productivity loss from reduced skill investment (minus its costs). The second term is the gain from reducing consumption dispersion across agents with different preferences and labor productivity.

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

5.4 Optimal marginal tax rate at the top

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

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

where ζ^u and ζ^c are uncompensated and compensated labor supply elasticities.[32] Thus, as Saez and others have noted, the thicker is the right tail of the productivity distribution (i.e., the smaller is the Pareto coefficient θ), the higher is the optimal marginal tax rate at the top.

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

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

Our model with $\chi = 0$ and $v_\varphi = v_\omega = v_\varepsilon = 0$ is a version of the Mirrlees environment, but with one important difference: the Pareto distribution for labor productivity is *endogenously* determined by skill investment, and its Pareto parameter is the elasticity of substitution across skill types in production. As we showed in Section 5.2.2, the optimal choice for τ is nonmonotonic in θ.[33] In particular, there is a range of values for θ close to unity in which reducing θ (and increasing inequality) lowers optimal progressivity, and thus marginal tax rates at high income levels, in sharp contrast to the familiar Mirrlees result.[34] Recall the logic for this result: the more complementary are skill types, the larger are the productivity gains from a more even skill distribution. Thus, the more costly are high tax rates at the top that discourage skill investment of high learning ability (high κ) individuals. We conclude that whether the Pareto right tail in the earnings distribution reflects exogenous luck or endogenous investments is quantitatively important for determining optimal top marginal tax rates.

6 Quantitative implications of the theory

After describing the model parametrization, we explore the quantitative implications of the theory. Next, we perform a robustness analysis with respect to (i) the weights used by the government in its social welfare function and (ii) the assumption that government expenditures are valued.

6.1 Parametrization

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 enable us to prove identification of all parameters and to estimate the model given empirical counterparts of these moments computed from commonly used micro-data on wages, hours worked, and consumption.

We begin by recognizing that in survey data, hours and consumption are measured with error,

[33] Figure 2 is drawn for the case $\sigma = \infty$, but the nonmonotonicity result is more general and applies with elastic labor supply.

[34] Strictly speaking, for any $\tau > 0$, marginal tax rates converge to one under our functional form for taxes as earnings go to infinity. But from eq. (3), it is easy to show that marginal tax rates are strictly increasing in progressivity for any $y \in (\exp(-1/(1-\tau)), \infty)$.

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 (21) and (22), and compute cross-sectional moments of their joint distribution, we obtain the following set of moment conditions:

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

These moments contain all the structural parameters of the model. The variance of the uninsurable innovation v_ω is implied by v_α, given a value for δ.[35]

Based on our previous work (Heathcote et al., 2014), we set $\sigma = 2.165$, 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 evidence on the progressivity of the U.S. tax/transfer system described in Section 2, we set $\tau = 0.151$.

It is easy to see that v_φ, v_ε, and $\left(v_\alpha + 1/\theta^2\right)$ 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 $var^0\left(\log w\right)$, $var^0\left(\log c\right)$, and $cov^0\left(\log w, \log c\right)$ at age $j = 0$, which reflect only variation in skills acquired before labor market entry, since $v_\alpha^0 = 0$. In Section 6.2.1 we pursue an alternative strategy for separating v_α and θ.

Our data are drawn from two surveys, the Panel Study of Income Dynamics (PSID) for years

[35]The variance of α at age j is $v_\alpha^j = j v_\omega$ so the cross-sectional uninsurable variance in the model is

$$v_\alpha = (1-\delta)\sum_{a=0}^{\infty} \delta^a v_\alpha^a = \frac{\delta}{1-\delta} v_\omega.$$

Table 1: Parametrization

	τ^{US}	χ	σ	$v_{\mu h}$	$v_{\mu y}$	$v_{\mu c}$	v_{φ}	v_{ε}	v_{α}	θ	v_{ω}
Baseline	0.151	0.233	2.165	0.036	0	0.040	0.035	0.166	0.097	3.144	0.003
	(0.003)	–	–	–	–	–	(0.002)	(0.010)	(0.007)	(0.115)	(0.0003)
Alternative	0.151	0.233	2.165	0.036	0	0.040	0.021	0.139	0	2.000	0
	(0.003)	–	–	–	–	–	(0.002)	(0.010)	(0.0000)	–	(0.0000)

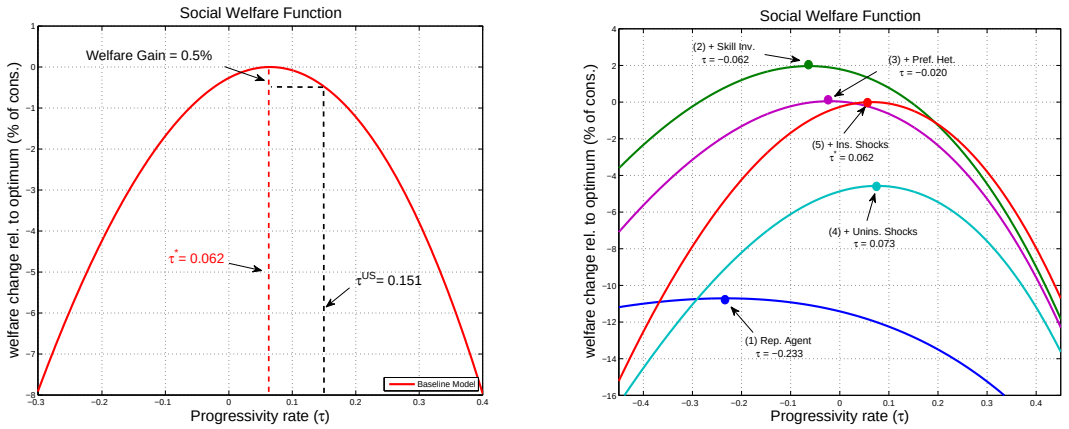
Source: PSID and CEX, 2000-2006. See the main text for details. Bootstrapped standard errors based on 500 replications in parentheses.

2000, 2002, 2004, and 2006 and the Consumption Expenditure Survey (CEX) for years 2000–2006. We apply the same sample selection criteria outlined in Section 2. We first regress individual log wages, individual log hours, and household log consumption on year dummies, a quartic in age, and (for consumption) household composition dummies. We then use the residuals from these regressions to construct the empirical counterpart of the moments in (43) plus the three moments at age "zero" (an average of ages 25-29 in the data). The minimum distance procedure 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. Given an estimate for v_{α}, the innovation variance v_{ω} is pinned down. The estimated parameter values are summarized in Table 1, together with the other pre-determined parameter values.

In the data, $var(\log w) = 0.43$, $var(\log h) = 0.11$, $var(\log c) = 0.18$, and $var^0(\log c) = 0.15$.[36] The estimates in Table 1 imply that, net of measurement error, (i) the insurable component accounts for 40 percent of the variance of wages, the uninsurable component accounts for one-quarter, and the heterogeneity in skills for the residual one-third; (ii) cross-sectional dispersion in the disutility of work effort explains almost half of the hours variation, and insurable shocks explain almost one-quarter; (iii) one-fifth of consumption inequality is due to dispersion in the disutility of work, and the residual is accounted for equally by uninsurable wage risk and skill heterogeneity; (iv) the growth in the variance of log consumption over the life cycle (ages 25-60) is around 0.10. These findings are broadly in line with the results in Heathcote et al. (2014).

To set the 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.

[36]The values of the other empirical moments used in the estimation are $cov(\log h, \log w) = -0.09$, $cov(\log h, \log c) = 0.03$, $cov(\log w, \log c) = 0.15$, $var^0(\log w) = 0.28$, and $cov^0(\log w, \log c) = 0.10$.



(a) Social Welfare Function

(b) Decomposition of the Social Welfare Function

Figure 3: Social welfare as a function of τ and welfare gain relative to the current U.S. system (left panel). Decomposition of social welfare into various components described in Section 5.2. The optimal value for τ is 0.062.

In Section 7.1 we show that if g was the outcome of voting, the median voter would pick the efficient level g^*, which provides a theoretical foundation for this calibration choice. Since, over the period 2000-2006, $g = G/Y = 0.189$, we set $\chi = 0.233$. Because the optimal τ is sensitive to the choice for χ, Section 6.4 discusses alternative scenarios.

6.2 Results

Once the optimality condition $g^* = -\chi$ is substituted into (29) and values have been assigned to all the structural parameters, one obtains social welfare $\mathcal{W}(\tau)$ as a function of τ only. Figure 3(a) plots this function, assuming $\gamma = \beta$. The value of progressivity that maximizes social welfare is $\tau^* = 0.062$. The welfare gain from reducing progressivity from the current value of $\tau^{US} = 0.151$ to τ^* is equivalent to 0.5 percent of lifetime consumption.

How different are the actual and optimal schemes? Note that the ratio of the variance of log disposable income to pre-government income is $(1-\tau)^2$. Moving to the optimal scheme would increase this ratio from 0.72 to 0.88. The average income-weighted marginal tax rate would drop from 31 to 24 percent. Section 6.5 compares the tax/transfer schedule under the two systems in more detail.

The right panel 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 (productivity gain from skill investment net of education costs minus the implied welfare loss from between-skill consumption inequality) pushes toward a more progressive system, and the optimal τ moves to the right to -0.062. The concern for additional consumption inequality induced by preference heterogeneity further raises the optimal τ to -0.02. Uninsurable shocks are a stronger source of consumption dispersion, which is reflected in the substantial upward jump in τ to 0.073 when this component is incorporated. Finally, adding the productivity gain from insurable shocks pulls τ back toward zero to its final value of 0.062.

Consumption dispersion generated by preference heterogeneity, skill dispersion, and uninsurable risk induces the government to choose a progressive scheme. If consumption inequality was the government's only concern, τ would be optimally set to one, and the tax/transfer scheme would fully equate post-government income and consumption across households. Besides the desire to provide public goods, which we analyze in Section 6.4, two forces limit progressivity in the model: the distortion to skill investment and the distortion to labor supply.

To measure the strength of these two channels, we compute the optimal τ (i) when $\sigma = \infty$ and labor supply is therefore inelastic and (ii) when skills are exogenous.[37] In the case $\sigma = \infty$, the optimal τ is 0.217, whereas in the case of an exogenous skill distribution, the optimal τ is 0.211. Therefore, the endogeneities of labor supply and skill investment play quantitatively similar roles in limiting progressivity, and in the absence of either one of these channels, optimal progressivity would be substantially higher. Note in particular that endogenous skill investment, a margin ignored in much of the literature, is a very important factor restraining progressivity in our model.[38] Absent this margin, the welfare-maximizing policy would imply an increase in progressivity relative to the current tax system, whereas once this margin is incorporated, reducing progressivity is optimal.

[37] This 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.

[38] Two recent papers that do explicitly model the skill investment margin are Guvenen, Kuruscu, and Ozkan (2014) and Krueger and Ludwig (2013).

6.2.1 Alternative calibration of θ and top marginal tax rates

A long-standing challenge in labor economics is understanding what fraction of wage dispersion reflects differential endogenous skill or human capital investments, and what fraction reflects exogenous differences in productivity. In the context of our model, the variance of the former is $1/\theta^2$ and the variance of the second is $v_\alpha = \delta v_\omega/(1-\delta)$. We now consider an alternative strategy for calibrating θ (and thus v_ω). Since the top end of the model distribution of income is approximately Pareto with parameter θ, we have that

$$\frac{E\left[p\left(s\right)|s>\bar{s}\right]}{p\left(\bar{s}\right)}=\frac{\theta}{\theta-1}. \tag{44}$$

From our PSID sample, we estimate that the ratio in (44) is stable and around 2 for income thresholds above \$250,000, which implies a value of $\theta = 2$.[39] If we use this moment to calibrate θ, instead of using consumption dispersion at labor market entry, and reestimate the other model parameters, we obtain the values in the second row of Table 1. Because the consumption dispersion implied by skill investment is so large with $\theta = 2$, the estimation sets the variance of the uninsurable lifetime shocks v_ω to zero.[40] The model still calls for positive preference heterogeneity and insurable productivity dispersion to account for the cross-sectional inequality of hours in the data. Under this alternative calibration, we obtain $\tau^* = 0.037$. Optimal progressivity is lower than in the baseline calibration, because a larger role for endogenous skills in generating wage dispersion implies a correspondingly smaller role for uninsurable life-cycle shocks, and thus reduces the incentive to provide social insurance against those shocks.

6.3 Nonutilitarian welfare criteria

The utilitarian social welfare function in (31) embeds both the desire to *insure* households against the privately uninsurable life-cycle shocks ω and to *redistribute* against income differentials due to initial heterogeneity in preferences (φ) and learning ability (κ). We now consider some alternative formulations for the social welfare function that retain the desire to insure against uninsurable

[39] Since the benchmark of comparison for our normative results is τ^{US} estimated on PSID data, we use PSID data for this alternative estimate of θ. Tax 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 choice for the threshold $\bar{s}$. Thus, our estimate falls within this range.

[40] As a result, under this calibration, the model cannot generate a rise in consumption inequality over the life cycle.

shocks, but switch off the desire to redistribute with respect to initial conditions. We first note that when there is no need to raise revenue to finance public expenditure ($\chi = 0$), and when all life-cycle productivity shocks are insurable ($v_\omega = 0$) the welfare theorems apply, and thus the laissez-faire competitive equilibrium with $\lambda = 1$ and $\tau = 0$ is efficient.

Proposition 6 [efficiency with $\chi = v_\omega = \tau = 0$]. *If $\chi = v_\omega = 0$, then competitive equilibrium allocations with $\tau = 0$ are efficient. The corresponding planner weights are*

$$\zeta(\varphi, \kappa) = \frac{\exp\left(-\varphi + \frac{\eta}{\theta}\kappa\right)}{\theta/(\theta - 1)}. \tag{45}$$

Proposition 6 describes the Pareto weights for the planner's problem whose solution coincides exactly with the competitive equilibrium with $\tau = 0$. These planner weights, by construction, do not incorporate any desire to redistribute with respect to income differentials reflecting fixed heterogeneity in preferences or learning ability. The social welfare function corresponding to these weights overweights agents with high ability and high diligence.[41] Thus, the competitive equilibrium with $\tau = 0$ does not deliver the allocations that would be chosen by a utilitarian planner with planner weights given by $\zeta(\varphi, \kappa) = 1$.

We label the planner with weights described by eq. (45) the "insurance-only" planner, because the only motive for progressivity with such weights is insuring against ex post realizations of the ω shock. Using similar logic, we also construct preference weights for a φ-neutral planner, who is indifferent to consumption inequality originating from heterogeneity in the taste for leisure, and for a κ-neutral planner who has no desire to respond to income inequality generated by the initial heterogeneity in learning ability κ and the ensuing skill inequality.

Evaluating equilibrium allocations given the weights in eq. (45), we can compute social welfare under the insurance-only social welfare function.

[41] Benabou (2002) uses an alternative approach to evaluate welfare gains while abstracting from purely redistributive effects. Instead of aggregating utilities across individuals, his social welfare function aggregates consumption certainty equivalents.

Table 2: Optimal progressivity under nonutilitarian welfare

	Utilitarian	κ-neutral	φ-neutral	Insurance-only
Redist. wrt κ	Y	N	Y	N
Redist. wrt φ	Y	Y	N	N
Insurance wrt ω	Y	Y	Y	Y
τ^*	0.062	0.003	0.035	-0.025
Welf. gain (pct of c)	0.52	1.40	0.88	1.98

Corollary 6.1 [nonutilitarian welfare]. *Welfare for the insurance-only planner is*

$$\begin{aligned} \mathcal{W}^{ins}(g,\tau) &= \mathcal{W}(g,\tau) \\ &\quad + (1-\tau)^2 \frac{v_\varphi}{2} + \left(1-\tau^2\right)\frac{v_\varphi}{2} \\ &\quad - \frac{1}{2\theta}(1-\tau) + \frac{1}{2(\theta-1)}(1-\tau). \end{aligned} \tag{46}$$

The second line adjusts the welfare component associated with preference heterogeneity which appears in the utilitarian welfare function. The new term that replaces (37) penalizes deviations from $\tau = 0$. The third line corrects the welfare component associated with skill investment (productivity gain, education cost, and consumption inequality).[42] As special cases, the first two lines in (46) give welfare for the φ-neutral planner, and the first and third lines give welfare for the κ-neutral planner.

Table 2 summarizes the results. We find that the insurance-only planner would set $\tau^* = -0.025$. Thus, concern for social insurance against life-cycle wage shocks offsets almost exactly the desire for regressivity linked to public good provision. The φ-neutral planner would set progressivity to 0.035 and the κ-neutral planner to 0.003. Overall, these governments with limited taste for redistribution choose tax systems that are nearly proportional. Under these alternative welfare criteria, the welfare gains of switching from the current progressive system to the optimal near-proportional system are much larger: the insurance-only criterion implies a gain of 2 percent of lifetime consumption.

[42] Also in this case, the term in τ_{-1} is separable, and hence we have omitted it. See the Appendix for details.

6.4 Alternative modeling of G

A theme of this paper is that there is an important interaction between the size of government expenditure and the progressivity of the tax and transfer system: the more spending the government wants to finance, the less progressive will be the optimal tax system. In a representative agent version of the model, Section 5.2.1 showed that the externality associated with public good provision calls for a regressive tax system, with the degree of regressivity increasing in the taste for the public good. In Section 5.3, we revisited the issue in the baseline heterogeneous agent model and characterized the conditions under which this force toward regressivity dominates the utilitarian planner's desire for progressivity to reduce inequality. The key comparative static is that, holding fixed all other structural parameters and the choice of social welfare function, the optimal tax and transfer system should be less progressive the more utility agents attach to publicly provided goods.

In contrast to this paper, most previous analyses of tax design abstract from the choice of public good provision and simply assume that an exogenous level of expenditure must be financed. However, given the interaction we have highlighted between the size of government and the progressivity of the tax and transfer system, one might expect the details of how expenditure is modeled to have a quantitatively significant impact on the conclusions for tax design. We now illustrate that this is in fact the case.

In the baseline model, we assumed that (i) households derive some utility from government expenditures ($\chi > 0$) and that (ii) the government chooses the fraction of output g to be transformed into the publicly provided good. We now set $\chi = 0$ throughout – which implies that the publicly provided good is not valued – and examine two alternative models for spending that are popular in the existing literature on tax design. In the first alternative model, g is fixed exogenously to $g = g^{US} = 0.189$. Thus, a fraction of output equal to government purchases' share in the baseline economy is wasted. In the second alternative, the *level* of expenditures is fixed to $G = G^{US} = g^{US}Y(\tau^{US})$.

Table 3 reports the optimal degree of progressivity in these two cases for the utilitarian and the insurance-only planners. Comparing the baseline model with $\chi > 0$ and $g^* = \chi/(1+\chi)$ with the case $\chi = 0$ and $g = g^{US} = 0.189$, we see that, absent the public good externality, the utilitarian

Table 3: Optimal progressivity under different scenarios for G

Case		Utilitarian				Insurance-only			
		$\frac{G}{Y}$	τ^*	T'	$\Delta\mathcal{W}$	$\frac{G}{Y}$	τ^*	T'	$\Delta\mathcal{W}$
g^* endog.	$\chi = 0.233$	0.189	0.062	0.239	0.52%	0.189	-0.025	0.169	1.98%
g exog.	$\chi = 0$	0.189	0.182	0.337	0.06%	0.189	0.089	0.261	0.22%
G exog.	$\chi = 0$	0.179	0.069	0.236	0.48%	0.171	-0.014	0.159	1.83%

Note: $T' = 1 - (1 - \tau)(1 - g)$ is the income-weighted average marginal tax rate. $\Delta\mathcal{W}$ is the welfare change in consumption equivalent units. G in the third row is fixed at the baseline absolute value, which is equal to 18.9% of GDP in the economy with $\tau^{US} = 0.151$.

planner desires substantially more progressivity – even more progressivity than is embedded in the current system. Note that the optimal τ in this case (0.182) is exactly the same as the optimal τ would be in an economy with $\chi = 0$ and $g = 0$.[43]

The second case (where the level of G is fixed) gives an optimal value for τ very similar to the baseline model, and the logic is similar: the planner internalizes that less progressive taxation encourages labor supply and makes it easier to finance public good provision. With a fixed amount G spent on public goods, more output makes financing government purchases easier – which acts as an incentive to limit progressivity.

Thus, we have shown that a seemingly minor difference in how government spending is modeled – fixed in level, versus fixed as a share of output – has a very large impact on optimal progressivity. The lesson we draw is that policy choices regarding the level of expenditure and the distribution of the net tax burden are intimately interrelated and are best studied in models that integrate the modeling of public good provision and tax design.

6.5 Taking stock

To visualize the differences between the actual and the optimal tax system, Figure 4 plots marginal and average tax rates in the current scheme $\left(\tau^{US} = 0.151\right)$, and in the optimal scheme under the utilitarian and insurance-only criteria, for the baseline model where G is valued and for the model where $\chi = 0$ and g is exogenous.

[43]This result is an application of Corollary 4.4, which stated that (holding fixed χ), the optimal τ is independent of g. One way to understand the result is that when the tax planner has to devote a fixed slice of output to government spending, the planner is effectively operating in an economy with lower total factor productivity (TFP). The optimal τ is invariant to g for the same reason that the optimal τ is invariant to TFP.

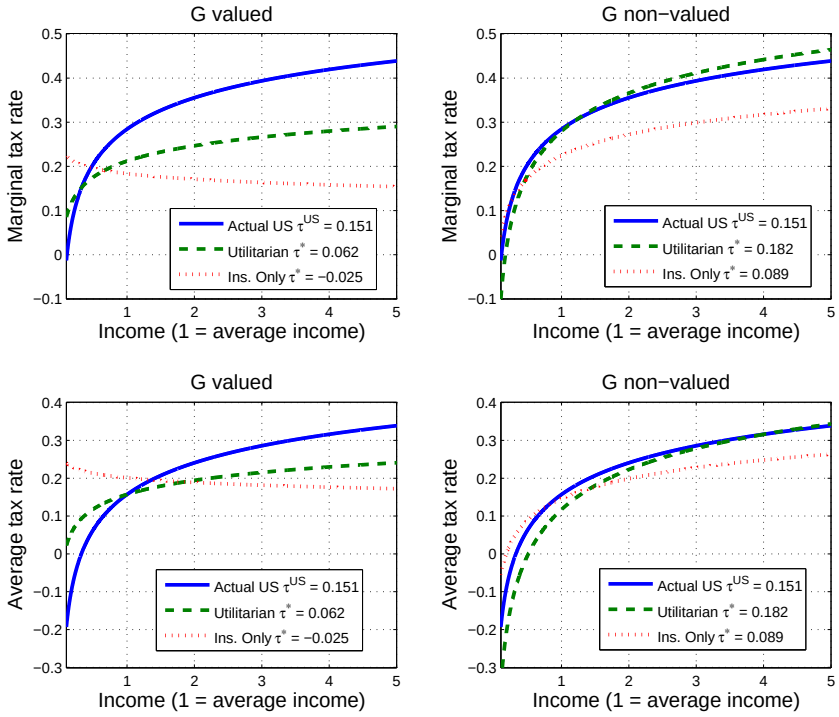


Figure 4: Top-left and Bottom-left panels: marginal and average tax rates in the baseline model with $\chi > 0$ and G valued. Top-right and bottom-right panels: marginal and average tax rates in the baseline model with $\chi = 0$ and $\bar{g}$ exogenous.

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

7 Extensions

This section contains four extensions: a politico-economic analysis and an analysis of a progressive consumption tax in the baseline model, a model in which skill investments are irreversible, and a version of the model with skill bias in the production technology.

7.1 Political-economic determination of progressivity

After our extensive 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 human capital accumulation decision is reversible, so the transition to a new steady state is immediate.[44]

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 that g should be set equal to $\chi/(1+\chi)$; (ii) notwithstanding multidimensional heterogeneity, the attitude of individual agents toward progressivity τ can be summarized by a single summary statistic, so voters are effectively heterogeneous in only one dimension. As a consequence, the median voter theorem applies.

We begin by proving that all agents agree on the optimal size of government.

Proposition 7 [agreement on G]. *When voting over g, every agent chooses $\hat{g} = g^* = \chi/(1+\chi)$, independently of the choice for τ.*

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

$$\chi \frac{1}{g} = \frac{c_i}{1-g} \cdot \frac{1}{c_i}. \tag{47}$$

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 (47). The second term is the marginal utility of private consumption. The key to the result in Proposition

[44] As explained previously, a time-varying τ would break the no-bond-trade result, and the wealth distribution would become a relevant part of the state space.

7 is that the marginal cost of increasing g is the same for every individual. A low earnings agent has a high marginal utility of private consumption, but also gains little additional consumption from reducing average tax rates. The two effects exactly cancel out, and all agents agree on the optimal choice for g.

We now consider voting over τ and make an additional simplifying assumption: voting occurs before the realization of the insurable shock ε. Since ε is i.i.d., its current value has little impact on an agent's preferred permanent degree of progressivity. The individual state vector relevant for a voter is, therefore, $(\varphi, \alpha, \kappa)$.[45] Expected lifetime utility for an agent with characteristics (a, φ, κ) (ignoring, as usual, the separable term in τ_{-1}) is given by

$$\begin{aligned} U(\alpha, \varphi, \kappa; g, \tau) = \mathcal{W}(g, \tau) \quad &- \quad (1-\tau)\frac{\beta\delta}{1-\beta\delta}\frac{v_\omega}{2} - (1-\tau)\left(-\frac{\beta\delta}{\gamma-\beta\delta}\frac{v_\omega}{2} - \frac{v_\varphi}{2} + \frac{1}{2\theta}\right) \\ &+ \quad (1-\tau)\left(\alpha - \varphi + \frac{\eta}{2\theta}\kappa\right). \end{aligned} \tag{48}$$

Proposition 8 [median voter]. *Under a majority rule voting system, the value for τ that would emerge is the value that maximizes eq. (48) for the agent with the median value for the random variable $x = \alpha - \varphi + \frac{\eta}{2\theta}\kappa$.*

The median voter theorem applies in our economy because preferences are single peaked in τ. This follows from our earlier result on the concavity of $\mathcal{W}(g, \tau)$ (Corollary 4.1). Concavity in $\mathcal{W}(g, \tau)$ translates into concavity in $U(\alpha, \varphi, \kappa; g, \tau)$ since the additional terms in eq. (48) are linear in τ. Note also that for the purposes of characterizing attitudes to progressivity, the three-dimensional vector over $(\varphi, \alpha, \kappa)$ 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 Exponentially-Modified Gaussian (EMG) random variable.

We find that the median voter – the agent with the median value for x – would choose $\tau^{med} = 0.084$. Relative to the utilitarian planner's choice $(\tau^* = 0.062)$, this modestly higher degree of progressivity translates into an increase in the average effective marginal tax rate of 1.8 percentage points. To understand the difference between the two values, note first that in the limiting case $\theta \to \infty$ and $v_\omega = 0$, $U\left(x^{med}; g, \tau\right) = \mathcal{W}(g, \tau)$ and thus the median voter would choose exactly the same degree of progressivity as the utilitarian planner. The logic for this result is that the id-

[45] κ is a state because agents can reoptimize their choice for skills after the change in τ.

iosyncratic preference term appears in individual log consumption in the form $-(1-\tau)\varphi$ (see eq. 22). Given a symmetric normal distribution for φ and utility that is logarithmic in consumption, average utility is then equal to utility of the median φ agent. Similar logic would apply to fixed heterogeneity in initial labor productivity α_0. 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 as insurance against future uninsurable shocks (the second term on the right-hand side of eq. 48).[46] With respect to skill heterogeneity, the agent with median ability κ has less than average ability, because κ is exponentially distributed. Thus, the median κ agent has more to gain from progressive taxation and would choose a higher value for τ.

We conclude by noting that the median voter prefers a lower value for τ than our empirical estimate. However, this result is predicated on our baseline value for χ. One could use the median voter model to identify how strong the preference for public goods must be so as to deliver $\tau = \tau^{US} = 0.151$ as a political economy outcome. The answer is $\chi = 0.057$. Thus, the model can rationalize the observed level of progressivity if agents have a weaker preference for public goods than we assumed in our baseline calibration.

7.2 Progressive consumption taxation

The comparison between income and consumption taxes has a long tradition in public finance and macroeconomics. The main argument set forth by the literature in favor of consumption taxes is *productive* efficiency. As explained by Nishiyama and Smetters (2007), a tax on consumption is an income tax where saving is exempt and, as a result, it imposes a smaller distortion on capital accumulation compared with an income tax. Correia (2010) shows that when a flat consumption tax is augmented with a universal lump-sum transfer, productive efficiency can be improved without any welfare loss in terms of higher inequality.

Here, we put forth a novel argument in favor of *progressive* consumption taxes.[47] We show that, compared with progressive earnings taxation, progressive consumption taxation reduces distortions

[46] The median value for α is $-\frac{\delta}{1-\delta}\frac{v_\omega}{2}$. Note that when $\gamma = \beta$, the third term on the right-hand side of eq. (48) therefore offsets the contribution to welfare from the term in α for the median voter.

[47] Implementation in our model would follow McCaffery's (2002) proposal of taxing income progressively with an exemption for savings.

to the distribution of labor supply and allows stronger protection against life-cycle earnings shocks that are privately uninsurable.

In our model, with a consumption expenditure tax/transfer scheme, the budget constraint (16) for household i of age a becomes

$$\lambda c_{ia}^{\frac{1}{1-\tau}} + \delta q b_{i,a+1} = p\left(s_i\right) \exp\left(\alpha_{ia} + \varepsilon_{ia}\right) h_{ia} + B\left(\varepsilon_{ia}\right), \tag{49}$$

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

$$\log h\left(\varphi, \varepsilon; \tau\right) = \log H^{RA}\left(\tau\right) - \varphi + \frac{1}{\sigma}\varepsilon - \frac{1}{\sigma}\mathcal{M}\left(v_\varepsilon; 0\right). \tag{50}$$

Compared with (21), we note two differences. First, the pass-through from insurable shocks to hours is undistorted by taxes and equal to the Frisch elasticity $1/\sigma$, *not* the tax-modified elasticity $1/\hat{\sigma}$. Second, the productivity gain from insurable dispersion v_ε is unaffected by τ. The new productivity gain term $\mathcal{M}\left(v_\varepsilon; 0\right) = \frac{1}{\sigma}\frac{v_\varepsilon}{2}$ is obtained, as the notation suggests, by simply setting $\tau = 0$ in $\mathcal{M}\left(v_\varepsilon; \tau\right)$.[48] To understand these differences, recall that the consumption allocation is independent of ε. Therefore, taxing consumption, instead of earnings, is a form of taxation that specifically targets the uninsurable shocks (which pass through to consumption) without distorting the response of hours worked to insurable shocks. In sum, progressive consumption taxation is more efficient than progressive income taxation. The next proposition summarizes this result by stating the form of the social welfare function.

Proposition 9 [welfare with progressive consumption tax]. *Under a progressive consumption tax/transfer system, social welfare is*

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

where $\mathcal{W}\left(g, \tau; \tau_{-1}\right)$ is defined in Proposition 4.

The productivity gain term $\frac{1}{\hat{\sigma}}v_\varepsilon$ becomes $\frac{1}{\sigma}v_\varepsilon$ and the welfare loss from hours fluctuation $-\sigma\frac{1}{\hat{\sigma}^2}\frac{v_\varepsilon}{2}$ becomes $-\sigma\frac{1}{\sigma^2}\frac{v_\varepsilon}{2}$.

[48] The consumption allocation is given by

$$\log c\left(\varphi, \alpha, s; \lambda, \tau\right) = \log C^{RA}\left(\lambda, \tau\right) + \left(1-\tau\right)\mathcal{M}\left(v_\varepsilon; 0\right) - \left(1-\tau\right)\varphi + \left(1-\tau\right)\log p\left(s; \tau\right) + \left(1-\tau\right)\alpha,$$

where $\log C^{RA}\left(\lambda, \tau\right) = \log\lambda + \frac{1}{\left(1+\hat{\sigma}\right)}\log(1-\tau)$.

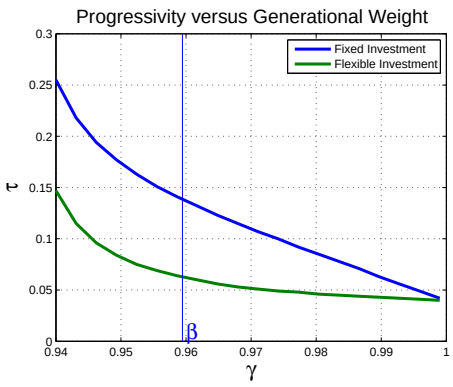


Figure 5: The figure shows how the optimal choice for τ varies with γ in the baseline flexible (reversible) investment model and the alternative fixed (irreversible) investment model.

The degree of consumption tax progressivity that maximizes $\mathcal{W}^{cons}(g, \tau; \tau_{-1})$ is 0.073. Because this tax system does not distort the distribution of labor supply across households, the planner chooses a slightly higher value for τ than under the baseline earnings tax system, thereby providing better insurance against life-cycle earnings shocks. The welfare gain associated with switching from the current earnings tax system to this progressive consumption tax system is 0.56 percent, slightly larger than the baseline welfare gain (0.52 percent).

7.3 Progressivity when past skill investment is sunk

In our baseline model, we assumed fully reversible investment. This implied that (i) the optimal choice for τ is independent of the pre-existing distribution of skills, and (ii) transition following any tax reform is instantaneous. Although the assumption that skill investment is fully reversible lends transparency to the analysis, it is extreme. We now consider the opposite extreme assumption, namely that skill investment is chosen once and for all at age zero and can never be adjusted thereafter. This introduces an additional force in the direction of more progressivity: by making the tax system more progressive, the social planner can immediately reduce consumption inequality in the economy without simultaneously distorting skill investments for agents who entered the economy in the past.

As in our baseline model, we consider an unanticipated once-and-for-all change in the tax system, defined as permanent changes in the fiscal policy parameters at date 0 from (τ_{-1}, g_{-1}) to (τ, g). With irreversible investment, transition is no longer instantaneous because, given an

unanticipated change in τ, output gradually evolves over time as the population share of agents who make skill investments under the new tax regime rises.

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

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

When we set $\gamma = \beta = 0.9594$, and assume that the initial steady state corresponds to our estimated progressivity value, $\tau_{-1} = \tau^{US} = 0.151$, we find that the optimal new permanent choice for τ is 0.137. This value is larger than the optimum in our baseline model with fully reversible investment ($\tau^* = 0.062$). The intuition is that the utilitarian planner now uses progressive taxation to tax away past sunk skill investments by high ability (high κ) individuals, a result that is analogous to the desire to tax initial capital under the Ramsey approach to optimal taxation in the standard growth model.

Figure 5 shows how the optimal choice for τ varies with the planner's intergenerational discount factor, γ, for both the baseline flexible investment model and the alternative fixed investment model. In both models, the more weight the planner puts on agents who entered the economy in the past (i.e., the lower is γ), the larger is the optimal choice for τ. In the flexible investment model, this reflects the contribution of the sixth line of eq. (29), which captures the incentive to redistribute to offset wage dispersion due to permanent uninsurable productivity shocks. A lower γ implies more relative weight on older agents who exhibit high idiosyncratic productivity dispersion and thus translates into a higher optimal τ. The optimal choice for τ is more sensitive to γ in the fixed investment model because now another force comes into play: a smaller value for γ means that the planner cares more about reducing consumption inequality among existing cohorts (whose skill investments are sunk) and cares less about distorting skill investment for future cohorts (whose investments will adjust). As $\gamma \to 1$, the planner attaches so much weight to future

generations that the short-run gains from expropriating past skill investments become irrelevant, so the optimal choices for τ in the models with reversible and irreversible investment converge. In both cases, the expression for social welfare corresponds to expected lifetime utility for a newborn agent in steady state, and the value for τ that maximizes expected lifetime utility for a newborn is $\tau = 0.040$.

The baseline model with fully reversible investment and the alternative with irreversible investment are two extremes. A compromise between the two models would be one in which it is impossible for agents to reduce past skill investments, but previous investments can be supplemented for free. In this hybrid model, the welfare effects of unanticipated progressivity reductions would be identical to those in the flexible investment model, whereas the welfare effects of progressivity increases would be identical to those in the fixed investment model. It follows that, under our baseline calibration with $\gamma = \beta$, the optimal choice for τ in this hybrid model would again be $\tau = 0.062$. The logic is that in the fixed investment model, leaving progressivity unchanged is preferred to increasing progressivity (recall that the optimal choice in that model involves reducing τ from 0.151 to 0.137), and thus in the hybrid model the planner will not want to increase progressivity.

7.4 Skill bias in production

So far we have emphasized a theory in which all differences in skill prices are due to differences in the relative scarcity of different skill types. However, the literature has also explored an alternative source of wage dispersion due to skill differences, namely, bias in production in favor of certain types of skills (e.g., Katz and Murphy, 1992). For example, the production technology may put relatively high weight on high skilled workers. We now generalize our analysis to incorporate both sources of skill price dispersion by extending the production function in (5) as follows:

$$Y = \left\{ \int \exp(\varrho s) \cdot [N(s) \cdot m(s)]^{\frac{\theta-1}{\theta}} \, ds \right\}^{\frac{\theta}{\theta-1}}, \tag{51}$$

where the weight $\exp(\varrho s)$ captures the skill bias in production for skill level s. The technology in (51) is, in essence, a multiskill version of the two-skill production function of Katz and Murphy (1992). The skill bias must be bounded to ensure finite production, and we therefore make the following assumption.

Assumption 1 $\varrho < \frac{\theta-1}{\theta}/\sqrt{2}$.

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

$$p(s) = \exp(\varrho s)\left(\frac{Y}{N \cdot m(s)}\right)^{\frac{1}{\theta}}, \tag{52}$$

where N is the effective hours supplied by every skill type (defined in Corollary 3.2).

The following amended version of Proposition 3 summarizes how the skill prices and skill choices are affected by the presence of skill bias in production.[49]

Proposition 10 [equilibrium prices with skill bias]. *In equilibrium, the skill price is given by*

$$\log p(s;\tau) = \pi_0(\tau) + \pi_1(\tau) \cdot s(\kappa;\tau), \tag{53}$$

where the marginal return to skill is $\pi_1(\tau) = \frac{\varrho}{2} + \sqrt{\left(\frac{\varrho}{2}\right)^2 + \frac{1}{\theta(1-\tau)}}$. *The equilibrium distribution of log skill premia* $\pi_1(\tau) \cdot s(\kappa;\tau)$ *is exponential with parameter* $(1-\tau)^{-1}(\pi_1(\tau))^{-2}$.

As is clear from Proposition 10, the model with skill bias in production preserves the result that in equilibrium, the price of skills is log-linear in s. This can be anticipated from the fact that the logarithm of the marginal product of labor is now the skill bias ϱ plus the scarcity term from before (see eq. 52). As before, the upper tail of the wage distribution is Pareto, the marginal return to skill, $\pi_1(\tau)$, is increasing in τ, and the base price $\pi_0(\tau)$ is falling in τ.[50] However, the sensitivity of π_1 and π_0 to τ is smaller the larger is the exogenous skill bias ϱ. The logic is that, in this model, some of the price differences are due to ϱ and are therefore independent of the effect of τ on the relative supply of different skill types.

The variance of pre-tax log skill prices is now

$$var(\log p(s;\tau)) = (1-\tau)^2(\pi_1(\tau))^4. \tag{54}$$

[49] To simplify the algebra, we assume $\mu = \eta$ here. This assumption is innocuous for the purposes of computing the optimal degree of progressivity since, as in the baseline model, there is no interaction between τ and either μ or η in the expression for social welfare (recall Corollary 4.5).

[50] The base skill price is $\pi_0(\tau) = -\frac{1}{\theta-1}\log\left(\frac{(\theta-1)}{\theta}\frac{1}{(1-\tau)}\left(\frac{\varrho}{2} + \sqrt{\left(\frac{\varrho}{2}\right)^2 + \frac{1}{\theta(1-\tau)}}\right)^{-1} - \varrho\right)$.

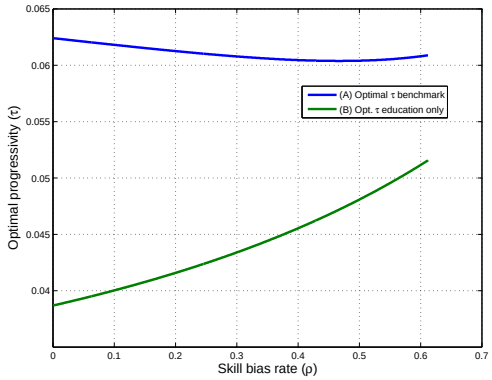


Figure 6: The figure plots the optimal degree of progressivity (τ) under all combinations of θ and $\varrho > 0$ consistent with a cross-sectional variance of skill prices of 0.1012.

It is straightforward to show that the variance is strictly falling (increasing) in τ whenever $\varrho > 0$ ($\varrho < 0$). The logic is that when $\varrho > 0$, the quantity effect of more progressivity (the skill distribution is more compressed toward zero) dominates the price effect (higher skill types are relatively more scarce and command a higher wage premium). The model in Guvenen, Kuruscu, and Ozkan (2014) has this property: they focus on the limiting case in which $\theta = \infty$ and all skill price inequality is driven by skill bias ϱ only, so that the price effect is entirely absent.

The social welfare expression in the model with skill bias is unchanged relative to the baseline model expression up to the skill-related terms in $\mathcal{W}(g, \tau; \tau_{-1})$ (the second, third, and fourth lines in equation 29). In the Appendix we provide exact expressions for how these terms depend on ϱ.

Note that all the model moments we used for estimation (in Section 6) are the same as before, except for the variance of skill prices (54). The identification of the model parameters is therefore the same, except for the estimate of θ. From eq. (54) it is clear that skill price dispersion can stem from either a large complementarity between skills, i.e., a low θ, or a large skill bias in production, i.e., a large ϱ.

The possible combinations of (ϱ, θ) that are consistent with the calibration in Section 6 can therefore be derived by assuming the same variance of skill prices, i.e., $var(\log p(s; \tau)) = 0.1012$, imposing $\tau^{US} = 0.151$, and using the expression for the variance in eq. (54), which is a function of ϱ and θ. Clearly, larger values of ϱ are associated with larger values of θ.

Next, we compute and plot in Figure 7.4 the optimal τ for a range of possible values for ϱ. We construct two plots. The first, labeled "benchmark," is the full model with all sources of heterogeneity. The second, labeled "education only," is a stripped-down version of the model in which agents are heterogeneous only with respect to κ. In both cases, as we change ϱ we also simultaneously adjust θ, so as to hold fixed the total variance of skill prices.

This exercise yields a number of interesting insights. The first is that if skill acquisition ability κ were the only source of heterogeneity, then the optimal τ would be steeply increasing in ϱ. The logic is that as skill price differentials become more exogenous and less sensitive to τ, the utilitarian planner prefers more progressivity. Conversely, when general equilibrium effects are important (ϱ and θ are low), the planner chooses modest progressivity, understanding that a higher value for τ would reduce skill investment and raise the skill premium, thereby undoing to some extent the redistribution the planner desires.

Second, in the full model that incorporates all the redistribution motives, the optimal value for τ is (i) uniformly higher than in the "education-only" model and (ii) largely insensitive to ϱ. From the first observation we can infer that other sources of heterogeneity (preference heterogeneity, uninsurable productivity shocks) incline the utilitarian planner toward strong progressivity, whereas heterogeneity in skill acquisition ability inclines the planner toward more moderate progressivity – because the planner understands that progressivity crowds out skill investment. Comparing the optimal τ plots across the two models, we can infer that when ϱ is large, the planner is more concerned about crowding out and thus chooses a value for τ in the full model that is close to the optimal τ in the education-only model. Crowding out is a bigger concern when ϱ is large because a higher value for τ directly reduces the private return to skill investment. Conversely, when ϱ is low, crowding out is less of a concern, because as skill investment falls the skill premium rises, and thus the decline in investment is mitigated.

We conclude that under the benchmark calibration, the optimal value for τ is largely insensitive to whether the variance in skill prices is primarily driven by relative scarcity (low θ and low ϱ) or by a high-skill-biased production technology (high θ and high ϱ). The finding that the optimal τ is robust to these alternative models of endogenous skill investment is reassuring, since separately identifying θ and ϱ is a challenging task. However, it would be wrong to infer that one can neglect the skill investment margin when thinking about how progressive taxes should be. Recall from

Section 6.2 that, when skills are exogenous, making taxes much more progressive is optimal. Rather, the conclusion here is that two different textbook models of endogenous skill investment deliver similar normative implications.

8 Conclusions

This paper has developed a tractable equilibrium framework to study the optimal degree of progressivity of the tax and transfer system. Our main result is an expression for social welfare as a function of (i) policy parameters defining the degree of progressivity and the level of public expenditure and (ii) structural parameters defining preferences, technology, and households' access to private insurance. These parameters regulate the relative strength of the economic forces pushing for and against progressivity. A utilitarian planner wants a progressive tax system to redistribute with respect to 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 translates into lower aggregate output by discouraging skill investment and labor supply, and thereby exacerbates the externality associated with the provision of public goods.

We approached this problem by restricting attention to a particular class of tax/transfer functions that fits the U.S. household-level cross-sectional relationship between disposable and gross income remarkably well. In addition, the optimal policy in this class comes very close, in welfare terms, to decentralizing the constrained-efficient Mirrlees allocations in a stripped-down version of our framework without skill investment.

The presence of skill investment in the model yields several interesting insights. First, tax progressivity is not an effective way to compress inequality in pre-tax wages. Although higher progressivity does reduce skill investment, and thereby compresses inequality in skills, it also has an offsetting general equilibrium price effect: as high skill types become relatively more scarce, the equilibrium pre-tax skill premium increases. Second, and relatedly, when the income distribution is right skewed because of skill complementarity in production, a thicker Pareto tail does not necessarily translate into higher optimal marginal tax rates at the top, in contrast to a well-known result in the Mirrlees literature.

When we parameterize the model and quantify the net impact of the various forces for and against progressivity, we conclude that there would be welfare gains from making the U.S. tax and transfer system less progressive. The associated gains do not exceed two percent of lifetime consumption. This finding is robust to the degree of irreversibility of skill investment and to the relative importance of technological skill bias versus skill complementarity in determining the skill premium. However, there is a significant interaction between the level of spending and the progressivity of the tax system: the more valued are public goods, and thus the more net revenue must be collected, the less progressive should be taxes and transfers. If one takes the view that government expenditures are not valued by households, then the current tax/transfer scheme is appropriately progressive.

We have also made some progress toward a positive theory of tax progressivity by investigating the fiscal policy parameters that would emerge from a majority voting model. We found that the median voter prefers higher progressivity than a utilitarian social planner, in part because the median agent has below-average aptitude for skill investment. This result helps rationalize why the current tax system appears overly progressive relative to the utilitarian optimum.

Finally, we find that switching from a progressive earnings tax to a progressive consumption tax would generate welfare gains, because this would be a way for the government to directly compress dispersion in the uninsurable component of labor income shocks – which transmit to consumption – without distorting responses to insurable shocks – which affect earnings but do not pass through to consumption.

References

ATKINSON, A. B., T. PIKETTY, AND E. SAEZ (2011): "Top Incomes in the Long Run of History," *Journal of Economic Literature*, 49(1), 3–71.

ATTANASIO, O., AND S. J. DAVIS (1996): "Relative Wage Movements and the Distribution of Consumption," *Journal of Political Economy*, 104(6), 1227–1262.

BAKIS, O., B. KAYMAK, AND M. POSCHKE (2012): "On the Optimality of Progressive Income Redistribution," Working Paper.

BENABOU, R. (2000): "Unequal Societies: Income Distribution and the Social Contract," *American Economic Review*, 90(1), 96–129.

——— (2002): "Tax and Education Policy in a Heterogeneous-Agent Economy: What Levels of Redistribution Maximize Growth and Efficiency?," *Econometrica*, 70(2), 481–517.

CALVO, G. A., AND M. OBSTFELD (1988): "Optimal Time-Consistent Fiscal Policy with Finite Lifetimes," *Econometrica*, 56(2), 411–432.

CHEN, S.-H., AND J.-T. GUO (2013): "Progressive Taxation and Macroeconomic (In) stability with Productive Government Spending," *Journal of Economic Dynamics and Control*, 37(5), 951–963.

CHETTY, R., AND E. SAEZ (2010): "Optimal Taxation and Social Insurance with Endogenous Private Insurance," *American Economic Journal: Economic Policy*, 2(2), 85–114.

COCHRANE, J. H. (1991): "A Simple Test of Consumption Insurance," *Journal of Political Economy*, 99(5), 957–976.

CONESA, J. C., AND D. KRUEGER (2006): "On the Optimal Progressivity of the Income Tax Code," *Journal of Monetary Economics*, 53(7), 1425–1450.

CONSTANTINIDES, G. M., AND D. DUFFIE (1996): "Asset Pricing with Heterogeneous Consumers," *Journal of Political Economy*, 104(2), 219–240.

CORREIA, I. (2010): "Consumption Taxes and Redistribution," *American Economic Review*, 100(4), 1673–1694.

DIAMOND, P., AND E. SAEZ (2011): "The Case for a Progressive Tax: From Basic Research to Policy Recommendations," *Journal of Economic Perspectives*, 25(4), 165–190.

EATON, J., AND H. S. ROSEN (1980): "Optimal Redistributive Taxation and Uncertainty," *Quarterly Journal of Economics*, 95(2), 357–364.

FARHI, E., AND I. WERNING (2013): "Insurance and Taxation over the Life Cycle," *Review of Economic Studies*, 80(2), 596–635.

FEENBERG, D., AND E. COUTTS (1993): "An Introduction to the TAXSIM model," *Journal of Policy Analysis and Management*, 12(1), 189–194.

FELDSTEIN, M. S. (1969): "The Effects on Taxation on Risk Taking," *Journal of Political Economy*, 77(5), 755–764.

——— (1973): "On the optimal progressivity of the income tax," *Journal of Public Economics*, 2(4), 357–376.

GOLOSOV, M., M. TROSHKIN, AND A. TSYVINSKI (2011): "Optimal Dynamic Taxes," NBER Working Paper 17642.

GORRY, A., AND E. OBERFIELD (2012): "Optimal Taxation Over the Life Cycle," *Review of Economic Dynamics*, 15(4), 551–572.

GUNER, N., R. KAYGUSUZ, AND G. VENTURA (2012a): "Income Taxation of U.S. Households: Facts and Parametric Estimates," CEPR Discussion Paper.

——— (2012b): "Taxation and Household Labour Supply," *Review of Economic Studies*, 79(3), 1113–1149.

GUVENEN, F., B. KURUSCU, AND S. OZKAN (forthcoming): "Taxation of Human Capital and Wage Inequality: A Cross-Country Analysis," *Review of Economic Studies*.

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.

HEATHCOTE, J., K. STORESLETTEN, AND G. L. VIOLANTE (2008): "Insurance and Opportunities: A Welfare Analysis of Labor Market Risk," *Journal of Monetary Economics*, 55(3), 501–525.

——— (forthcoming): "Consumption and Labor Supply with Partial Insurance: An Analytical Framework," *American Economic Review*.

HEATHCOTE, J., AND H. TSUJIYAMA (2013): "Optimal Income Taxation: Mirrlees Meets Ramsey," *Mimeo*.

HECKMAN, J., L. LOCHNER, AND C. TABER (1998): "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.

HUGGETT, M. (1993): "The Risk-Free Rate in Heterogeneous-Agent Incomplete-Insurance Economies," *Journal of Economic Dynamics and Control*, 17(5-6), 953–969.

KAKWANI, N. C. (1977): "Measurement of Tax Progressivity: An International Comparison," *Economic Journal*, 87(345), 71–80.

KATZ, L. F., AND K. M. MURPHY (1992): "Changes in Relative Wages, 1963-1987: Supply and Demand Factors," *Quarterly Journal of Economics*, 107(1), 35–78.

KEANE, M. P. (2011): "Labor Supply and Taxes: A Survey," *Journal of Economic Literature*, 49(4), 961–1075.

KRUEGER, D., AND A. LUDWIG (2013): "Optimal Progressive Labor Income Taxation and Education Subsidies When Education Decisions and Intergenerational Transfers Are Endogenous," *American Economic Review*, 103(3), 496–501.

KRUEGER, D., AND F. PERRI (2010): "Public versus Private Risk Sharing," *Journal of Economic Theory*, 146(3), 920–956.

MCCAFFERY, E. J. (2002): *Fair not Flat: How to Make the Tax System Better and Simpler*. Chicago: University of Chicago Press.

MEGHIR, C., AND L. PISTAFERRI (2010): "Earnings, Consumption and Lifecycle Choices," in *Handbook of Labor Economics*, ed. by O. C. Ashenfelter, and D. E. Card, vol. 4B of *Handbooks of Labor Economics*. Elsevier.

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

MUSGRAVE, R. A. (1959): *The Theory of Public Finance: A Study in Public Economy*. New York: McGraw-Hill.

NISHIYAMA, S., AND K. SMETTERS (2007): "Does Social Security Privatization Produce Efficiency Gains?," *Quarterly Journal of Economics*, 122(4), 1677–1719.

PERSSON, M. (1983): "The Distribution of Abilities and the Progressive Income Tax," *Journal of Public Economics*, 22(1), 73–88.

PIKETTY, T., AND E. SAEZ (2003): "Income Inequality in the United States, 1913-1998," *The Quarterly Journal of Economics*, 118(1), 1–39.

RAMSEY, F. P. (1927): "A Contribution to the Theory of Taxation," *Economic Journal*, 37(145), 47–61.

ROTHSCHILD, C., AND F. SCHEUER (2013): "Redistributive Taxation in the Roy Model," *The Quarterly Journal of Economics*, 128(2), 623–668.

SAEZ, E. (2001): "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 68(1), 205–229.

SAMUELSON, P. A. (1954): "The Pure Theory of Public Expenditure," *Review of Economics and Statistics*, 36(4), 387–389.

STANTCHEVA, S. (2013): "Optimal Taxation and Human Capital Policies over the Life Cycle," *Mimeo*.

STIGLITZ, J. E. (1982): "Utilitarianism and Horizontal Equity : The Case for Random Taxation," *Journal of Public Economics*, 18(1), 1–33.

——— (1985): "Equilibrium Wage Distribution," *Economic Journal*, 95(379), 595–618.

UHLIG, H. (1996): "A Law of Large Numbers for Large Economies," *Journal of Economic Theory*, 8(1), 41–50.

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

APPENDIX

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

A.1 Proof of Proposition 1 [equilibrium]

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

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

where $\bar{c}(g, \tau)$ is a constant that does not depend on any individual state variables, and f is a function of individual age-invariant states (φ, s). We now show that this class of allocations implies that agents do not want to trade the bond. In the proof of the next proposition, we show that, absent bond trade, allocations do indeed take this form, verifying the original conjecture.

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

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

The expected marginal rate of substitution between states is

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

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

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

as stated in Proposition 1 (in log terms).

A.2 Proof of Proposition 2 [hours and consumption]

We follow the proof in Heathcote, Storesletten, and Violante (2013). We first guess that there is no bond trade in equilibrium. Given no bond trade across agents, the only securities that are traded are insurance claims against ε shocks. Without loss of generality, we can therefore think of our economy as an island economy where each island is populated by agents indexed by their fixed effects (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 social 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 planner's problem, taking the aggregate fiscal variables (G, λ, τ) and the skill price $p(s)$ as given, is

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

subject to the resource constraint

$$\int_E c(\varepsilon)\, dF_\varepsilon = \lambda \int_E \exp\left[(1-\tau)\left(p(s)+\alpha+\varepsilon\right)\right] h\left(\varepsilon\right)^{1-\tau} dF_\varepsilon. \tag{A2}$$

Taking first-order conditions and substituting in the resource constraint, one obtains

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

where the first term is hours worked by the representative agent and the last term is the constant $\mathcal{M}(v_\varepsilon;\tau)/(\sigma+\tau)$. Similarly,

$$\log c = \log\lambda + \frac{1-\tau}{1+\sigma}\log(1-\tau) + (1-\tau)\left[\log p(s) + \alpha - \varphi\right] + \frac{(1-\tau)\left[1-2\tau-\sigma\tau\right]}{(\sigma+\tau)}\frac{v_\varepsilon}{2}.$$

The consumption allocation (22) has the form in $(A1)$, which confirms the no bond trade guess.

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.

A.3 Proof of Proposition 3 [skill price and skill choice]

Recall from eq. (14) in the main text that the optimality condition for skill investment is

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

The skill level s affects only the consumption allocation (not the hours allocation) and only through the price $p\left(s;\tau\right)$, which is fixed over time. Hence, (14) can be simplified as

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

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

$$\log p\left(s;\tau\right) = \pi_0(\tau) + \pi_1(\tau)\cdot s, \tag{A4}$$

which implies that the skill allocation has the form

$$s(\kappa;\tau) = (1-\tau)\,\mu\pi_1(\tau)\cdot\kappa. \tag{A5}$$

Since the exponential distribution is closed under scaling, skills inherit the exponential density shape from κ, with parameter $\zeta \equiv \frac{\eta}{(1-\tau)\mu\pi_1(\tau)}$, i.e., $m\left(s\right) = \zeta\exp\left(-\zeta s\right)$.

We now turn to the production side of the economy. Effective hours worked N are independent of skill type s (see Proposition 2). Aggregate output is therefore

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

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

$$\begin{aligned}\log p(s) &= \log\left[\frac{\partial Y}{\partial\left[N\cdot m(s)\right]}\right] = \frac{1}{\theta}\log Y - \frac{1}{\theta}\log\left[N\cdot m(s)\right] \\ &= \frac{1}{\theta}\log\left(\frac{Y}{N}\right) - \frac{1}{\theta}\log\zeta + \frac{\zeta}{\theta}s.\end{aligned} \tag{A6}$$

Equating coefficients across equations (A4) and (A6) implies $\pi_1(\tau) = \frac{\zeta}{\theta} = \frac{1}{\theta}\frac{\eta}{\mu(1-\tau)\pi_1(\tau)}$, which yields $\pi_1(\tau) = \sqrt{\frac{\eta}{\theta\mu(1-\tau)}}$ and thus

$$m(s) = \sqrt{\frac{\eta\theta}{(1-\tau)\mu}}\exp\left(-\sqrt{\frac{\eta\theta}{(1-\tau)\mu}}s\right). \tag{A7}$$

Similarly, the base skill price is

$$\pi_0(\tau) = \frac{1}{\theta}\log\left(\frac{Y}{N}\right) - \frac{1}{2\theta}\log\left(\frac{\eta\theta}{(1-\tau)\mu}\right). \tag{A8}$$

We derive a fully structural expression for $\pi_0(\tau)$ below in the proof of Corollary 3.2.

A.4 Proof of Corollary 3.1 [distribution of skill prices]

From eq. (A5) and the expression for $\pi_1(\tau)$, the skill premium for an agent with ability κ is

$$\pi_1(\tau)\cdot s(\kappa;\tau) = \pi_1(\tau)\cdot(1-\tau)\mu\pi_1(\tau)\cdot\kappa = \frac{\eta}{\theta\mu(1-\tau)}(1-\tau)\mu\cdot\kappa = \frac{\eta}{\theta}\cdot\kappa,$$

and thus skill premia are exponentially distributed with parameter θ.

The variance of log skill prices is

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

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.

A.5 Proof of Corollary 3.2 [aggregate quantities]

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

$$\begin{aligned}H(\tau) &= \int\int\int h(\varphi,\varepsilon;\tau)\,dF_\varphi dF_\alpha dF_\varepsilon, \\ N(s;\tau) &= N(\tau) = \int\int\int \exp(\alpha+\varepsilon)h(\varphi,\varepsilon;\tau)\,dF_\varphi dF_\alpha dF_\varepsilon.\end{aligned}$$

Using the expression for individual hours in Proposition 2 and integrating over the normal distributions for φ, α, and ε gives

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

Aggregate output is equal to aggregate labor earnings

$$\begin{aligned} Y(\tau) &= \int\int\int\int p(s;\tau)\exp(\alpha+\varepsilon)h(\varphi,\varepsilon;\tau)\,dF_s dF_\varphi dF_\alpha dF_\varepsilon \\ &= \int p(s;\tau)\,dF_s \cdot N(\tau) \\ &= \frac{\theta}{\theta-1}\cdot\exp(\pi_0(\tau))\cdot N(\tau), \end{aligned} \tag{A9}$$

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

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

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

$$\begin{aligned} Y &= \left\{\int_0^\infty \left[N\cdot\zeta\exp(-\zeta s)\right]^{\frac{\theta-1}{\theta}}ds\right\}^{\frac{\theta}{\theta-1}} \\ &= N\cdot\sqrt{\frac{\eta\theta}{(1-\tau)\mu}}\left(\left(\frac{\theta-1}{\theta}\right)\sqrt{\left(\frac{\eta\theta}{(1-\tau)\mu}\right)}\right)^{-\frac{\theta}{\theta-1}}. \end{aligned} \tag{A10}$$

Comparing this equation to eq. (A9) it is immediate that

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

which is the expression reported in Proposition 3.

A.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 variable λ. Second, we substitute the allocations into the social welfare function and show how to obtain eq. (29).

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

$$\lambda = \frac{(1-g)\,Y}{\tilde{Y}}. \tag{A11}$$

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 :

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

Substituting in the hours allocation (21), the expression for the skill price (23), the density function $m(s)$ (A7), and integrating, we arrive at

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

where

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

Note that

$$\begin{aligned}&\int_0^\infty \exp\left[(1-\tau)\,\pi_1(\tau)s\right] m(s)\, ds \\ &= \int_0^\infty \exp\left((1-\tau)\left(\frac{\eta}{\theta\mu(1-\tau)}\right)^{\frac{1}{2}} s\right)\left(\frac{\eta\theta}{\mu(1-\tau)}\right)^{\frac{1}{2}} \exp\left(-\left(\frac{\eta\theta}{\mu(1-\tau)}\right)^{\frac{1}{2}} s\right) ds \\ &= -\frac{\theta}{(1-\tau)-\theta},\end{aligned}$$

and recall that

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

Now sum across age groups to obtain

$$\begin{aligned}\tilde{Y} &= (1-\delta)\sum_{a=0}^{\infty}\delta^a\tilde{Y}^a = \mathcal{K}\times(1-\delta)\sum_{a=0}^{\infty}\delta^a \exp\left(-\tau\left(1-\tau\right)\frac{v_\alpha^a}{2}\right) \\ &= \mathcal{K}\times\frac{(1-\delta)\exp\left(-\tau(1-\tau)\frac{v_\alpha^0}{2}\right)}{1-\delta\exp\left(\frac{-\tau(1-\tau)}{2}v_\omega\right)}. \qquad \text{(A12)}\end{aligned}$$

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

$$\begin{aligned}\log\lambda &= \log(1-g) + \frac{\tau(1-\tau)}{\sigma+\tau}\left(\frac{1+\sigma}{\sigma+\tau}+2+\sigma\right)\frac{v_\varepsilon}{2} + \frac{\tau}{1+\sigma}\log(1-\tau) \qquad \text{(A13)} \\ &\quad +\tau(1-\tau)\frac{v_\varphi}{2} \\ &\quad -\log(1-\delta) + \tau(1-\tau)\frac{v_\alpha^0}{2} + \log\left[1-\delta\exp\left(\frac{-\tau(1-\tau)}{2}v_\omega\right)\right] \\ &\quad +\frac{1}{2}\frac{\tau}{\theta-1}\log\left(\frac{1-\tau}{\theta}\right) + \frac{1}{2}\frac{\tau}{\theta-1}\log\left(\frac{\mu}{\eta}\right) + \frac{\theta-1+\tau}{\theta-1}\log\left(\frac{\theta}{\theta-1}\right) + \log\left(\frac{\theta-1+\tau}{\theta}\right).\end{aligned}$$

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

$$\begin{aligned} u(c_a, h, G) &= \log\lambda + \frac{1-\tau}{1+\sigma}\log(1-\tau) - (1-\tau)\varphi + \frac{1-\tau}{2(\theta-1)}\left(\log(1-\tau) + \log\left(\frac{\theta}{(\theta-1)^2}\frac{\mu}{\eta}\right)\right) \\ &\quad + \mathcal{M} - \exp\left(-\frac{1+\sigma}{\widehat{\sigma}(1-\tau)}\mathcal{M}\right)\exp\left(\frac{1+\sigma}{\widehat{\sigma}}\varepsilon\right)\frac{(1-\tau)}{(1+\sigma)} + (1-\tau)\kappa\frac{\eta}{\theta} + \chi\log G + (1-\tau)\alpha_a. \end{aligned}$$

The disutility cost from investing in education is

$$v(s(\kappa)) = -\frac{1}{\kappa}\frac{s(\kappa;\tau)^2}{2\mu} = -(1-\tau)\kappa\frac{\eta}{2\theta}. \tag{A14}$$

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

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

where

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

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

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

Substituting in the expressions above for λ and Y (A13andA10) gives

$$\begin{aligned} \bar{u} &= \log(1-g) + \chi\log g + (1+\chi)\frac{\log(1-\tau)}{(1+\hat{\sigma})(1-\tau)} - \frac{1}{(1+\hat{\sigma})} \\ &\quad + (1+\chi)\left[\frac{\log(1-\tau)}{2(\theta-1)} + \frac{1}{2(\theta-1)}\log\left(\frac{\mu}{\eta}(\theta-1)^{-2\theta}\theta^{(2\theta-1)}\right)\right] \\ &\quad - \left[-\log\left(1-\left(\frac{1-\tau}{\theta}\right)\right) - \left(\frac{1-\tau}{\theta}\right)\right] - (1-\tau)^2\frac{v_\varphi}{2} \\ &\quad + \log\left(\frac{1-\delta\exp\left(\frac{-\tau(1-\tau)}{2}v_\omega\right)}{1-\delta}\right) + (1+\chi)\left[\frac{1}{\hat{\sigma}}v_\varepsilon - \sigma\frac{1}{\hat{\sigma}^2}\frac{v_\varepsilon}{2}\right]. \end{aligned} \tag{A15}$$

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

$$\bar{c}_Y = \frac{1-\tau}{2\theta},$$

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

$$\bar{c}_o = \frac{1-\tau}{2\theta} - \frac{1-\tau_{-1}}{2\theta}$$

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,

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

Now

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

so

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

and thus

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

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

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

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

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

The term in v_ω is

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

Now we can add together the contributions to social welfare from agents born before and after the reform. The term in $\bar{u}$ is simply $\bar{u}$. The terms in $\bar{c}_Y$ and $\bar{c}_0$ simplify to give $-\left(\frac{1-\tau}{2\theta}\right)+\frac{\beta\delta}{\gamma}\frac{(1-\gamma)}{(1-\beta\delta)}\frac{1}{2\theta}(1-\tau_{-1})$. 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 5.2.

A.7 Proofs of Corollaries 4.1, 4.3, 4.4, 4.5

4.1: In eq. (29), the term in τ_{-1} is additively separable from all the others containing g and τ.
4.3: Differentiating eq. (29) 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. (30).
4.4: Differentiating eq. (29) with respect to τ, the first-order condition has no terms involving g. Thus, the optimal choice for τ is independent of g.
4.5: The parameters μ and η only appear in an additively separable constant in eq. (29). Thus, these parameters do not appear in the first-order conditions defining the optimal choices for g and τ.

A.8 Proof of Corollary 4.2 [concavity]

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 τ.

A.9 Proof of Corollary 4.6 [$\gamma = \beta$ case]

In eq. (28) 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

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

where $\left(1+\beta+\beta^2+...\right)$ 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

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

where the term $(1-\delta)\,\delta$ is the size of the population at age 1. And so on.

Now we need to compute how education costs factor into social welfare. Education costs for the new and future cohorts are

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

If education 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 $\frac{1}{2\kappa\mu}\left(s(\kappa;\tau)^2 - s(\kappa;\tau_{-1})^2\right)$. Thus, the contribution to social welfare from net skill investments from cohorts who entered the economy in the past is

$$\begin{aligned}
&(1-\beta)\frac{1-\delta}{1-\beta\delta}\cdot\sum_{a=1}^{\infty}\left(\frac{\beta\delta}{\beta}\right)^a\mathbb{E}\left[\frac{1}{2\kappa\mu}\left(s(\kappa;\tau)^2 - s(\kappa;\tau_{-1})^2\right)\right] \\
= \;& (1-\beta)\frac{\delta}{1-\beta\delta}\cdot\left(\mathbb{E}\left[\frac{1}{\kappa}\frac{s(\kappa;\tau)^2}{2\mu}\right] - \mathbb{E}\left[\frac{1}{\kappa}\frac{s(\kappa;\tau_{-1})^2}{2\mu}\right]\right).
\end{aligned}$$

Adding the two pieces gives

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

Adding up these various welfare components gives the expression for social welfare in eq. (31).

A.10 Proof of Corollary 4.7 [efficiency in the RA model]

The planner's problem in the representative agent economy is

$$\begin{aligned} & \max_{C,H,g} \left\{ \log C - \frac{H^{1+\sigma}}{1+\sigma} + \chi \log (gY) \right\} \\ & s.t. \\ Y &= H \\ C &= (1-g)Y, \end{aligned} \tag{A16}$$

where the first constraint is the production function and the second is the feasibility constraint. Substituting these two constraints into the objective and taking first-order conditions gives $g^* = \chi/(1+\chi)$ and $\log H^* = \frac{1}{1+\sigma} \log (1+\chi)$. Comparing this first-best expression for hours with the competitive equilibrium allocation in eq. (18), $\log H^{RA}(\tau) = \frac{1}{1+\sigma} \log(1-\tau)$, it is immediate that $\tau^* = -\chi$ implements the first best.

A.11 Proof of Proposition 5 [condition for progressivity]

Assume $\gamma = \beta$ and approximate the sixth line of the social welfare expression (eq. 29) by $-(1-\tau)^2 \frac{v_\alpha}{2}$. Then the derivative of the social welfare expression with respect to τ is

$$\begin{aligned} \frac{\partial \mathcal{W}(g,\tau;\tau_{-1})}{\partial \tau} &= -\frac{(1+\chi)}{2(\theta-1)(1-\tau)} + \frac{1}{\theta-1+\tau} - \frac{1}{2\theta} - \frac{(\tau+\chi)}{(\sigma+1)(1-\tau)} \\ &\quad + (1-\tau)(v_\varphi + v_\alpha) - \tau \frac{(1+\sigma)^2}{(\sigma+\tau)^3} v_\varepsilon (1+\chi). \end{aligned}$$

We want to sign this derivative at $\tau = 0$ in order to ascertain whether a marginal increase in progressivity is welfare improving:

$$\frac{\partial \mathcal{W}(g,\tau;\tau_{-1})}{\partial \tau}|_{\tau=0} = -\frac{1+\chi}{2(\theta-1)} + \frac{1}{\theta-1} - \frac{1}{2\theta} - \frac{\chi}{\sigma+1} + v_\varphi + v_\alpha.$$

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

A.12 Proof of Proposition 6 [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.

Given $\gamma = \beta$, social welfare is equal to average period utility in cross section (recall Corollary 4.6). Moreover, absent uninsurable life-cycle shocks, expected 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

$$\max \int\int\int \zeta(\varphi,\kappa,\varepsilon)\left\{\log c(\varphi,\kappa,\varepsilon) - \frac{\exp\left[(1+\sigma)\varphi\right]}{1+\sigma}h(\varphi,\kappa,\varepsilon)^{1+\sigma} - \frac{s(\varphi,\kappa,\varepsilon)^2}{2\mu\kappa}\right\} dF_\kappa dF_\varphi dF_\varepsilon$$

subject to

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

where effective hours by skill type z is given by

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

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

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

where ψ is a constant.

Now the average Pareto weight must equal one:

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

Thus,

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

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

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

Thus, 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:

$$\begin{aligned} \max_{\{c(\kappa), s(\kappa), s'(\kappa), \lambda, \psi(\kappa)\}} \Lambda &= \int \zeta(\kappa) \left\{ \log c(\kappa) - \frac{s(\kappa)^2}{2\mu\kappa} \right\} dF_\kappa \\ &\quad + \lambda \left\{ N \cdot \left[\int_0^\infty f_\kappa(\kappa)^{\frac{\theta-1}{\theta}} s'(\kappa)^{\frac{1}{\theta}} d\kappa \right]^{\frac{\theta}{\theta-1}} - \int c(\kappa) dF_\kappa \right\} \\ &\quad + \int_0^\infty \psi(\kappa) \left[s(\kappa) - \left(\int_0^\kappa s'(x)dx + s(0) \right) \right] d\kappa, \end{aligned}$$

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

$$-\frac{s(\kappa)}{\mu\kappa}\zeta(\kappa)f(\kappa) + \psi(\kappa) = 0$$

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

which we can rewrite as

$$\begin{aligned} \lambda N^{1-\frac{1}{\theta}} \frac{\theta}{\theta-1} Y^{\frac{1}{\theta}} f_\kappa(\kappa)^{\frac{\theta-1}{\theta}} \frac{1}{\theta} s'(\kappa)^{\frac{1-\theta}{\theta}} &= \int_\kappa^\infty \psi(x)dx \\ \frac{1}{\theta-1} \left(\frac{Y}{N} \frac{s'(\kappa)}{f_\kappa(\kappa)} \right)^{\frac{1-\theta}{\theta}} &= \int_\kappa^\infty \frac{s(x)}{\mu x} \zeta(x) f_\kappa(x) dx. \end{aligned}$$

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

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

$$s(\kappa) = \sqrt{\frac{\eta\mu}{\theta}}\kappa.$$

Substituting these into the first-order condition gives

$$\begin{aligned}
\frac{1}{\theta-1}\left(\frac{Y}{N}\frac{\sqrt{\frac{\eta\mu}{\theta}}}{f_\kappa(\kappa)}\right)^{\frac{1-\theta}{\theta}} &= \int_\kappa^\infty \left\{\frac{\sqrt{\frac{\eta\mu}{\theta}}x}{\mu x}\exp\left(\frac{\eta}{\theta}x - \log\left(\theta/(\theta-1)\right)\right)\eta\exp\left(-\eta x\right)\right\}dx \\
\frac{Y}{N} &= \left(\frac{\theta}{\theta-1}\right)^{\frac{\theta}{\theta-1}}\theta^{\frac{1}{2}\left(\frac{-1}{\theta-1}\right)}\left(\frac{\eta}{\mu}\right)^{\frac{1}{2}\left(\frac{-1}{\theta-1}\right)},
\end{aligned}$$

which is exactly the expression for productivity per efficiency unit of labor supply in the text (eq. 34). Thus the planner's first-order condition is satisfied at the competitive equilibrium allocation.

A.13 Proof of Corollary 6.1 [nonutilitarian welfare]

Given $\gamma = \beta$, the contribution to social welfare from a particular individual of age j can be written as

$$u_j\left(\varphi, \alpha_j, \kappa, \varepsilon\right) = (1-\delta)\delta^j\left[\begin{array}{c}\tilde{u} - (1-\tau)\varphi + (1-\tau)\alpha_j + (1-\tau)\frac{\eta}{\theta}\kappa \\ -\frac{1-\tau}{1+\sigma}\exp\left(\frac{(1+\sigma)}{\hat{\sigma}}\left(\varepsilon - \frac{\mathcal{M}}{(1-\tau)}\right)\right) - (1-\tau)\kappa\frac{\eta}{2\theta}\end{array}\right],$$

where $\tilde{u}$ collects all terms that are common across agents.[51] In this expression, the terms $-(1-\tau)\varphi$ and $(1-\tau)\alpha_j$ capture the impact of preference heterogeneity and uninsurable productivity shocks on log consumption. The term $(1-\tau)\kappa\frac{\eta}{\theta} = (1-\tau)\pi_1(\tau)s(\kappa;\tau)$ captures the impact of skill investment on log consumption. Disutility from hours worked is given by the next term, which varies with ε but is independent of φ. The final term is the disutility cost of gross skill investment and is equal to $-s(\kappa;\tau)^2/(\kappa 2\mu)$ (note that we again ignore the fact that net investment may be smaller for agents who entered before date 0).

Given planner weights $\xi\left(\varphi,\kappa\right) = \frac{\theta-1}{\theta}\exp\left(-\varphi + \frac{\eta}{\theta}\kappa\right)$, the contribution to social welfare from all agents of age j is $(1-\delta)\delta^j\tilde{u}$ plus

$$(1-\delta)\delta^j\int\int\int\int \xi\left(\varphi,\kappa\right)(1-\tau)\left\{-\varphi + \alpha_j + \frac{\eta}{2\theta}\kappa - \frac{1}{1+\sigma}\exp\left[\frac{(1+\sigma)}{\hat{\sigma}}\left(\varepsilon - \frac{\mathcal{M}}{(1-\tau)}\right)\right]\right\}\,dF_\varphi dF_\kappa dF_\alpha^j dF_\varepsilon.$$

We evaluate the integral piece by piece.

The term in $-\varphi$ is

$$\begin{aligned}
&(1-\delta)\delta^j\frac{\theta-1}{\theta}(1-\tau)\int\int\int \exp\left(-\varphi + \frac{\eta}{\theta}\kappa\right)(-\varphi)\;dF_\varphi dF_\kappa dF_\alpha^j \\
= \;& (1-\delta)\delta^j(1-\tau)\int \exp\left(-\varphi\right)(-\varphi)\;dF_\varphi \;=\; (1-\delta)\delta^j(1-\tau)\frac{v_\varphi}{2},
\end{aligned}$$

[51] For the sake of simplicity, we assume a degenerate skill distribution at zero at the date of reform, which allows us to ignore the last term in the expression for social welfare involving τ_{-1} (see eq. 31). Note that this assumption has no impact on the optimal choice for τ, given reversible skill investment.

where the last equality uses a standard formula for Gaussian integrals.

The term in α_j is

$$(1-\delta)\delta^j\frac{\theta-1}{\theta}(1-\tau)\int\int\int\exp\left(-\varphi+\frac{\eta}{\theta}\kappa\right)\alpha_j\,dF_\varphi dF_\alpha^j dF_\kappa=-(1-\delta)\delta^j(1-\tau)\frac{jv_\omega}{2}.$$

The term in κ is

$$\begin{aligned}&(1-\delta)\delta^j\frac{\theta-1}{\theta}(1-\tau)\int\int\exp\left(-\varphi+\frac{\eta}{\theta}\kappa\right)\frac{\eta}{2\theta}\kappa\,dF_\varphi dF_\kappa\\ =\;&(1-\delta)\delta^j\frac{\theta-1}{\theta}(1-\tau)\int\exp\left(\frac{\eta}{\theta}\kappa\right)\frac{\eta}{2\theta}\kappa\;(\eta\exp(-\eta\kappa))\,d\kappa\\ =\;&(1-\delta)\delta^j\frac{(1-\tau)}{2(\theta-1)}.\end{aligned}$$

The term in ε is simply $-(1-\delta)\delta^j\frac{1-\tau}{1+\sigma}$.

Collecting terms, the contribution to social welfare from agents of age j is therefore

$$(1-\delta)\delta^j\tilde{u}+(1-\delta)\delta^j(1-\tau)\left(-\frac{1}{(1+\sigma)}-\frac{jv_\omega}{2}+\frac{v_\varphi}{2}+\frac{1}{2(\theta-1)}\right).$$

The corresponding expression for the utilitarian social planner was

$$(1-\delta)\delta^j\tilde{u}+(1-\delta)\delta^j(1-\tau)\left(-\frac{1}{(1+\sigma)}-\frac{jv_\omega}{2}-\frac{v_\varphi}{2}+\frac{1}{2\theta}\right).$$

Thus,

$$\mathcal{W}^{ins}(g,\tau;\tau_{-1})=\mathcal{W}(g,\tau;\tau_{-1})+(1-\tau)v_\varphi+(1-\tau)\left(\frac{1}{2(\theta-1)}-\frac{1}{2\theta}\right).$$

A.14 Proof of Proposition 7 [agreement on G]

Expected utility for an individual with state $(\varphi,\alpha,\kappa,\varepsilon)$ can be written as

$$\begin{aligned}U(\varphi,\alpha,\kappa,\varepsilon;g,\tau)\;=\;&-\frac{1}{\kappa}\frac{s(\kappa;\tau)^2}{2\mu}+(1-\beta\delta)\,\mathbb{E}\sum_{a=0}^{\infty}(\beta\delta)^a\left[\log c(\varphi,\alpha,\kappa;g,\tau)\right.\\ &\left.-\frac{\exp[(1+\sigma)\varphi]}{1+\sigma}h(\varphi,\varepsilon;\tau)^{1+\sigma}+\chi\log(gY(\tau))\right],\end{aligned}\tag{A17}$$

where we have made explicit the facts that in equilibrium, skill investment s, hours h, and aggregate output Y are independent of g.

Recall from eq. (22) that

$$\log c(\varphi,\alpha,s^*(\kappa;\tau);g,\tau)=\log\lambda(g,\tau)+\frac{1-\tau}{1+\sigma}\log(1-\tau)+(1-\tau)\left[\log p(s(\kappa;\tau))+\alpha-\varphi\right]+\mathcal{M}(v_\varepsilon;\tau),$$

where, from eq. (A13),

$$\log\lambda(g,\tau)=\log(1-g)+\Lambda(\tau).$$

If the individual with state $(\varphi, \alpha, \kappa, \varepsilon)$ were able to dictate the choice for g, taking τ as given, his optimal choice would satisfy the following first-order condition:

$$\frac{1}{c}\frac{\partial c}{\partial \lambda}\frac{\partial \lambda}{g} + \chi\frac{1}{g} = 0 \tag{A18}$$

or

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

which implies the desired result

$$g = \frac{\chi}{1+\chi}.$$

A.15 Proof of Proposition 8 [median voter]

Substituting the allocations $s(\kappa;\tau)$, $h(\varphi,\varepsilon;\tau)$, and $c(\varphi,\alpha,s;g,\tau)$ into expected utility $(A17)$ yields

$$\begin{aligned} U &= -(1-\tau)\frac{\eta\kappa}{2\theta} + (1-\beta\delta)\,\mathbb{E}\sum_{j=0}^{\infty}(\beta\delta)^j\left\{\log\lambda(g,\tau) + \frac{\log(1-\tau)}{(1+\hat{\sigma})} + \mathcal{M} + (1-\tau)\left[\alpha_j - \varphi + \kappa\frac{\eta}{\theta}\right]\right. \\ &\quad \left. + \frac{(1-\tau)}{2(\theta-1)}\left(\log(1-\tau) + \log\left(\frac{\theta}{(\theta-1)^2}\frac{\mu}{\eta}\right)\right) - \frac{(1-\tau)}{1+\sigma}\exp\left(-\frac{1+\sigma}{\widehat{\sigma}(1-\tau)}\mathcal{M} + \frac{1+\sigma}{\widehat{\sigma}}\varepsilon\right)\right\} \\ &\quad + \chi\log G. \end{aligned}$$

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,

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

Recall that the baseline social welfare function is

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

Therefore, we can express expected utility as

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

Note that $U(\varphi,\alpha,\kappa;g,\tau)$ is strictly concave in τ, since $\mathcal{W}(g,\tau)$ is concave in τ and the additional terms in $U(\varphi,\alpha,\kappa;g,\tau)$ are linear in τ.

We need to determine the median voter. A useful property is that the three individual states (α,φ,κ) enter as a linear combination. Let

$$x=\alpha-\varphi+\frac{\eta}{2\theta}\kappa.$$

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.

A.16 Proof of Proposition 9 [welfare with progressive cons. tax]

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

$$\lambda\int c(\varepsilon)^{\frac{1}{1-\tau}}dF_\varepsilon=\int_E \exp\left(p(s)+\alpha+\varepsilon\right)h(\varepsilon)\,dF_\varepsilon\,. \tag{A19}$$

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

A.17 Proof of Proposition 10 [equilibrium prices with skill bias]

We guess and will verify that equilibrium skill prices are still linear in s, i.e., $p(s)=\pi_0+\pi_1 s$. As in Proposition 3, the optimal skill choice is then given by $s=\kappa\eta(1-\tau)\pi_1$ and the cumulative distribution for skills $F_s(S)=\Pr(s\le S)=1-\exp\left(-\frac{1}{(1-\tau)\pi_1}S\right)$. This is the CDF of an exponential distribution with parameter $\zeta=1/[(1-\tau)\pi_1]$. The density function is then given by $f_s(s)=\zeta\exp(-\zeta s)$. This allows us to compute output per effective hour:

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

Below we verify that $\frac{(\theta-1)}{\theta}\zeta>\varrho$, so output is finite.

Substituting the expressions for Y/N and $m(s)$ into eq. (52) and taking logs yields

$$\log p(s)=\varrho s+\frac{1}{\theta}\log\left(\zeta\left(\frac{(\theta-1)}{\theta}\zeta-\varrho\right)^{-\frac{\theta}{\theta-1}}\right)-\frac{1}{\theta}\log[\zeta\exp(-\zeta s)].$$

Substituting the expression for ζ gives

$$\log p(s)=-\frac{1}{\theta-1}\log\left(\frac{(\theta-1)}{\theta}\frac{1}{(1-\tau)\pi_1}-\varrho\right)+\left(\frac{1}{\theta}\frac{1}{(1-\tau)\pi_1}+\varrho\right)s.$$

Note that the logarithm of the marginal product of skill type s is linear in s, confirming the guess that $p(s)$ is linear. Matching up coefficients, i.e., setting $\pi_1 = 1/(\theta(1-\tau)\pi_1) + \varrho$, implies[52]

$$\pi_1 = \frac{\varrho}{2} + \sqrt{\left(\frac{\varrho}{2}\right)^2 + \frac{1}{\theta(1-\tau)}} \geq 0.$$

Note that Assumption 1 ensures that π_1 is real.

Using the solution for π_1 and equating coefficients for the constant yields

$$\pi_0 = -\frac{1}{\theta - 1}\log\left(\frac{(\theta-1)}{\theta}\frac{\eta}{\mu(1-\tau)\mathcal{T}} - \varrho\right),$$

where

$$\mathcal{T} = \frac{\varrho}{2} + \sqrt{\left(\frac{\varrho}{2}\right)^2 + \frac{1}{\theta(1-\tau)}}.$$

Note that the logarithm of the skill premium is $\pi_1 \cdot s = \eta(1-\tau)(\pi_1)^2\kappa$. It follows that the equilibrium distribution of log skill premia, $\pi_1(\tau)\cdot s$, is exponential with parameter $\zeta = (1-\tau)^{-1}(\pi_1(\tau))^{-2}$.

Finally, we verify that $\varrho < \frac{(\theta-1)}{\theta}\zeta$. Given π_1 we can characterize in terms of structural parameters the condition under which

$$\varrho < \frac{\theta-1}{\theta}\zeta = \frac{(\theta-1)}{\theta}\frac{1}{(1-\tau)\mathcal{T}},$$

Now note that the left-hand side is increasing in ϱ while the right-hand side is decreasing in ϱ. So setting the two sides equal defines an upper bound for ϱ satisfying the inequality. It is straightforward to verify that this upper bound is given by

$$\bar{\varrho} = \frac{\theta-1}{\theta}\frac{1}{\sqrt{1-\tau}}.$$

Since $\tau \geq -1$, it is clear that Assumtion 1 ensures that $\varrho \leq \bar{\varrho}$.

With skill bias in production ($\varrho \neq 0$), the skill-related terms in the equally weighted social welfare function $\mathcal{W}(g, \tau; \tau_{-1})$ (second, third and fourth lines in eq. 29) become as follows (abstracting from the term in τ_{-1}):

$$\begin{aligned}
&-(1+\chi)\log\left[(1-\tau)\mathcal{T}\right] - (1+\chi)\frac{\theta}{\theta-1}\log\left[\frac{(\theta-1)}{\theta}\frac{1}{(1-\tau)\mathcal{T}} - \varrho\right] \\
&+\log\left[1-(1-\tau)^2\mathcal{T}^2\right] + (1-\tau)^2\mathcal{T}^2 \\
&-\frac{1}{2}(1-\tau)^2\mathcal{T}^2.
\end{aligned}$$

Note that in order for social welfare to be finite (i.e., larger than $-\infty$) consumption dispersion must not be too large. In particular, the term in the logarithm (in the second line) must be positive, which implies a lower bound on τ given θ and ϱ :

$$1-(1-\tau)^2\mathcal{T}^2 > 0 \Rightarrow \tau > 1 - \frac{1}{\frac{1}{\theta} + \varrho}.$$

[52]Note that since $s \geq 0$, we can rule out the second root, which is always negative: $\pi_1 = \varrho/2 - \sqrt{(\varrho/2)^2 + 1/(\theta(1-\tau))} < 0$.

Finally, consider the special case in which $\theta \to \infty$. Then $\mathcal{T} = \varrho$ and the terms above simplify to

$$
\begin{aligned}
&- (1+\chi) \log\left[(1-\tau)\,\varrho\right] - (1+\chi) \log\left(\frac{1}{(1-\tau)\,\varrho} - \varrho\right) \\
&+ \log\left[1 - (1-\tau)^2\,\varrho^2\right] + (1-\tau)^2\,\varrho^2 - \frac{1}{2}\,(1-\tau)^2\,\varrho^2.
\end{aligned}
$$