How Should Tax Progressivity Respond to Rising Income Inequality?*

Jonathan Heathcote Kjetil Storesletten Giovanni L. Violante

October 1, 2020

Abstract

We address this question in a heterogeneous-agent incomplete-markets model featuring exogenous idiosyncratic risk, endogenous skill investment, and flexible labor supply. The tax and transfer schedule is restricted to be log-linear in income, a good description of the US system. Rising inequality is modeled as a combination of skill-biased technical change and growth in residual wage dispersion. When facing shifts in the income distribution like those observed in the US, a utilitarian planner chooses higher progressivity in response to larger residual inequality but lower progressivity in response to widening skill price dispersion reflecting technical change. Overall, optimal progressivity is approximately unchanged between 1980 and 2016. We document that the progressivity of the actual US tax and transfer system has similarly changed little since 1980, in line with the model prescription.

Keywords: Optimal Taxation; Redistribution; Tax Progressivity; Income Distribution; Skill Investment; Labor Supply; Incomplete Markets; Skill-biased Technical Change; Inequality.

JEL Codes: D30, E20, H20, I22, J22, J24.

*Heathcote: Federal Reserve Bank of Minneapolis and CEPR, heathcote@minneapolisfed.org; Storesletten: University of Oslo and CEPR, kjetil.storesletten@econ.uio.no; Violante: Princeton University, CEBI, CEPR, IFS, IZA and NBER, violante@princeton.edu. This paper was presented as Kjetil Storesletten's Presidential Address at the Meeting of the European Economic Association in Manchester, August 2019. 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. We thank Johannes Fleck for research assistance and useful comments. Storesletten thanks Oslo Fiscal Studies for support during this project.

1 Introduction

Income inequality has risen sharply in the US over the last four decades. At the heart of this trend is a marked widening of the wage distribution. Figure 1 shows that less than half of the rise in wage inequality occurred between demographic groups based on age and education. The bulk of this upward trend is within group, or residual. Income inequality rose across many developed economies over the same period. The magnitude of this phenomenon, however, has been much more pronounced in Anglo-Saxon countries, such as the US and the UK, than in Continental Europe, Japan, and the Nordic countries (see, for example, Table 3 in Krueger et al., 2010).

This paper asks two questions regarding the US experience. How has the government responded to this widening of the income distribution? And how should the government have responded? The natural tool for a government that wants to compress income inequality is to rely on tax and transfer policies to achieve the desired amount of redistribution. There are, of course, additional instruments that governments could employ, including investment in education and training programs; labor market regulations such as minimum wages and employment protection; and competition policies aimed at trade, market power, and migration. While these additional interventions often complement traditional government redistribution, the lion's share of redistribution in practice occurs through taxes and transfers. This fiscal lever is the focus of our paper.

One might conjecture that a utilitarian government would respond to rising inequality with more redistribution. In particular, to the extent that the government's objective in designing the tax and transfer system is trading off equity and efficiency considerations, one might expect that greater inequality would lead the government to put more emphasis on equity and to therefore choose a more progressive tax system, even if that would reduce efficiency somewhat. In contrast, the consensus view in the literature is that the US tax system has become less progressive over the past 40 years. For example, Piketty and Saez (2007), Saez and Zucman (2019), Ferriere and Navarro (2020), and Wu (forthcoming) all find declines in various measures of progressivity over time.

This apparent contradiction raises something of a puzzle and has left economists searching for explanations. One possible explanation has been that social preferences have changed over time, with the US government effectively becoming more willing to tolerate inequality (Lockwood and Weinzierl, 2016). Similarly, Saez and Zucman (2019) argue that changes in politics and ideology have led to reductions in capital taxation and to acceptance of tax evasion. We think that such explanations are not fully satisfying

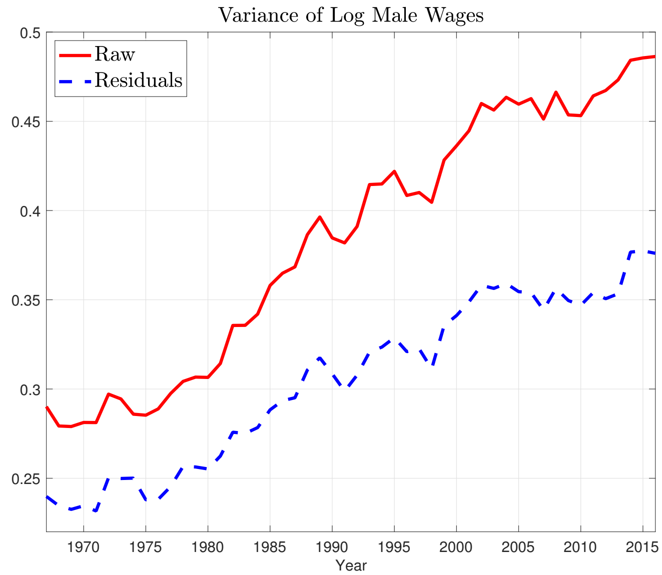

Figure 1: The figure shows variance of log hourly wages for US male workers (solid red line) and residual wage dispersion (dashed blue line). Residual inequality is the dispersion left over after controlling for years of education and a quartic in age. Source: Current Population Survey (CPS).

absent a deeper theory of the drivers of shifts in social preferences.

In this paper, we challenge on two levels the traditional narrative on inequality and redistribution. First, we argue that the tax and transfer system has *not* in fact become less progressive over time. On the contrary, we argue that the amount of redistribution embedded in the tax and transfer system has been rather stable on net between 1980 and 2016. Second, we argue that the appropriate policy to address rising inequality depends on why income inequality is going up. One needs a theory of inequality that can address the empirical patterns documented in Figure 1. When we model both the rise in returns to skills and the increase in residual wage dispersion, counteracting forces emerge in the optimal taxation problem. We conclude that progressivity should have remained roughly *stable* over time.

1.1 Our two key results in the context of the existing literature

Measurement. The literature has long recognized that measuring overall tax and transfer progressivity and its change over time is challenging. Studies that abstract from the role of transfers (e.g., Guner et al., 2014; Saez and Zucman, 2019) find the system has become less redistributive over time. Studies that focus on the dynamics of statutory marginal tax rates (e.g., Lockwood and Weinzierl, 2016; Ferriere and Navarro, 2020) also tend to estimate declines in progressivity over time. Trends in progressivity also depend on where in the household income distribution you focus: Saez and Zucman (2019) and Hubmer et al. (forthcoming) emphasize declining marginal tax rates within the top 1% of the income distribution.

Our approach is to (i) include transfers alongside taxes; (ii) measure taxes actually paid by households rather than measure statutory rates; and (iii) examine global progressivity of the fiscal system across the distribution, not just at the top. Specifically, we estimate progressivity from Congressional Budget Office (CBO) data, largely following the CBO's own definitional choices. We include most transfers in our measure of post-government income, with the exception of certain transfers in kind (Medicaid and Medicare). Our measure of taxes follows the CBO practice and includes all federal taxes, but not state and local ones, with estimates based on IRS-derived data on taxes actually paid.[1]

As in much of our previous research, we summarize the US tax and transfer system through a parsimonious log-linear relationship between gross income and disposable income, in which progressivity is a function of one parameter only (Heathcote et al., 2014, 2017, 2020b). Through the lens of this tax function, the progressivity coefficient τ is stable at 0.186 (corresponding to an income-weighted average marginal tax rate of 0.34) between the early 1980s and the mid 2010s.

Theory. On the theory side, we analyze optimal taxation within the analytical framework developed in Heathcote et al. (2017). Our Ramsey-style normative analysis restricts the search for optimal progressivity within a given parametric class of tax and transfer schemes.[2]

Our model incorporates a range of benefits of higher tax progressivity. Individuals are born unequal and face additional shocks over the life cycle, some of which cannot be insured privately. Thus a progressive tax and transfer system both provides redistribution with respect to unequal initial conditions and substitutes for missing private insurance against life-cycle shocks. The model also captures three key costs of higher tax progressivity: a static distortion to labor supply, a dynamic distortion to skill investment, and an effect on public good provision. Through our general equilibrium model, taxes and

[1] The CBO has very recently started studying how to further improve its data by allocating state and local taxes to households. See https://www.cbo.gov/publication/54685 for calculations. Preliminary results suggest that state and local income taxes are progressive, whereas consumption taxes and property taxes are regressive.

[2] We refer to Heathcote and Tsujiyama (2019) for a comparison between the solution of the optimal taxation problem under our tax function and the solution of a full Mirrlees problem. See also the survey by Stantcheva (2020).

transfers affect the level of output; the pre-tax income distribution; and, ultimately, the equilibrium distributions of consumption, labor supply, and welfare.

Most papers on progressive income taxation have focused exclusively on distortions to labor supply with an exogenous wage distribution. They all conclude that the current tax system appears to offer too little redistribution relative to what a utilitarian planner would choose (see, e.g., Saez, 2001; Heathcote and Tsujiyama, 2019; Bakis et al., 2015; Kindermann and Krueger, 2014). Rising uninsurable labor productivity dispersion in these models always calls for more rather than less redistribution. We generalize this channel, as in Heathcote et al. (2014), and argue that the data call for a share of the rise in labor market risk to be privately insurable, a force that limits the increase in optimal progressivity.

Exogenous labor market risk is only one of the channels at work in our model. The other one is endogenous skill investment. Guvenen et al. (2014), Krueger and Ludwig (2016), Findeisen and Sachs (2016), Stantcheva (2017), and Badel et al. (2020) are recent papers studying optimal taxation in models with human capital. Optimal taxation now depends on the details of how skill investment is modeled, but because progressive taxation distorts choices along an additional margin, the efficiency costs of progressivity will tend to be larger, and optimal progressivity will typically be reduced. We find that when we add this margin to the model, the optimal system for a utilitarian planner is almost identical to the system we observe for the US at the federal level in 1980.

The main analysis in this paper is a comparative static exercise. We focus on two structural shifts that have widened cross-sectional inequality in wages and earnings in the last four decades. First, we interpret the observed rise in residual wage dispersion as reflecting an increase in the variance of idiosyncratic life-cycle labor productivity shocks. The appropriate policy response here is well understood: if this extra exogenous wage dispersion is privately uninsurable, a more progressive tax system that offers more generous social insurance is called for. Second, we interpret the rise in between-group wage dispersion as reflecting an increase in the return to skills. How should this phenomenon be modeled? We consider two alternative hypotheses. The first is akin to conventional "skill-biased technical change" according to which the relative production weights on high skill labor inputs have increased over time, possibly capturing, in reduced form, increasing complementarity of high skill workers with ICT capital (Krusell et al., 2000). The second hypothesis, more novel in the literature, is that different skill types have potentially become more complementary in production over time, which – holding the distribution of skills fixed – will tend to increase the wage dispersion between worker skill types that are relatively scarce and those that are relatively abundant. One possible interpretation is that workers have become more specialized in narrower skills sets over time (see, e.g., Alon, 2018). We label this force "specialization-biased technical change".

A key message from the paper is that a utilitarian planner's optimal response to either source of increasing skill price dispersion is to reduce tax progressivity. Why is this? On the one hand, the more progressive taxation is, the lower equilibrium skill investment is, which depresses aggregate output and consumption. On the other hand, more progressive taxation reduces inequality in consumption due to differences in skills. In our baseline calibration, it turns out that the net contribution to social welfare from these two strong, but countervailing, forces is maximized at a positive but modest level of tax progressivity. When we feed in our estimated shifts to all components of the wage structure, the efficiency costs associated with distorted skill investment loom larger in the overall welfare calculations, and the model thus calls for a modest decline in optimal tax progressivity between 1980 and 2016.

This result is closely related to three recent contributions in the optimal taxation literature. Ales et al. (2015) simulate widening income inequality in a skill-to-task assignment model with an endogenous wage distribution and find that only moderate changes to the tax system are optimal. In a similar vein, Scheuer and Werning (2017) argue that when the rise of income inequality at the top of the distribution is generated through a stronger Rosen-style superstar effect, optimal taxes remain unaltered. Through a Ben-Porath style technology, Wu (forthcoming) incorporates human capital accumulation, and similarly argues that rising inequality implies declining optimal progressivity.[3]

In sum, recognizing that the wage distribution is an endogenous equilibrium object that is affected by the tax structure is paramount when thinking about the optimal design of government policy. This insight is also related to a result we obtained in previous work (Heathcote et al., 2010b, 2013). There, we showed that allowing for an endogenous wage structure when modeling the rise in US income inequality affects the quantification of its welfare implications.

The rest of the paper is organized as follows. Section 2 describes our measurement of the historical changes in the progressivity of the US tax and transfer system over the past

[3]His model also features two other changes over time that work against increasing optimal progressivity. The first, and most important, is increasing fiscal pressure on the government to raise revenue, which he models as an aging population and a rising dependency ratio. Heathcote et al. (2017) and Heathcote and Tsujiyama (2019) show that increasing the government revenue requirement lowers optimal tax progressivity. The second force he points to is rising female labor force participation. If women's labor supply is more elastic than men's, a rising share of women in the labor force increases the cost of high levels of tax progressivity.

40 years. Section 3 outlines the model. Section 4 calibrates the model and explains the key forces at work. Section 5 describes the results of our main comparative static exercise. Section 6 concludes. An online appendix contains some technical derivations.

2 Measuring tax progressivity

The US system of taxes and transfers is complex, featuring a wide array of social insurance programs and means-tested benefits and taxation at different levels of government (federal, state, and local). This makes it challenging to concisely summarize how the tax-and transfer system should respond to changes in inequality and what this response has been empirically.

Is there a way to summarize the tax and transfer system in a simplified way? In Heathcote et al. (2017), we illustrate the US tax and transfer system non-parametrically by dividing households into percentiles of pre-government income. For each household, we calculate a measure of disposable income, defined as pre-government income plus transfers minus taxes. We then calculate average disposable income for each percentile. A scatter plot of pre- versus post-government income shows that the log of post-government income is approximately a linear function of the log of pre-government income, except at the lowest income percentiles, where there is more redistribution. In sum, the following log-linear tax and transfer function yields a remarkably good fit

$$\begin{aligned} \log\left[y - T(y)\right] &= \log(\lambda) + (1-\tau)\log\left[y\right] \\ &\Rightarrow \\ y - T(y) &= \lambda y^{1-\tau}, \end{aligned} \tag{1}$$

where y is pre-government income and $T(y)$ is taxes minus transfers.

Such a log-linear tax and transfer function has long been a tradition in public economics, including in the work of Musgrave (1959); Jakobsson (1976); Kakwani (1977); and, more recently, Bénabou (2000) and Heathcote et al. (2017). The parameter λ captures the level of taxation, while the parameter τ can be interpreted as a measure of tax progressivity. To see this, note that when $0 < \tau < 1$, the tax system features *progressivity* in the sense that the marginal tax rate $T'(y)$ is larger than the average tax rate $T(y)/y$ for any positive income level $y > 0$. Conversely, when $\tau < 0$, the marginal tax rate is lower than the average tax rate, $T'(y) < T(y)/y$, implying that taxes are regressive. When $\tau = 0$, the tax system is flat, with a constant marginal tax rate, $T'(y) = T(y)/y = 1 - \lambda$, and when

$\tau = 1$, there is full redistribution ($T(y) = y - \lambda$). This tax system has a break-even income level $y^0 = \lambda^{1/\tau}$ at which point pre-government income equals post-government income (zero average tax rate).

This tax function imposes that marginal taxes are monotone in income. In reality, at the bottom of the income distribution, marginal tax rates can be high in the region where means-tested programs are phased out. Moreover, this system has no lump-sum transfers or floor for disposable income (the post-government income of those with zero pre-government income is also zero). In the US, there exist programs that guarantee a floor. An example of such programs is the Supplemental Nutrition Assistance Program (SNAP), formerly known as food stamps. For these two reasons, the log-linear fit worsens in the bottom decile of the distribution. However, for the rest of the income distribution, it offers a very good fit.

2.1 Progressivity in the US, 1979-2016

We now set out to measure the progressivity of the US tax and transfer system in line with eq. (1), and to explore how this has changed since 1979.

We use data from the CBO.[4] The CBO regularly produces reports on the distribution of income, using various data sources including the Internal Revenue Service (IRS) Statistics of Income (SOI) sample of tax returns and the Annual Social and Economic (ASEC) Supplement of the CPS. The CBO reports average pre- and post-government income for various quantiles of the income distribution for several different income concepts. We focus on their measurements where households are ranked by total income (adjusted for household size) before means-tested transfers and taxes. Income rankings are reported for various sub-groups of the US population. Given our interest in tax progressivity for households of working age, we focus on the sample of households with children and non-elderly childless households.[5] Relative to the PSID or other micro data sets, a key advantage of the CBO data is that they contain comprehensive estimates of taxes paid and transfers received.

The CBO breaks income into three broad components: (i) *market income* includes wage income, business income, capital income, and other non-governmental sources of in-

[4] In Heathcote et al. (2017), we relied on the Panel Study of Income Dynamics (PSID) and focused on households of working age with a strong connection to the labor market.

[5] These two groups are of similar size. In 2016, there were 38.5 million households with children, containing 154.5 million individuals (of which 71.7 million were below the age of 18). In the same year, there were 55.6 million non-elderly childless households, containing 103.3 million individuals (all over 18, with a head or spouse below age 65).

come (e.g., private transfers); (ii) *social insurance benefits* include Social Security, Medicare, unemployment insurance, and workers' compensation; (iii) *means-tested transfers* include Medicaid, SNAP (formerly food stamps), the Children's Health Insurance Program (CHIP), TANF (formerly ADFC), and Supplemental Security Income. The CBO uses ASEC data as the starting point for their transfer estimates but imputes transfer income to non-reporters to address under-reporting in the survey. Taxes are broken down into individual income taxes, payroll taxes, and corporate taxes, where 75% of corporate taxes are allocated in proportion to household capital income, and 25% of them are in proportion to labor income. The CBO tax measures exclude state and local taxes.

To estimate tax progressivity using eq. (1), we need to take a stand on definitions for pre-government income and post-government income. We are interested in redistribution and progressivity induced both through taxes and through transfers. Since transfers are simply negative taxes, the level of transfers and how transfers vary with income are just as important for skill investment and labor supply choices as the level and income sensitivity of taxes. However, while it is straightforward to characterize how taxes contribute to public redistribution, dealing with transfers is messier. One reason is that many important transfers, such as Medicaid and SNAP, are transfers in kind. A second reason is that the Social Security system mixes forced individual saving (simple intertemporal reallocation of income within an individual life) with redistribution across households.

The CBO's own current baseline income measure for measuring tax and transfer rates – what they label "income before taxes and transfers" – is market income plus social insurance benefits (see Perese, 2017). We will use this as our starting point for defining pre-government income and CBO "income after taxes and transfers" – which adds means-tested transfers and subtracts taxes – as the basis for post-government income.

We make two adjustments to the CBO measures of transfers by subtracting Medicare transfers from social insurance benefits and Medicaid and CHIP benefits from means-tested transfers. These are transfers in kind, which the CBO estimates based on the cost to the government of providing the benefits. For low income households, the magnitudes of these transfers are very large. For the bottom quintile of the sample of households with children, the average value of Medicaid and CHIP was $14,400 in 2016. However, these transfers do not constitute a standard notion of disposable income, and including them would paint an overly rosy picture of income at the bottom of the distribution. Moreover, if one were to include the value of free health care, one might also want to include the value of public education and other public services.[6]

[6] Public health care and public education provide benefits to their recipients, but there are also signifi-

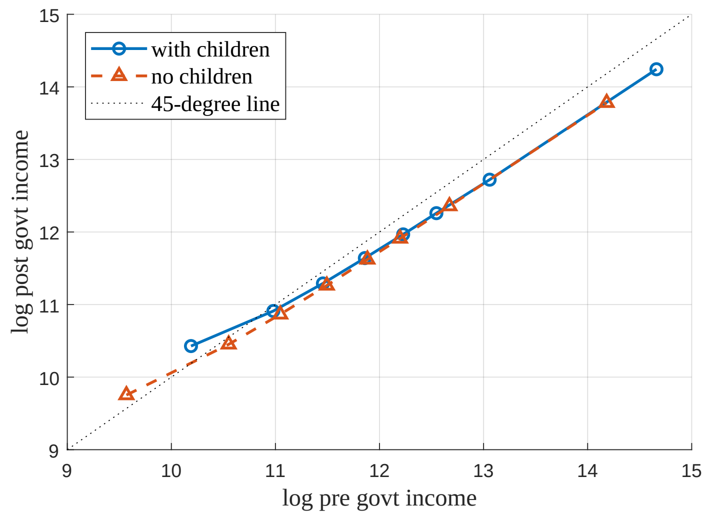

Figure 2: Pre- and post-government income for various quantiles of households with children (blue line) and non-elderly households without children (red dashed line) in 2016. Source: Congressional Budget Office (CBO).

It is debatable whether other social insurance benefits should be included in pre-government income. By far the most important component here is Social Security income. One rationale for including Social Security in pre-government income (our baseline choice) is that most of the Social Security benefits received by working-age households reflect returns to forced saving made earlier in life.[7] But we will also report results for a case in which we exclude from pre-government income all social insurance benefits (while retaining them in post-government income).

Figure 2 plots our baseline pre-government income measure against post-government income in 2016 for our two samples. Each dot corresponds to the log of average pre- and post-government income for the following eight quantiles of the income distribution, which are the ones the CBO reports: [p0-p20], [p21-p40], [p41-p60], [p61-p80], [p81-p90], [p91-95], [p96-p99], and [p100]. As is clear from the figure, the relationship between log pre-government income and log post-government income is quite close to linear above

cant positive externalities to having a healthy and well-educated population. An interesting but ambitious project would be to estimate the private value of all public services to households at different points in the income distribution and to use these estimates to generate a comprehensive measure of government redistribution.

[7]There is some redistribution embedded in the Social Security system, but it is more modest than one might think. First, the system favors married couples at the expense of singles (Groneck and Wallenius, 2020). Second, while higher income households in principle receive lower replacement rates, some of the associated redistribution is undone by the fact that higher income households tend to live and collect benefits for longer.

Table 1: Alternative Estimates for Progressivity τ

Specification	Estimates of τ 1979-83	2012-16	Change
Baseline			
With children	0.218	0.207	-0.011
Without children	0.155	0.164	+0.009
Average	0.186	0.186	-0.001
Alternative income measures			
Pre-govt. inc. = market income	0.236	0.216	-0.020
Pre-govt. inc. = post-govt. + taxes	0.089	0.109	+0.019
Alternative samples			
First quintile dropped	0.083	0.112	+0.029
Top 5% only	0.043	0.051	+0.008

Estimates based on CBO data of progressivity τ for various samples and various periods. The baseline pre-government income measure equals market income + social insurance benefits (except for Medicare). The baseline post-government income equals pre-government income plus means-tested transfers (except for Medicaid and CHIP) minus taxes.

the bottom quintile of the income distribution, consistent with the log-linear functional form for the tax and transfer system. As discussed above, for low pre-government income levels, post-government income is more generous than a linear relationship would predict. Note, finally, that the tax and transfer system offers notably more support to low income households with children than to similarly poor households without children.

Based on the quantiles in Figure 2, we estimate the progressivity parameter τ from equation (1) year by year. We take the logarithm of the average pre- and post-government income measures and estimate τ by a simple least squares regression, using weights proportional to the number of households in each quantile bin.[8] Table 1 reports the average estimated values for τ over the 1979-1983 period and the 2012-2016 period. We report estimates for the sample of households with children, the sample without children, and the simple average across the two samples, which will serve as our baseline progressivity estimate. The possibly surprising takeaway is that given our baseline income definitions, overall progressivity in the United States has not changed over these 35 years. We find that the average progressivity value is $\tau = 0.186$ in both periods.

[8]Strictly speaking, what matters for marginal household decisions is the progressivity of the *statutory* tax and transfer system. In Heathcote et al. (2017) we estimate statutory progressivity by subtracting estimates for tax deductions from pre- and post-government income measures when estimating τ. We abstract from that refinement in this paper.

Alternative specifications. The next row of Table 1 reports progressivity estimates when pre-government income excludes social insurance benefits – in particular, Social Security income. This specification delivers higher estimates for τ. The reason is that Social Security is an important income component for many low income households and excluding it reduces pre-government income disproportionately at the bottom. Since our definition for post-government income is unchanged, the tax and transfer system now appears more redistributive at the bottom, which in turn translates into higher estimates for τ. This measure of progressivity falls only very slightly over time.

Next, we ask how much progressivity is embedded in taxes alone, by defining pre-government income as market income plus social insurance benefits plus means-tested transfers (we continue to exclude from these transfers the values of Medicaid and Medicare).[9] Now, taxes are the only difference between pre- and post-government income. The estimates for τ are now around 0.10, with a modest increase in progressivity over time. The fact that the progressivity estimate substantially falls relative to the baseline value of 0.186 indicates that much of the effective progressivity in the US system operates through transfers rather than taxes.

The next row of Table 1 reports progressivity estimates using our baseline pre-government income definition, but excluding from the regression the bottom quintile of the income distribution. Individuals with strong attachment to the labor force, for whom tax progressivity mediates skill investment and labor supply choices (as in our model), are mostly above the bottom quintile of the income distribution. As can be expected from Figure 2, these estimates for τ are notably lower. Thus, the estimates for progressivity reported in the baseline specification of Table 1 reflect a compromise between trying to match the high degree of redistribution at the bottom of the US tax and transfer system and the lower degree of redistribution everywhere else. Note that the progressivity estimate increases somewhat over time when the bottom quintile is excluded.[10]

Finally, the last row of the table shows estimates for τ using only the top two quantiles reported by the CBO, involving pre and post-government income for households in the top 1% and the next 4% of the income distribution. These estimates are lower, but still positive, indicating that marginal tax rates are increasing in income even at the top of the distribution.

[9] Many papers in the public finance literature focus on progressivity from taxes alone. For example, Guner et al. (2014) report estimates for τ in 2000 of around 0.05 using IRS data. Ferriere and Navarro (2020) estimate τ to be around 0.1 after 1986.

[10] This finding is robust to the alternative measure of pre-government income: when pre-government income is defined as market income, the average estimated τ, excluding the bottom quintile, increases from 0.121 in 1979-83 to 0.148 in 2013-2017.

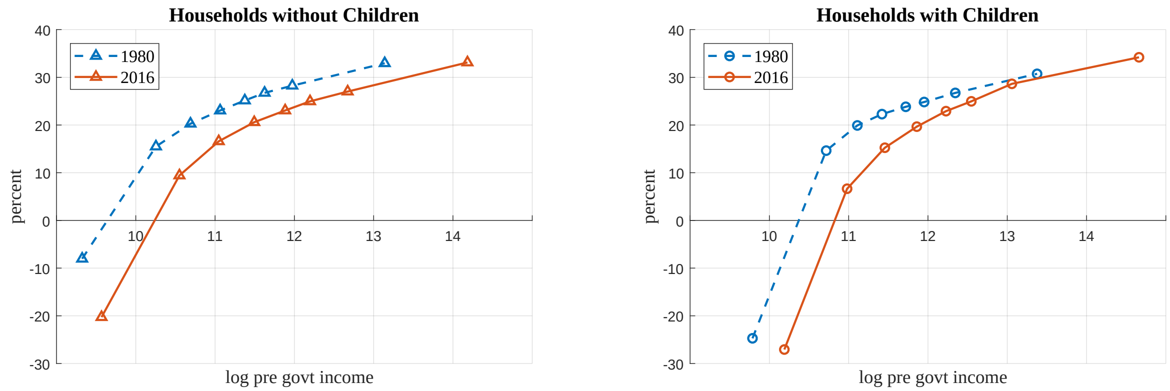

Figure 3: Average tax rates by income, 1980 and 2016. Data source: CBO.

To summarize, Table 1 indicates that the degree of progressivity has remained approximately constant since 1979. Indeed, for none of the different income measures or samples we have considered do we find evidence of economically significant changes in progressivity over time. However, one must be cautious in characterizing the extent of progressivity of the US tax and transfer system. The treatment of different components of transfers matters, and different types of households face different mixes of taxes and transfers.

Note that the estimates we have reported are for tax progressivity, not for tax rates. Figure 3 plots average net tax rates by income in 1980 and 2016, to give a feeling for the mapping between our estimates for τ and actual redistribution across the income distribution. In particular, for each income bin reported by the CBO, we report pre-government income minus post-government income (i.e., taxes net of transfers) divided by pre-government income, given our baseline income definitions.[11] The picture clearly illustrates that the US tax and transfer system is progressive. Federal taxes net of transfers are lower in 2016 relative to 1980, across the entire income distribution. Note that the dots for 2016 are generally to the right of those for 1980, indicating real income growth over this 36 year period, especially at the top of the distribution.

Net average tax rates are below 30% for all quantiles except for the very top of the income distribution. One reason is that we are measuring taxes net of transfers, which are necessarily a smaller share of income than taxes alone. In addition, recall that our calculations exclude taxes at the state and local levels. Including those would push up average tax rates. Finally, note that because the tax and transfer system is progressive,

[11] Income here is inflation-adjusted using the PCE deflator and is in 2016 dollars.

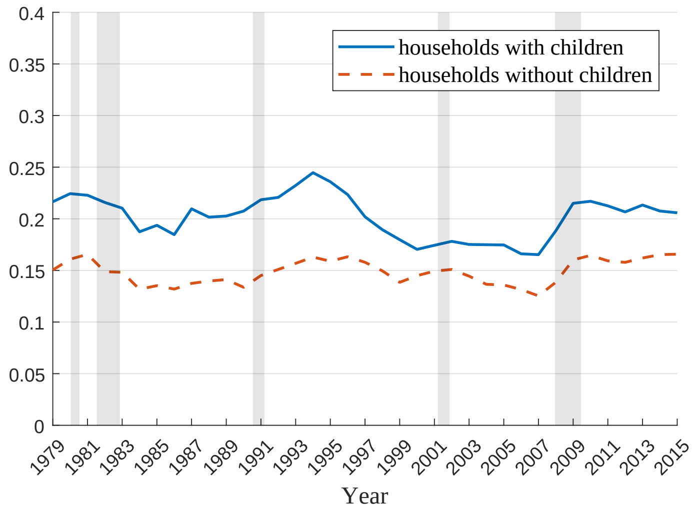

Figure 4: Estimates of progressivity (τ) of the US tax and transfer system, 1979–2016. NBER recessions are shaded. Data source: CBO.

marginal tax rates are larger than average tax rates. Given our log-linear parametric tax and transfer schedule, the average income-weighted marginal-tax rate is given by $1-(1-\tau)(1-g)$, where g is the ratio of government purchases (consumption plus investment) relative to GDP. For 1980 and 2016, the average value for this ratio was $g = 0.192$ (for federal, state, and local levels combined). Given that value and our baseline estimate for τ of 0.186, the implied (income-weighted) average marginal tax rate is 34.2%.

Low-frequency changes in progressivity. Figure 4 further explores time changes in tax progressivity, by plotting the time paths of our baseline τ estimates for every year from 1979 through 2016. The plot reinforces the message from Table 1: the overall progressivity of the tax and transfer system has been remarkably stable over time. One might expect to see more of an imprint of some of the tax reforms that occurred over this period. For example, the Tax Reform Act of 1986 lowered the top individual income tax rate from 50% in 1986 to 38.5% in 1987 and to 28% in 1988. Our estimates for progressivity, however, barely move in those years. Zooming in on the top 1% of households ranked by pre-government income, the share of income this group paid in taxes in fact changed little over these years. This share was actually higher in 1988 than in 1986, for example. The reason is that the tax reform package also modified many other provisions for deductions and exemptions, and relatively few households ever faced the top marginal rate.

High-frequency changes in progressivity. While estimated tax progressivity does not appear to have changed much over time in response to explicit tax reforms, it clearly does vary over the business cycle. In particular, measured progressivity rises in all but one recession over this period: in 1980, in 1990-91, in 2001, and in the Great Recession of 2007-09. During expansions, and especially during the late 1990s, estimated progressivity declines.[12] This cyclical variation reflects the fact that the US tax and transfer system is especially progressive at low income levels. During recessions, income declines for households experiencing unemployment, causing inequality to widen sharply at the bottom of the distribution (see Heathcote et al., 2020a). Thus, more households come to benefit from the extensive redistribution the US system delivers – primarily via transfers – to low income households, raising the estimate for τ. Between 2007 and 2010, for example, average household income before means-tested transfers and taxes for the bottom quintile of households with children fell from $26,800 to $24,500, while over the same period receipt of SNAP benefits for the same group rose from $2,000 to $3,400. During expansions, income growth at the bottom undoes this effect.

2.2 Comparison with other studies

Our finding that overall tax and transfer progressivity has changed little over the last 40 years is consistent with the narrative in Slemrod and Bakija (2017). See, for example, their Figure 3.2. In contrast, other studies in the literature have found more evidence of declining progressivity over time.

Wu (forthcoming) uses different sources than we use to estimate the progressivity parameter τ. In particular, he applies ASEC data for income and transfers and the TAXSIM model to estimate taxes. His pre-government income measure excludes all government transfers, so his estimates should be compared to our specification in which pre-government income is equal to market income. He finds a decline in τ from 0.19 in 1978-80 to 0.14 in 2014-16. Thus, while his estimate for the level for progressivity is similar to ours, he finds a decline over time, while we do not.[13]

Piketty and Saez (2007) argue that between 1960 and 2004, the US tax system became less progressive. The main distinction between their IRS-based analysis and our CBO-based one is that their interest is primarily in the very top of the income distribution (the top 0.1% and above). They also incorporate estate taxes (which are significant at

[12] The Personal Responsibility and Work Opportunity Reconciliation Act of 1996, which profoundly reformed the welfare system, may also have contributed to lower measured progressivity in this period.

[13] Cyclical fluctuations in Wu's estimated path for τ (his Figure A.1) are similar to those in our Figure 4.

the very top of the income distribution) and make different assumptions than the CBO on the incidence of corporate taxes. Similarly, Saez and Zucman (forthcoming) estimate that tax rates have declined for households within the top 1% of the income distribution (their Figure 5). But at the same time, they find (as we do) that taxes net of transfers have declined for households at the bottom of the income distribution (their Figure 6).

Ferriere and Navarro (2020) find a sharp drop in tax progressivity in 1986 (their Figure 12), with stable progressivity thereafter. Their approach is based on the idea that the progressivity parameter τ can be estimated given two inputs: an estimate for the economy-wide average tax rate and one for the average marginal tax rate. They exploit the following result. Suppose the tax and transfer function is log-linear given by equation (1). Then, the progressivity parameter τ can be estimated as

$$\tau = \frac{\text{(Average Marginal Tax Rate)} - \text{(Average Tax Rate)}}{1 - \text{(Average tax rate)}}. \tag{2}$$

The more different these two rates are, the more progressive the tax system is. If the Average Tax Rate (ATR) is zero, then τ equals the Average Marginal Tax Rate (AMTR). If the average tax is positive, then τ is smaller than the AMTR. This lemma is useful because it provides a simple strategy for estimating τ, provided one has estimates of the AMTR and the ATR. Ferriere and Navarro (2020) do not include transfers in their measurements and rely on estimates of the average *statutory* marginal rates from Mertens and Olea (2018). It is the compression of statutory marginal rates in the 1986 reform that lowers estimated progressivity at that time. But recall that we found – based on the CBO data – that this change did not materially affect the distribution of actual taxes paid (see line 5 in Table 1). This suggests a disconnect between the profiles for statutory and effective marginal tax rates. More work is clearly required to fully reconcile these different approaches and estimates.

2.3 Progressivity across states, countries, and age groups

The measurement of tax progressivity above focuses on taxes and transfers at the federal level in the US. Fleck et al. (2020) apply the same methodology to state and local taxation across US states. Taking into account all taxes at state and local levels, they document marked differences in tax progressivity across states. Moreover, they find that Democrat-leaning states tend to have higher tax progressivity than Republican-leaning states.

Measuring tax progressivity across countries by modeling the detailed tax and transfer programs for the entire population is beyond the scope of this paper. Instead, we pursue

Table 2: Estimates for τ across countries

Country	Tax progressivity in 2005
USA	$\tau = 0.09$
France	$\tau = 0.13$
Japan	$\tau = 0.14$
UK	$\tau = 0.16$
Germany	$\tau = 0.16$
Denmark	$\tau = 0.23$

Estimates for progressivity τ of tax systems across countries in 2005, based on equation (2) and tax rate estimates from the World Tax Indicator database.

an exercise along the line of Ferriere and Navarro (2020) and focus on taxes only. Using data on the AMTR and ATR for 2005 from World Tax Indicator database, we estimate tax progressivity based on equation (2) for a handful of countries.

Table 2 shows that tax progressivity is similar across Continental European countries, Japan, and the United Kingdom. However, the United States has lower progressivity, and Scandinavia (represented by Denmark) has higher progressivity. Note that the estimate of τ for the United States is in line with what we estimated using CBO data when we excluded transfers from disposable income (row 5 in Table 1).

Finally, in Heathcote et al. (2020b), we document that the degree of progressivity is stable over the life cycle. To reach this conclusion, we use PSID data on married households and estimated τ_a for each age group a following the procedure in Heathcote et al. (2017). The result is that τ_a is very close to the average τ for all age groups of working age.

3 A tractable macro model

We established that tax progressivity in the US has not changed much over time. But how *should* the tax system have responded to rising inequality? To address this question, we lay out a tractable macroeconomic model with heterogeneous households and partial consumption insurance which builds closely on Heathcote et al. (2014, 2017). The model incorporates the key drivers of the observed rise in inequality in the United States.

Demographics and preferences. Demographics follow a perpetual youth model where all individuals have a constant survival rate δ and $1 - \delta$ new individuals are born every period. Households have preferences over consumption, c; hours, h; publicly-provided goods, G; and a skill investment effort, s. Preferences are time-additive with discount factor β,

$$U_i = -v_i(s_i) + \mathbb{E}_0 \sum_{t=0}^{\infty} (\beta\delta)^t u_i(c_{it}, h_{it}, G), \tag{3}$$

where the period utility function u_i is given by

$$u_i\left(c_i, h_{it}, G\right) = \log c_{it} - \exp\left[(1+\sigma)\varphi_i\right] \frac{h_{it}^{1+\sigma}}{1+\sigma} + \chi \log G.$$

The individual-specific parameter φ_i captures cross-sectional heterogeneity in the trade-off between consumption and labor supply. This heterogeneity could reflect differences in diligence, sickness, disability, and the like.[14] We assume that φ_i is drawn from a normal distribution with variance v_φ, that is, $\varphi_i \sim \mathcal{N}\left(-v_\varphi/2, v_\varphi\right)$. This choice simplifies the analytical expressions. Log-utility delivers balanced growth. Households also value a government-provided good G that enters separable in preferences.[15]

Consider now the term $v_i(s_i)$, which captures the cost (expressed in utility terms) of individual i's skill choice s_i. We assume a power disutility with a skill elasticity ψ:

$$v_i(s_i) = \frac{1}{(\kappa_i)^{1/\psi}} \cdot \frac{s_i^{1+1/\psi}}{1+1/\psi}. \tag{4}$$

The parameter κ_i captures individual heterogeneity in this disutility. We interpret κ as an index of learning ability. We assume that κ is exponentially distributed with parameter η. This assumption is important in order to preserve tractability. As we shall see, this assumption ensures that (log of) the return to skill is linear in skill s, as in a standard Mincerian model. We assume that φ_i and κ_i are independent.

Technology. Following a long line of work in macroeconomics, we assume that different skills are imperfect substitutes in production. In particular, we assume there is a continuum of skills $s \in (0, \infty)$ with a constant elasticity of substitution across skills, as assumed by, for example, Katz and Murphy (1992) and Abbott et al. (2019). In particular, the production function is given by

[14]For individual choices of consumption and labor supply, it does not matter whether we model this preference heterogeneity as a weight on consumption or labor supply in the utility function. However, this choice matters when formulating the social welfare function.

[15]It would be interesting to study an extension of this model in which a share of public goods (e.g., public parks and infrastructure) are partial substitutes to private consumption. This would make public good provision an alternative tool for redistribution.

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

where $N\left(s\right)$ is the number of aggregate effective hours supplied by skill s and $m\left(s\right)$ is the density of workers with skill s. The parameter $\theta > 1$ is the elasticity of substitution between skills. The parameter $\widetilde{\varrho}$ determines the relative importance of different skill types in production. When $\widetilde{\varrho} = 0$, all skill types are intrinsically equally important in production, while $\widetilde{\varrho} > 0$ (< 0) corresponds to a case in which technology is high-skill (low-skill) biased.

Competitive firms all have access to this technology and the equilibrium wage for skill s, $p(s)$, is the marginal product of the skill, that is,

$$\log p\left(s\right) = \frac{1}{\theta} \log\left(\frac{Y}{N}\right) + \widetilde{\varrho} s - \frac{1}{\theta} \log\left[m\left(s\right)\right]. \tag{6}$$

The price (per efficiency unit) of skill s increases more swiftly in s the larger $\widetilde{\varrho}$ is and the more swiftly the density $m\left(s\right)$ declines with s. This scarcity effect is stronger the lower θ is – that is, the more complementary in production are different skill types.

This model nests two different views of skill prices. When $\widetilde{\varrho} = 0$ (no inherent skill bias in technology), the return to skill is driven entirely by relative scarcity, as in Abbott et al. (2019) and Heathcote et al. (2017). When $\theta \to \infty$, skills are perfect substitutes in production, and skill price differentials are driven entirely by a skill-biased technology, as in Guvenen et al. (2014). To ensure existence of equilibrium and finite output, we need to impose an upper bound on $\widetilde{\varrho}$ and a lower bound on θ.

Assumption 1. Assume that $\theta > 1$ (skills are more substitutable than Cobb-Douglas) and $\widetilde{\varrho} < \sqrt{2\eta}(\theta - 1)/\theta$.

The number of aggregate effective hours worked by skill type s is given by

$$N(s) = \int_0^1 \mathbb{I}_{\{s_i = s\}} \, z_i h_i \, di,$$

where h_i is hours worked and z_i, described below, is exogenous productivity per hour worked.

Because there is no capital and thus no capital accumulation in the model, the aggregate resource constraint implies that output is spent either on consumption or on public goods,

$$Y = \int_0^1 c_i \, di + G.$$

Individual efficiency units of labor. These efficiency units are exogenous to individual choices and reflect two components, α and ε:

$$\log z_{it} = \alpha_{it} + \varepsilon_{it}.$$

The α component follows a random walk, $\alpha_{it} = \alpha_{i,t-1} + \omega_{it}$. Both ω_{it} and ε_{it} are i.i.d. over time and across households.[16] For analytical tractability, we assume that both innovations are normally distributed, $\omega_{it} \sim \mathcal{N}(-v_\omega/2, v_\omega)$ and $\varepsilon_{it} \sim \mathcal{N}(-v_\varepsilon/2, v_\varepsilon)$.

Pre-government earnings are then determined by the product of skill price times efficiency units times hours worked,

$$y_{it} = \underbrace{p(s_i)}_{\text{skill price}} \times \underbrace{\exp(\alpha_{it} + \varepsilon_{it})}_{\text{efficiency}} \times \underbrace{h_{it}}_{\text{hours}} \tag{7}$$

This formulation determines earnings as a result of human capital investment (captured by $p(s_i)$), luck (captured by the exogenous efficiency units z_{it}), and work effort (captured by hours worked).

Market structure. Financial markets are incomplete. Individuals can save in terms of a risk-free bond that is in zero net supply, subject to a natural borrowing limit. There is no explicit insurance against the (permanent) ω shocks while individuals can trade state-contingent claims offering perfect insurance against the (transitory) ε shocks. We refer to ε as *insurable* risk and to ω as *uninsurable* risk. This market structure is simple but flexible. When both $var(\omega) > 0$ and $var(\varepsilon) > 0$, the economy features partial insurance. When $var(\varepsilon) = 0$, the model is a standard incomplete markets model à la Huggett (1993). When $var(\omega) = 0$, the economy features complete markets with ex-ante heterogeneity. Finally, when there is no cross-sectional dispersion ($var(\omega) = var(\varepsilon) = var(\varphi) = 0$) and skills are perfect substitutes ($\theta \to \infty$), the economy is a standard representative-agent economy.

Finally, we assume competitive markets for labor and for the final goods and allow standard annuity markets against survival risk.

Government. The tax and transfer system is assumed to be of the log-linear form

[16]The assumption that ε is i.i.d. over time is for expositional simplicity, and none of the results depend on it. As we show in Heathcote et al. (2014), the model allows for any stochastic process for this component. However, the assumption that α is a unit-root process is important.

described in equation 1. The government chooses the fiscal parameters λ and τ and also chooses the level of expenditure G on public goods. The budget must be balanced period by period.

3.1 Equilibrium allocations

Consumption and hours. During working life, individuals choose consumption, savings, and hours, given their taste for work effort φ and their skill level s. All agents start with zero financial wealth. As we show in Heathcote et al. (2014), the equilibrium allocation of consumption and hours worked are log-linear in the (latent) factors α, ε, φ, and s. Moreover, in equilibrium all households choose to hold zero risk-free bonds. This result builds on Constantinides and Duffie (1996) and hinges on our assumptions about market structure, preferences, and wealth's being in zero net supply. The equilibrium allocations can be derived analytically as

$$\log h_{it} = \frac{\log(1-\tau)}{1+\sigma} + \left(\frac{1-\tau}{\sigma+\tau}\right)\varepsilon_{it} - \varphi_i - \mathcal{H}, \tag{8}$$

$$\log c_{it} = \log\lambda + (1-\tau)\left[\frac{\log(1-\tau)}{1+\sigma} + \log p(s_i) + \alpha_{it} - \varphi_i\right] + \mathcal{C}, \tag{9}$$

where λ and τ are policy variables and $\mathcal{H}$ and $\mathcal{C}$ are constants common for all households. Note that $\mathcal{C}$ and $\mathcal{H}$ depend on policy τ and will be fully incorporated in the welfare analysis.

Hours worked are increasing in ε and falling in τ and φ. It is optimal for the household to work harder in states when the wage rate is higher – the household wants to *make hay when the sun shines*. This effect is stronger the larger the tax-modified Frisch elasticity $(1-\tau)\,/\,(\sigma+\tau)$ is. Note that this elasticity is falling in τ. The larger tax progressivity is, the less strongly the individual wants to react to variation in ε, because changes in hours affect the marginal tax rate more when τ is large. Since the utility function is of the balanced-growth preference form and households in equilibrium hold zero wealth, the income effect of wage differentials exactly offsets the substitution effect for the uninsurable components of wages. Permanent heterogeneity through $p(s)$ and permanent uninsurable risk (through α) therefore have no effect on labor supply.

Consumption is increasing in α and $p(s)$ and falling in φ. Note that tax progressivity τ mitigates the pass-through of shocks and inequality to consumption. For example, if taxes were proportional ($\tau = 0$), the heterogeneity α and $p(s)$ would have full pass through to consumption since hours worked are unaffected by this heterogeneity. This illustrates that tax progressivity provides the insurance against life-cycle risk α that missing markets fail to deliver. It also provides redistribution of inequality at birth – that is, dispersion in φ and κ. Naturally, insurable risk ε has no effect on equilibrium consumption because households can fully hedge this risk.

Skill prices. At birth (i.e., before entering the labor market) individuals choose a skill level, given their initial draw of (κ_i, φ_i). Taking the first-order condition of the objective function (3) with respect to skill s implies

$$\frac{\partial v_i(s)}{\partial s} = \left(\frac{s}{\kappa_i}\right)^{\frac{1}{\psi}} = \mathbb{E}_0 \sum_{a=0}^{\infty} (\beta\delta)^a \frac{\partial u_i(c_{ia}, h_{ia}, G)}{\partial s}. \tag{10}$$

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

We now solve for the market price of skills $p(s)$. We guess (and will verify) that the equilibrium density $m(s)$ is exponential. Under this guess, equation (6) implies that the skill price has a standard Mincerian form – namely, that there exist coefficients π_0 and π_1 so that for any s, the skill price is

$$\log p(s) = \pi_0 + \pi_1 s. \tag{11}$$

The coefficient π_1 is the marginal return to an additional unit of skills. Combining the first-order condition (10) and the skill price function (11) yields the optimal skill choice

$$s(\kappa; \tau) = [(1-\tau)\,\pi_1]^{\psi} \cdot \kappa. \tag{12}$$

The optimal skill investment $s(\kappa;\tau)$ therefore has elasticity ψ to the after-tax return and is linear in learning ability κ. Since κ is an exponential random variable, the equilibrium distribution of skills $m(s)$ will also be exponential, which confirms our guess. It follows from equation (11) that the skill price $p(s)$ follows a Pareto distribution and earnings in (7) follow a mixture between a log-normal and a Pareto distribution. In particular, the upper tail of the earnings distribution will be Pareto.

Solving for π_0 and π_1 boils down to equating coefficients using equations (6) and (11). We show in the Online Appendix that the equilibrium return to skills is the solution to the following equation:

$$\pi_1 = \widetilde{\varrho} + \frac{1}{\theta} \frac{\eta}{((1-\tau)\,\pi_1)^{\psi}}. \tag{13}$$

By implicitly differentiating equation (13), one can show that the return to skill π_1 is increasing in $\widetilde{\varrho}$ and falling in θ. Thus, the return to skills is larger the more skill-biased technology is, and the less substitutable different skill types are (recall higher skill types are always scarcer than lower ones).

Equation (13) has analytical solutions for the equilibrium return to skill in three special cases: (i) $\widetilde{\varrho} = 0$, (ii) $\psi = 1$, and (iii) $\theta \to \infty$.

In the first case, which we analyzed in Heathcote et al. (2017),

$$\pi_1 = \left(\frac{\eta}{\theta}\right)^{\frac{1}{1+\psi}} (1-\tau)^{-\frac{\psi}{1+\psi}}. \tag{14}$$

In the second case, which is our focus in this paper,

$$\pi_1 = \frac{\tilde{\varrho}}{2} + \sqrt{\left(\frac{\tilde{\varrho}}{2}\right)^2 + \frac{\eta}{\theta\,(1-\tau)}}. \tag{15}$$

In the third case,

$$\pi_1 = \widetilde{\varrho}.$$

How do changes in tax progressivity affect equilibrium skill investment? The elasticity of skill investment to $(1-\tau)$ is given by

$$\frac{\partial s}{\partial(1-\tau)} \frac{(1-\tau)}{s} = \psi \; + \psi \frac{\partial \pi_1}{\partial(1-\tau)} \frac{(1-\tau)}{\pi_1}.$$

The first term here, ψ, is the partial equilibrium elasticity of skill investment with respect to $(1-\tau)$, holding constant the pre-tax return to skill π_1; this partial equilibrium elasticity follows directly from the skill investment rule (12). The second term is the elasticity of the pre-tax return to skill π_1 to $(1-\tau)$. It captures the fact that in general equilibrium, changing progressivity changes the skill price. Because increasing $(1-\tau)$ (reducing τ) increases skill investment, it reduces the relative scarcity of high skill types, which in turn depresses the pre-tax skill return π_1. We label this the *Stiglitz effect*, after Stiglitz (1985).

The general equilibrium Stiglitz effect dampens the direct partial skill investment response to an increase in $(1-\tau)$. The magnitude of this dampening effect depends on the nature of the production technology. Implicitly differentiating equation (13) to evaluate

$\partial \pi_1 / \partial(1-\tau)$, the full general equilibrium elasticity of skill investment to $(1-\tau)$ can be written as

$$\frac{\partial s}{\partial(1-\tau)} \frac{(1-\tau)}{s} = \frac{1}{1 - \frac{\widetilde{\varrho}}{\pi_1} + \frac{1}{\psi}}.$$

Note first that the general equilibrium elasticity is increasing in ψ, as is the partial one. It is instructive to consider the three special cases described above.

First, when $\widetilde{\varrho} = 0$, the general equilibrium elasticity simplifies to $\psi/(1+\psi)$. Thus, if $\widetilde{\varrho} = 0$ and $\psi = 1$, the general equilibrium elasticity is $1/2$, so that the Stiglitz effect cuts the partial equilibrium elasticity in half.

Second, if $\psi = 1$ (the case we study in this paper), then the elasticity is $1/(2 - \widetilde{\varrho}/\pi_1)$. In this case, holding fixed π_1, the elasticity is increasing in $\widetilde{\varrho}$. The logic is that the model can generate a given return to skill via a range of combinations of θ and $\widetilde{\varrho}$ (see equation 13). Holding fixed π_1, a higher value for $\widetilde{\varrho}$ (a more skill-biased technology) implies lower skill complementarity (a higher θ). This in turn implies a weaker Stiglitz effect (i.e., a smaller general equilibrium response of π_1) when progressivity is modified and thus a larger general equilibrium elasticity.

Third, in the limiting case when $\theta \to \infty$, $\pi_1 \to \widetilde{\varrho}$. Thus, the Stiglitz effect vanishes and the general equilibrium elasticity collapses to the partial equilibrium elasticity ψ.

The cross-sectional variance of equilibrium log skill prices is given by the variance of $\pi_1 s$. We can solve for this in closed form when $\psi = 1$, using the expression for π_1 in equation (15) and the skill investment rule (12).

We show below that the parameters $\widetilde{\varrho}$ and η affect welfare only through their impact on the parameter ϱ, which is defined by

$$\varrho \equiv \frac{1}{\sqrt{\eta}} \frac{\widetilde{\varrho}}{2}.$$

Thus, from now on, we use use ϱ as our indicator of skill bias in technology. Given $\psi = 1$, dispersion in skill prices is given by

$$v_p \equiv var\left(\log p\left(s\right)\right) = (1-\tau)^2 \pi_1^4 = (1-\tau)^2 \left(\varrho + \sqrt{\varrho^2 + \frac{1}{\theta\left(1-\tau\right)}}\right)^4. \qquad (16)$$

Equation (16) shows that dispersion in log skill prices is falling in the elasticity of substitution θ and increasing in ϱ. How does skill price dispersion vary with progressivity, τ? If we differentiate the expression for v_p with respect to $(1-\tau)$, it is straightforward to show that skill price dispersion is decreasing in progressivity when $\varrho > 0$ and is increasing in progressivity when $\varrho < 0$. In the knife-edge case when $\varrho = 0$, $v_p = 1/\theta^2$ and is thus independent of τ. In the case of perfect substitutability ($\theta \to \infty$), $v_p = (1-\tau)^2(2\varrho)^4$.

The logic for these results is that when $\varrho = 0$, an increase in τ does reduce dispersion in skills, $s = (1-\tau)\pi_1\kappa$. But the Stiglitz effect increasing π_1 is just large enough that there is no change in dispersion in $\log(p(s)) = \pi_0 + \pi_1 s$. When $\varrho > 0$ the Stiglitz effect is weaker (see above), π_1 increases less when progressivity goes up, and skill price dispersion therefore goes down. When $\varrho < 0$, the Stiglitz effect is stronger, and skill price dispersion goes up.

3.2 Planner and aggregate allocations

Given a balanced budget requirement, the government budget constraint is

$$G = \int_0^1 T(y_i \mid \lambda, \tau)\, di. \tag{17}$$

Given equation (17), the government can freely choose two of the instruments (G, τ, λ). We focus on the government's choosing (G, τ), with λ determined residually through the budget constraint. Without loss of generality, we define $g = G/Y$ and let the government choose the share of output g devoted to public goods.

For convenience, we assume that human capital investments are fully reversible. The economy will then immediately transition to the new steady state after a tax reform. It is therefore appropriate to focus on steady-state comparisons and a once-and-for-all choice of taxes, transfers, and spending on public goods.[17]

To study optimal public policy, it is necessary to take an explicit stand on the planner weights for different households. In line with a long literature, we focus on a planner who puts equal weight on all individuals belonging to the same cohort. In our context, in which agents attach different relative weights to consumption versus work effort, we take *equal weight* to mean that the planner cares equally about the utility from consumption of all agents.[18] Moreover, we assume that the utilitarian planner discounts future

[17] In Heathcote et al. (2020b) we assume that skill investments are irreversible and allow the parameters τ and λ to vary across age and time. We study the transition from an initial steady state to a future one. Allowing τ and λ to vary across time and age yields welfare gains, part of which are due to being able to tax the irreversible pre-existing human capital stock without distorting the accumulation of new human capital (see also Hassler et al., 2008).

[18] This is not an obvious choice as the planner's taste for redistribution across people with different disutility weight φ could in principle differ from taste for redistribution against uninsurable risk α. See Lockwood and Weinzierl (2015) and Piacquadio (2017) for thorough analyses of the welfare criterion and

generations by the same discount factor β that households use to discount utility over the life cycle. In our model with perpetual-youth demography, this implies that the planner puts equal weight on all individuals who are alive at any point in time.

The contribution to the social welfare function from any cohort is the average discounted future utility of the cohort members, where equation (3) defines expected lifetime utility at age zero. Social welfare evaluated as of date 0 is then given by

$$\mathcal{W}(g,\tau;\tau_{-1}) \equiv \Gamma \sum_{j=-\infty}^{\infty} \beta^j U_{j0}(g,\tau;\tau_{-1}), \tag{18}$$

where $U_{j,0}(g,\tau;\tau_{-1})$ is remaining expected lifetime utility (discounted back to date of birth) as of date 0 for the cohort that entered the economy at date j, and Γ is a constant.[19] When investments are reversible, social welfare $\mathcal{W}(g,\tau)$ is equal (up to an additive constant) to average period utility in the cross section,

$$\begin{aligned} \mathcal{W}(g,\tau;\tau_{-1}) &= (1-\delta)\sum_{a=0}^{\infty} \delta^a \mathbb{E}\left[u\left(c\left(\varphi,\alpha_a,s(\kappa;\tau);g,\tau\right),h\left(\varphi,\varepsilon;\tau\right),G\left(g,\tau\right)\right)\right] \\ &\quad -\mathbb{E}\left[v\left(s(\kappa;\tau),\kappa\right)\right]+\Xi\left(\tau_{-1}\right), \end{aligned}$$

where c, h, and s are individuals' optimal policy rules in equilibrium and G is the equilibrium public good provision.[20]

The first expectation is taken with respect to the equilibrium cross-sectional distribution of $(\alpha_a,\varepsilon,\kappa)$ and the second expectation with respect to the distribution of κ.[21] The value for τ_{-1} enters only via the additive term $\Xi(\tau_{-1})$. Because this term does not interact with the choices for g and τ, we can ignore it when computing optimal policy.[22]

Substituting the expressions for equilibrium allocations into equation (19) and evaluating the expectations yields social welfare as a function of the two policy instruments

choice of planning weights in the presence of preference heterogeneity.

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

[20] G is written as a function of τ because the government chooses g, the output share, and aggregate output Y depends on τ.

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

[22] This property relies on the assumption that skill investments are reversible.

(g,τ). In the special case when $\psi = 1$, social welfare can be expressed analytically as

$$\mathcal{W}(g,\tau) = \begin{cases} (1) & \log(1-g) + \chi \log g + (1+\chi)\frac{\log(1-\tau)}{(1+\hat{\sigma})(1-\tau)} - \frac{1}{1+\hat{\sigma}} \\ (2) & +(1+\chi)\left(-\frac{1}{2}\log\left[f(\tau)\right] - \frac{\theta}{\theta-1}\log\left(\frac{\theta-1}{\theta}\left[f(\tau)\right]^{-\frac{1}{2}} - 2\varrho\right)\right) \\ (3) & -\frac{1}{2}f(\tau) \\ (4) & +\log\left[1-f(\tau)\right] + f(\tau) \\ (5) & -(1-\tau)^2\frac{v_\varphi}{2} \\ (6) & -(1-\tau)^2\frac{v_\alpha}{2} \\ (7) & +(1+\chi)\left(\frac{1}{\hat{\sigma}}v_\varepsilon - \sigma\frac{1}{\hat{\sigma}^2}\frac{v_\varepsilon}{2}\right), \end{cases} \tag{19}$$

where $\hat{\sigma} = (\sigma+\tau)/(1-\tau)$ is the inverse of the tax-modified Frisch elasticity, and the term $f(\tau)$ is given by[23]

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

As for earnings, the cross-sectional distribution for consumption is a mixture of a log-normal and a Pareto distribution. The term $1/f(\tau)$ is the Pareto parameter for the component of consumption that is driven by skill prices.

In the next section, we calibrate the model, solve for optimal policy, and give an explicit interpretation of each term in the welfare expression (19).

4 Quantitative results

4.1 Calibration

To quantify the optimal tax policy, we must set values for all those parameters that appear in the social welfare function above. The preference parameters $\sigma, \chi, v_\varphi, \psi$ are assumed to be time invariant, whereas we let the technology and risk parameters $\theta, \varrho, v_\alpha, v_\varepsilon$ vary over time. For our numerical experiment, we pick 1980 as our initial point, because this is the first year when household-level consumption data are available from the CEX, and 2016 as our final point.

[23] Note also that the sixth line is an approximation. The exact expression is $-(1-\tau)\delta/(1-\delta)\cdot v_\omega/2 + \log\left[1-\delta\exp\left(-\tau(1-\tau)v_\omega/2\right)\right] - \log(1-\delta)$. Given realistic values for δ and v_ω, this term is approximately equal to $-(1-\tau)^2 v_\alpha$ where $v_\alpha = v_\omega \cdot \delta/(1-\delta)$.

4.1.1 Preference parameters held constant over time

We set the curvature parameter on hours worked to $\sigma = 2$. This value is in line with our previous estimates in Heathcote et al. (2014) and implies a modified Frisch elasticity of around 0.4, given our estimate of the empirical $\tau_{US} = 0.186$. We set the survival probability to $\delta = 0.971$, which ensures an average duration of 35 years for a working life.

To calibrate the weight on the (government-provided) public good χ, we assume that the empirical fraction of output devoted to publicly provided goods g is efficient – that is, welfare maximizing. From (19), it is immediate that the efficient level is $g^* = \chi/(1+\chi)$. Note that g^* depends only on households' relative taste for the public good χ. The intuition for this result is that all individuals have the same logarithmic preferences over private and public consumption and hence the same trade-off between the goods. Therefore, there is no disagreement about how much of output should be allocated to G, irrespective of the level of inequality and risk. The ratio of government consumption and investment to GDP was $g = 0.192$ on average for the years 1980 and 2016, which implies $\chi = 2.37$. In Section 5.2, we allow χ to change over time.

The variance of preference heterogeneity, v_φ, can be identified from the cross-sectional covariance between consumption and hours worked according to the allocations in (9):

$$cov\left(\log h, \log c\right) = (1-\tau)\, v_\varphi. \tag{20}$$

Using data from the CEX, we find that $cov\left(\log h, \log c\right)$ is very stable over time, with an average value of 0.034.[24] Given $\tau = 0.186$, equation (20) implies $v_\varphi = 0.0425$.

We set the elasticity of skill investment parameter ψ to one, as this allows us to study the case with $\varrho \neq 0$ and different types of technological change. At the same time, we now argue that $\psi = 1$ is broadly consistent with estimates of how aggregate skill investment in the United States has evolved in response to observed changes in the rate of return to skill. From equation (12), the average skill choice is $\mathbb{E}[s] = [(1-\tau)\pi_1]^{\psi}\, \mathbb{E}[\kappa]$. This implies that ψ can be estimated from aggregate changes in s and π_1,

$$\log\left(\frac{\mathbb{E}[s_{2016}]}{\mathbb{E}[s_{1980}]}\right) = \psi \log\left(\frac{(1-\tau_{2016})\pi_{1,2016}}{(1-\tau_{1980})\pi_{1,1980}}\right). \tag{21}$$

To measure the change in skills and the return to skills over time we must take a stand

[24]Following Heathcote et al. (2010a), we drop households that work fewer than 260 hours per year, measure consumption as nondurable consumption equivalized using the OECD scale, and measure hours as hours for the head of household. We use this selection criterion because our theory focuses on working households.

on how to measure skills. Given our focus on discretionary schooling and how skill investment choices respond to changes in tax progressivity, we will measure s as years of education over and above mandatory schooling. In most US states, school is mandatory up to age 16. With mandatory schooling starting at age six, anything above 10 years of schooling reflects an active investment choice. Using data from the CPS, we find that average years of education for individuals between the ages of 26 and 30 were 12.6 and 14.0 in 1980 and 2016, respectively. A standard Mincer regression on this CPS sample that controls for gender and a quartic in age implies that the return to an additional year of schooling increased from 8.2% in 1980 to 11.7% in 2016.

Since progressivity has remained constant over this period, the implied elasticity is then

$$\psi = \log\left(\frac{14.0-10}{12.6-10}\right) / \log\left(\frac{(1-0.186)0.117}{(1-0.186)0.082}\right) = 1.21.$$

We conclude that if the entire change in aggregate skill attainment were driven by the observed increase in the return to skill, the implied elasticity ψ would be close to our assumed value of unity. In Section 5.2, we perform a sensitivity analysis on the value for ψ.

4.1.2 Technology and risk parameters changing over time

We take the view that differences in wages by education and by age reflect differences in skill investments as reflected in $p(s)$, while residual wage dispersion in log wages within age-education groups reflects exogenous labor market risk (α and ε). The logic for including age as well as education as a proxy for skills is that standard theories of life-cycle wage growth emphasize skill acquisition via human capital investment on the job and via learning by doing. Heathcote et al. (2010a) show that the variance of log wages explained by education and age in a standard Mincerian regression was 0.051 in 1980. Extending their methodology to 2016, we find that this dispersion increases to 0.110.[25]

We set $\theta_{1980} = 3.3$ in line with the estimate in Abbott et al. (2019) who use longitudinal US survey data to estimate the elasticity of substitution between three education groups. We then calibrate ϱ_{1980} so that the variance of $\log p(s)$ as defined in (16) matches between-group inequality. With $\tau_{1980} = \tau_{2016} = 0.186$ and $\theta_{1980} = 3.3$, equation (16) implies $\varrho_{1980} = -0.0891$.

[25] Education and age are coarse proxies for skill investment. For example, investment likely varies significantly by choice of college, choice of major, and by grades achieved. Thus, our estimates potentially understate the share of wage dispersion reflecting differential skill investment.

We will consider two alternative ways to model the increase in between-group inequality (skill price dispersion) between 1980 and 2016. First, we consider the case in which the entire increase is due to skill-biased technical change caused by an increase in ϱ, while θ is held constant at $\theta = 3.3$. This is our baseline assumption. From equation (16), this implies that the skill-bias parameter increases to $\varrho_{2016} = 0.0276$. Second, we consider the other extreme case, in which the growth in between-group inequality entirely reflects a rise in θ, with ϱ held constant at its 1980 value. This alternative view of technical change implies that $\theta_{2016} = 2.357$, indicating that skill types have become more complementary over time.[26]

Consider now the change in residual wage dispersion. Blundell and Preston (1998) and Heathcote et al. (2014) show that the variance of uninsurable risk can be identified from the cross-sectional consumption dispersion. From equation (22),

$$var\left(\log c\right) = \left(1-\tau\right)^2 \left(v_\varphi + v_p + v_\alpha\right). \tag{22}$$

Using the CEX data for consumption described above, we find that the variance of log consumption (nondurables plus services), $var\left(\log c\right)$, increased from 0.215 to 0.275 between 1980 and 2016. This measurement assumes that empirical log consumption has time-invariant classical measurement error with a variance of 0.041, in line with our estimates in Heathcote et al. (2014). Given our assumption that the variance of between-group wage inequality is v_p and our calibration of v_φ, equation (22) then identifies v_α and implies that v_α increased from 0.231 to 0.263 between 1980 and 2016. The variance of uninsurable innovations, v_ω, is then determined by the relation $v_\alpha = v_\omega \times \delta / (1-\delta)$.

The magnitude of insurable risk can then be identified from the variance of log wages as residual wage dispersion $(v_\alpha + v_\varepsilon)$ minus uninsurable risk v_α. This approach implies that v_ε increased from 0.025 to 0.114 between 1980 and 2016. All parameter values are summarized in Table 3.

It is useful to take a short detour on the interpretation of v_α and v_ε. We have modeled and interpreted them as labor market uncertainty, and our calculations imply that this source of wage volatility has increased in the last 40 years. Early work from survey data (see, for example, the discussion in Heathcote et al., 2010b) is consistent with this view. However, more recent research relying on administrative data (Guvenen et al., 2017), and comparing survey and administrative data (Moffitt, 2020), has disputed these findings and suggested that the rise in residual wage inequality is mostly due to larger dispersion

[26] Recall that increased complementarity can be interpreted as workers having specialized in acquiring narrower sets of skills, so that workers have become effectively less substitutable.

Table 3: Parameterization

		Time-invariant parameters				
		σ	δ	χ	ψ	v_φ
		2	0.971	0.237	1	0.0425
		Time-varying parameters				
	Year	ϱ	θ	v_α	v_ω	v_ε
	1980	−0.0891	3.3	0.2307	0.0068	0.0245
ϱ changes	2016	0.0276	3.3	0.2625	0.0078	0.1135
θ changes	2016	−0.0891	2.357	0.2625	0.0078	0.1135

Calibration of model parameters. See the main text for details.

in initial conditions at labor market entry (which cannot be explained by demographics) as opposed to higher wage volatility. This viewpoint is supported by evidence from administrative data in (Guvenen et al., 2017). This alternative view poses no challenge for our welfare exercise, since what enters in the welfare expression (19) is the cross-sectional variance; whether this is generated by longitudinal volatility or initial dispersion makes no difference.

To illustrate the magnitudes of the parameter estimates in Table 3, we do a variance decomposition of wages, consumption, and hours in our model. Table 4 lists the results. We find that (i) the uninsurable component α accounts for more than half the cross-sectional variance of wages, while skill prices account for about one-fifth; (ii) cross-sectional dispersion in the disutility of work effort explains at least three-quarters of model hours variation, while insurable shocks explain the rest;[27] (iii) the dispersion in the disutility of work accounts for around one-tenth of model consumption inequality, uninsurable wage shocks account for about two-thirds, and skill price dispersion accounts for around one-fifth.

It is also useful to illustrate the drivers of the rise in labor income inequality in the US according to the model. The parameters in Table 3 imply that (i) the insurable component accounts for about half the increase in the variance of log wages, while skill prices and the uninsurable component account for 30% and 20% of the increase, respectively; (ii) the increase in skill price dispersion accounts for two thirds of the growth in consumption inequality, while uninsurable wage risk accounts for the residual third. Finally, a

[27] A key reason for modeling heterogeneity in φ is to be able to account for the empirical dispersion in hours worked. Note that the dispersion in model hours is smaller than the dispersion in empirical hours. In Heathcote et al. (2014), we allow for measurement error and target the dispersion in empirical hours. There, the estimates of preference heterogeneity are slightly larger than in Table 3.

Table 4: Variance decomposition

	1980				2016			
	v_p	v_α	v_ε	v_φ	v_p	v_α	v_ε	v_φ
$var(\log w)$	0.17	0.75	0.08	0	0.23	0.54	0.23	0
$var(\log c)$	0.16	0.71	0	0.13	0.27	0.63	0	0.10
$var(\log h)$	0	0	0.07	0.93	0	0	0.27	0.73

Variance decomposition for wages, consumption, and hours in the model in 1980 and 2016. v_p: skill prices; v_α: uninsurable labor market risk; v_ε: insurable labor market risk; v_φ: preference heterogeneity.

change in tax progessivity could have affected consumption inequality. However, with an approximately constant τ, this contribution is zero.

4.2 Optimal tax progressivity in 1980

We begin by analyzing optimal progressivity in the economy calibrated to 1980.[28] To start with, we assume that all skills are perfect substitutes ($\theta \to \infty$) and abstract from all other sources of heterogeneity ($v_\alpha = v_\varepsilon = v_\varphi = 0$). This case corresponds to a representative-agent economy. The social welfare function $\mathcal{W}$ then simplifies to the first line in equation (19). The first three terms on this line reflect the value of private and public consumption. The last term is the average disutility of hours worked. Maximizing the objective (19) for this case yields a regressive tax, $\tau^*_{RA} = -\chi = -0.237$. An appropriately regressive tax system delivers a zero marginal tax rate at the equilibrium hours choice for the representative agent, while at the same time, the average tax rate is positive and delivers sufficient revenue to finance the optimal level of government purchases.

The second, third, and fourth lines in equation (19) capture the welfare contributions of skill investment choices and equilibrium skill price dispersion. The second line captures the efficiency aspect of skill accumulation. To see this, note that the term on this line is equal to $1 + \chi$ times aggregate productivity, measured as output per efficiency unit, plus a constant. Productivity is falling in τ, since progressivity discourages skill accumulation. The third line captures the utility cost of skill accumulation and is increasing in τ (discouraging skill investment saves on investment costs). Finally, the fourth line is the utility cost of the consumption dispersion caused by skill price dispersion. Naturally, this term is increasing in τ, since more compressed after-tax wages mitigate consumption

[28] Heathcote et al. (2017) discuss conditions under which $\mathcal{W}$ is globally concave in g and τ, in which case the first-order approach is sufficient to derive optimality conditions.

dispersion (see equation 9). In summary, skill choice introduces a trade-off in the choice of τ between incentives to accumulate skills (a force for lower τ) and the redistribution toward individuals with low learning ability κ (a force for larger τ). Setting θ and ϱ to their 1980 values implies that once we add the skill investment dimension to the representative agent model, optimal progressivity increases to $\tau^* = -0.15$ for our utilitarian planner.

The fifth line captures the welfare cost of preference heterogeneity. This term is the familiar Lucas expression for the welfare cost of consumption dispersion when shocks are log-normal: one-half of the variance of log consumption due to φ times the relative risk aversion, which is equal to unity. The fact that preference heterogeneity is a force for a larger τ is related to our assumption that equal planning weights mean equal weights on the consumption term across individuals in the social welfare function. Adding preference heterogeneity to the model increases optimal progressivity to $\tau^* = -0.061$.

The sixth line in (19) captures another key source of consumption dispersion: uninsurable shocks. Again, this corresponds to one-half of the variance of log consumption attributable to this source of heterogeneity. Naturally, this is unambiguously a force for higher progressivity, since this source of dispersion in consumption is falling in τ. The optimal value for progressivity increases to $\tau^* = 0.185$ when uninsurable risk is added to the model.

Finally, the last line in equation (19) captures the welfare effect of incorporating insurable risk. As illustrated by the allocations in (9), insurable risk has no impact on consumption but influences labor supply. The sum of these two terms is maximized at $\tau = 0$, which is the value at which hours respond efficiently to insurable shocks. Insurable wage risk therefore pushes τ^* toward zero. Adding insurable risk to the calibrated economy lowers τ^* slightly to 0.181. This implies an average income-weighted marginal tax rate of 33.8%.

In sum, when all model features are incorporated, optimal progressivity for the equal-weights planner is very close to the empirical estimate for 1980 of $\tau = 0.186$. We take this as evidence that our choice of a utilitarian equal-weights objective function for the planner offers a reasonable description of US society's taste for redistribution in 1980. In what follows we maintain this specification of the welfare objective.[29]

[29]To ensure that the optimal progressivity in the model is exactly equal to its empirical counterpart in 1980, we could introduce inequality aversion along the lines of Bénabou (2002). Pursuing this exercise under our calibration would imply that the planner has a degree of inequality aversion (equivalent to the planner's relative risk aversion for consumption inequality) of approximately 1.025, only slightly larger than the risk aversion of the individuals in the economy. See Heathcote et al. (2017) for details on how inequality aversion can be introduced in this model while retaining tractability and a closed-form expression

Table 5: Comparative statics on τ

	τ^*_{1980}	τ^*_{2016}	$\Delta\tau^*$	$AMTR_{1980}$	$AMTR_{2016}$
Change only ϱ	0.181	0.155	-0.026	0.338	0.317
Change only v_α	0.181	0.201	$+0.033$	0.338	0.354
Change only v_ε	0.181	0.167	-0.014	0.338	0.327
All combined	0.181	0.161	-0.020	0.338	0.322

The table shows the optimal τ^* in 1980 and 2016. In the first three lines, only one parameter changes to its 2016 level, while the others remain at their 1980 calibration values. In the last line, all parameters change to their 2016 levels. AMTR refers to the income-weighted average marginal tax rate, calculated using the formula $AMTR = 1 - (1-g)(1-\tau)$.

5 The optimal tax response to rising income inequality

We now address the main question of the paper: How should taxes respond to the widening of the income distribution observed since 1980? To answer this question, we introduce one mechanism at the time. The parameter shifts we consider are those estimated in Table 3.

The first exercise we consider is skill-biased technical change – that is, an increase in the skill-bias parameter ϱ to its 2016 value. The qualitative effect of increasing ϱ on the optimal τ^* is theoretically ambiguous and reflects the planner's trade-off between equality and efficiency. On the one hand, a larger ϱ increases the value of human capital accumulation in terms of aggregate output. This efficiency channel is a force for a lower τ^*. On the other hand, a larger ϱ increases the dispersion of skill prices and hence of consumption. This redistribution channel pushes the equal-weights planner to choose a higher τ^*. Imposing the estimated increase in ϱ while leaving the other parameters at their 1980 values induces the planner to *lower* tax progessivity: $\Delta\tau^* = -0.026$ (first row in Table 5). Thus, in our calibration, the efficiency effect dominates the redistribution effect. To illustrate the magnitude of this change in τ, the table also reports the average (income-weighted) marginal tax rate associated with each value for τ.

Consider now the drivers of residual (within-group) wage inequality. In response to the increase in v_α, the planner wants to increase progressivity by 0.033 (second row in Table 5) in order to provide more social insurance against uninsurable income risk. In contrast, the planner would lower the progressivity by $\Delta\tau^* = -0.014$ in response to the estimated rise in the insurable component of risk. Note that the quantitative effect on τ^*

for social welfare.

is small compared with the effect of changes in uninsurable risk, even though the rise in insurable risk is almost three times larger than the rise in uninsurable risk.

Combining together all sources of rising income inequality implies slightly less progressive taxes in 2016 than in 1980, $\tau^*_{2016} = 0.161$ (last row in Table 5). This implies a (modest) fall in the average marginal tax rate of about one percentage point. This somewhat surprising result is mainly the consequence of two forces. First, skill-biased technical change increases the return to human capital investment, and raising progressivity would excessively distort this margin. Second, the rise in consumption inequality in the data is lower than that of earnings inequality, implying that a large part of the latter was privately insurable and that additional social insurance through stronger progressivity would be redundant. The next section evaluates the robustness of this result and digs deeper into how an increase in the return to skill impacts optimal progressivity.

5.1 Skill price dispersion and optimal progressivity

To better understand the mechanisms, consider first an economy with no motives for redistribution other than skill price dispersion and no distortions besides skill investment. In particular, we make labor supply inelastic ($\sigma \to \infty$) and abstract from all heterogeneity beyond the dispersion in learning ability κ (i.e., we set $v_\alpha = v_\varepsilon = v_\varphi = 0$). We retain valued government purchases by keeping χ at its baseline value. In this stripped-down model, the planner trades off the benefit of more progressivity in terms of reduced skill price inequality against the efficiency loss in terms of less skill investment and lower output. The optimal τ given the 1980 values for θ and ϱ is $\tau = -0.01$ (second row in Table 6). Increasing ϱ to its 2016 value leads the planner to *raise* progressivity, contrary to what we found when running the same experiment in the baseline model (first row in Table 6).

The left panel of Figure 5 illustrates social welfare in this stripped-down economy as a function of τ for the 1980 calibration (solid blue line) and the 2016 calibration (dashed red line). In each case, welfare at each value for τ is measured in units of equivalent percentage decline in consumption at the welfare-maximizing value τ^*.

When ϱ increases (moving from blue to red), two things happen to the profile for welfare. First, the maximum shifts to the right, indicating that increasing ϱ raises the marginal distributional gains from higher progressivity more than it raises the marginal cost of lower productivity. Second, the welfare function becomes more concave in τ. The logic for this is that with a higher ϱ, changes in skill investment that arise when the planner alters τ have larger effects on efficiency and inequality.

Now consider the baseline version of the model with all model ingredients switched

Table 6: Robustness analysis

		τ^*_{1980}	τ^*_{2016}	$\Delta\tau^*$	*AMTR* 1980	*AMTR* 2016
(1)	Benchmark calibration (see Table 3)	0.181	0.161	−0.020	0.338	0.322
(2)	ϱ changes, only skill motives present	−0.010	0.043	+0.053	0.183	0.226
(3)	Technological change driven by θ	0.181	0.169	−0.012	0.338	0.328
(4)	θ changes, only skill motives present	−0.010	0.040	+0.050	0.183	0.224
(5)	Exogenous skills (no distortion to skill inv.)	0.275	0.317	+0.042	0.414	0.448
(6)	Low elasticity of skill investment ($\psi = 0.5$)	0.204	0.207	+0.003	0.356	0.359
(7)	Larger cons. inequality increase	0.181	0.188	+0.007	0.338	0.344
(8)	Taste for public good χ changes over time	0.171	0.172	+0.001	0.342	0.318

The table shows the optimal τ^* in 1980 and 2016 under various alternative calibrations. Row (1) is the baseline model in which technical change is driven by ϱ with $\theta_{2016} = \theta_{1980}$. Row (2) is an economy where $v_\alpha = v_\varepsilon = v_\varphi = 0$ and $\sigma \to \infty$ and technical change is driven by ϱ. Row (3) is the baseline model in which technical change is driven by θ with $\varrho_{2016} = \varrho_{1980}$. Row (4) is an economy where $v_\alpha = v_\varepsilon = v_\varphi = 0$ and $\sigma \to \infty$ and technical change is driven by θ. Row (5) assumes that the skill distribution is exogenous (no distortion to human capital). Row (6) assumes a lower ψ with technical change driven by θ. Row (7) assumes a larger increase in consumption dispersion, in line with Attanasio et al. (2007); $\Delta var\,(\log c) = 0.09$. Row (8) allows χ to change over time alongside ϱ. AMTR refers to the income-weighted average marginal tax rate, calculated using the formula $AMTR = 1 - (1-g)(1-\tau)$.

on. This case is plotted in the right panel of Figure 5. The first thing to note here is that while adding these new model elements activates additional terms in the welfare expression (equation 19), the terms that have to do with skill investment and skill price inequality—lines (2), (3) and (4)—are exactly the same as in the left panel. The second important observation is that introducing additional sources of heterogeneity now implies a higher optimal value for τ, given the 1980 values for ϱ and θ. So now the social welfare terms involving skills–those plotted in the left panel–are pulling progressivity down. Moreover, these terms pull the optimal τ down further when ϱ increases to its 2016 value, because the terms involving skills in social welfare are more concave in τ in 2016 than in 1980. Put differently, in 1980, skill investment considerations alone call for a lower τ relative to other model ingredients, but because the blue welfare expression in the left panel is quite flat, skill considerations have a relatively minor impact on the optimal policy. In contrast, in 2016, skill considerations matter more, and the optimal policy is pulled closer to the one that would be dictated by skill considerations alone – that is, the maximum in the left panel.

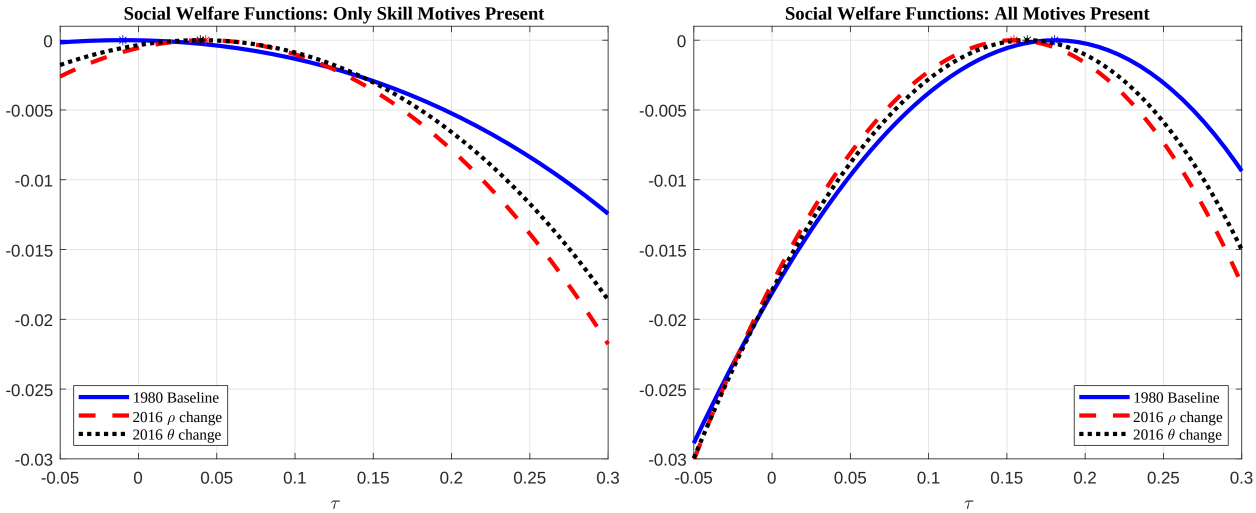

Figure 5: The figure illustrates social welfare functions (SWF) for various cases. The left panel abstracts from all motives for progressivity except skill price dispersion ($\sigma \to \infty$ and $v_\alpha = v_\varepsilon = v_\varphi = 0$). The right panel is the benchmark case when all motives for redistribution are present. The solid blue lines are SWF in 1980. The dashed red lines are SWF when $\varrho = \varrho_{2016}$ and other parameters are as in 1980. The dotted black lines are SWF when $\theta = \theta_{2016}$ and other parameters are as in 1980.

Specialization-biased technical change . The alternative model of technical change we discussed is one in which ϱ is constant but θ falls over time, implying greater complementarity between skill types. This version is plotted in the dotted black lines in Figure 5. Recall that given the empirically estimated value for τ (0.186), the red and black economies deliver identical equilibrium skill premium (π_1) and identical skill price dispersion. The left-hand panel of Figure 5 indicates that welfare is more concave in τ in the red, high ϱ economy than in the black, low θ economy. The reason is that the general equilibrium elasticity of skill investment to progressivity is larger in the high ϱ economy (see the discussion in Section 3.1). Because skill investment is less sensitive to τ in the low θ economy, from a skill investment perspective, the welfare costs of choosing the wrong τ are smaller. Thus, in this version of the model, skill investment considerations are a slightly weaker force pulling progressivity down, translating into a optimal τ of 0.169 in 2016 under the specialization-biased model for increased skill price dispersion, compared with 0.161 in the baseline economy.

5.2 Robustness analysis

Elasticity of skill investment . Our conclusion that progressivity should have remained approximately constant, or even declined slightly, in response to the substantial rise in inequality since 1980 is driven primarily by the fact that an increase in the return to skill increases the welfare cost of distorting skill investment through progressive taxation. To further illustrate this result, we now consider economies with less elastic skill choices.

Consider first an economy with a fully exogenous skill distribution, where all individuals' skills are exogenous and thus cannot be affected by progressivity. In this case, dispersion in $p(s)$ has exactly the same welfare consequences as dispersion in uninsurable risk α, and rising skill price dispersion is isomorphic to additional exogenous uninsurable income risk.[30] In this version of the model, progressivity changes from $\tau^*_{1980} = 0.275$ to $\tau^*_{2016} = 0.317$ (row [5] of the table). This experiment confirms that if the observed rise in consumption inequality were caused entirely by the uninsurable and exogenous wage component α, the optimal policy response would be to make the tax and transfer system substantially more progressive over time.

Consider now a version of the economy in which the skill investment elasticity is set to $\psi = 0.5$ – that is, half of its benchmark value.[31] We focus here on the case with no skill bias in technology ($\varrho = 0$) and assume that the observed increase in skill price dispersion is driven by a decline in θ.[32] The return to skill is now given by equation (14) and, as a consequence, the parts of the social welfare function (19) that involve skill investments (lines 2-4) are now modified in line with the model with $\psi \neq 1$ and $\varrho = 0$ (see Heathcote et al., 2017, for details). Implementing these changes implies that optimal progressivity now increases very slightly over time (row [6] in the table). Overall, our conclusion of roughly unchanged optimal progressivity over time appears robust to plausible variation in the elasticity of skill investment to the after-tax return to skill.

[30]In this special case, skill prices enter the social welfare function (19) in the same way as uninsurable risk. Namely, lines two and three in equation (19) are equal to zero, and the only effect of skill prices is through the cost of consumption inequality. This is still given by line four in equation (19). All other aspects of the calibration are unchanged. Note in particular that this alternative economy has the same changes in cross-sectional inequality in wages and consumption as the benchmark economy.

[31]To put this alternative value for ψ in context, $\psi = 0.5$ implies that following the observed increase in the return to education, average years of education would have increased from 12.6 to just 13.1, instead of the observed 14.0 between 1980 and 2016.

[32]The $\varrho = 0$ assumption allows us to solve equation (13) analytically with $\psi \neq 1$. This exercise also requires a different calibration of θ. The implied variance of log skill prices is $1/\theta^2$ (cf. equation [16]). Setting θ to match between-group inequality in 1980 and 2016 then implies $\theta_{1980} = 4.42$ and $\theta_{2016} = 3.04$. The rest of the calibration is unchanged.

Rise in consumption inequality . There is some disagreement in the literature as to the size of the increase in consumption inequality in recent decades. Some authors have argued that consumption inequality has likely increased by more than the estimates in Heathcote et al. (2010a). For example, Attanasio et al. (2007) combine data from both the Diary Sample and the Interview Sample in CEX and find that the variance of log consumption increased by 0.09 over the 1980-2005 period. Recall that the benchmark economy targets an increase of 0.06. This alternative estimate would imply a different calibration with a stronger increase in uninsurable risk ($v_{\alpha,2016} = 0.308$) and a more moderate increase in insurable risk ($v_{\varepsilon,2016} = 0.068$). Under this alternative view of the data (row (7) in Table 6), optimal progressivity increases slightly to $\tau^*_{2016} = 0.188$ in 2016. Thus, an accurate assessment of the true increase in uninsurable risk–which is manifested in higher consumption inequality–is an important input for computing how to optimally adjust progressivity in response to widening inequality.

Public expenditure . Finally, we also allow χ to change over time, together with the other parameters. Government spending as a share of GDP fell from 20.6% in 1980 to 17.6% in 2016. This implies $\chi_{1980} = 0.259$ and $\chi_{2016} = 0.214$. Holding the other parameters constant at their 1980 levels, this implies an increase in optimal progressivity from 0.17 to 0.19. As explained in Section 4.2, public good provision is a force for less progressive taxes, and a lower χ weakens this force. When implementing the change in χ alongside all other changes in parameters between 1980 and 2016, optimal progressivity is approximately constant, as in the data (row [8] in Table 6).

6 Conclusion

This paper asks how a utilitarian government that puts equal weight on all households in the economy should modify the tax and transfer system in response to rising income inequality. Answering this question within the log-linear class of tax and transfer systems–a specification that matches the data well–yields a closed-form and transparent solution to an otherwise computationally complex problem.

Our main finding is that the appropriate policy prescription hinges on the nature of the rise in inequality. If larger wage dispersion is caused by rising uninsurable labor market risk or ex-ante heterogeneity exogenous to individuals' choices, the recommendation is unambiguous: progressivity should increase in order to provide more social insurance. Conversely, progressivity should fall if the rise in risk is privately insurable. The optimal response to a higher skill premium depends on the magnitude of the human capital distortion. In our baseline calibration, skill-biased technical change that increases equilibrium skill price dispersion calls for lower optimal progressivity. Overall, the optimal response to the combination of factors shaping changes in wage, earnings, and consumption inequality in the United States is to keep progressivity approximately constant. This prescription is consistent with the empirical evolution of the US tax and transfer system over the last four decades.

Going forward, our analysis could be refined in a number of dimensions. We let the government use a limited set of policy instruments: income taxes and transfers only.[33] Some results in the existing literature (Bénabou, 2002; Krueger and Ludwig, 2013; Stantcheva, 2017) suggest that education subsidies could be an important component of the optimal policy. For example, in a model where most of human capital accumulation occurs before entering the labor market, the optimal policy response to rising inequality might combine an increase in tax progressivity with more generous subsidies to formal education. In a model where skills are mainly accumulated through on-the-job learning, however, such a policy mix would be less useful.[34]

We took the view that "biased" technological change is the driving force of the change in the wage structure. Within this view, we abstracted from two aspects. First, when technology is embodied in capital, another tool the government can employ to control the wage structure is a tax on capital or on "robots," machines that replace certain types of labor services, as in Guerreiro et al. (2017), Costinot and Werning (2018), Thuemmel (2019), and Moll et al. (2019). Second, when technological change reduces the return to work for a range of occupations (e.g., routine jobs in the manufacturing sector), it can end up pushing these workers out of the labor force altogether (Heathcote et al., 2020a). In this scenario, well designed active labor-market programs can be useful (Blundell, 2002; Pavoni and Violante, 2007; Pavoni et al., 2016).

Other forces have contributed to shape the wage structure beyond technology in the US. For example, increased trade exposure and the "China shock" are often heralded as drivers of wage inequality (see Autor et al., 2013). As we insisted throughout the paper, different sources of changes in inequality call for different prescriptions: a full normative analysis of taxation within a model where inequality is trade induced, with

[33]The introduction of progressive consumption taxes would not affect our results. In Heathcote et al. (2017), we show an isomorphism to income taxes within our framework.

[34]Kapička (2020) explores an alternative generalization of government policy by analyzing optimal taxation in a model where after-tax income is a log-linear weighted average of the entire history of past earnings instead of just current earnings, as in our paper.

heterogeneous gains and losses across the income distribution stemming from the free flow of goods and ideas would represent a welcome contribution to this literature. See Antrás et al. (2017) for an analysis incorporating inequality and progressive taxation in the context of international trade.

Finally, we specialized our analysis to the US. But what is valid for one country may not be valid for another. Different cultures, social norms, and institutions mean that technology and trade affect the wage structure differently across countries. A cross-country comparative normative analysis is yet another avenue that should be explored further.

References

Abbott, Brant, Giovanni Gallipoli, Costas Meghir, and Giovanni L. Violante (2019). "Education Policy and Intergenerational Transfers in Equilibrium." *Journal of Political Economy*, 127(6), 2569–2624.

Ales, Laurence, Musab Kurnaz, and Christopher Sleet (2015). "Technical Change, Wage Inequality, and Taxes." *American Economic Review*, 105(10), 3061–3101.

Alon, Titan (2018). "Earning More by Doing Less: Human Capital Specialization and the College Wage Premium." Mimeo, University of California, San Diego.

Antrás, Pol, Alonso De Gortari, and Oleg Itskhoki (2017). "Globalization, Inequality and Welfare." *Journal of International Economics*, 108(9), 387–412.

Attanasio, Orazio, Erich Battistin, and Hidehiko Ichimura (2007). "What Really Happened to Consumption Inequality in the United States?" In *Hard-to-Measure Goods and Services: Essays in Honor of Zvi Griliches*, edited by Ernst R. Berndt and Charles R. Hulten, pp. 515–543. University of Chicago Press.

Autor, David H., David Dorn, and Gordon H. Hanson (2013). "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review*, 103(6), 2121–2168.

Badel, Alejandro, Mark Huggett, and Wenlan Luo (2020). "Taxing Top Earners: A Human Capital Perspective." *Economic Journal*, 130(629), 1200–1225.

Bakis, Ozan, Barış Kaymak, and Markus Poschke (2015). "Transitional Dynamics and the Optimal Progressivity of Income Redistribution." *Review of Economic Dynamics*, 18(3), 679–693.

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

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

Blundell, Richard (2002). "Welfare-to-Work: Which Policies Work and Why?" *Proceedings of the British Academy*, 117, 477–524.

Blundell, Richard and Ian Preston (1998). "Consumption Inequality and Income Uncertainty." *Quarterly Journal of Economics*, 113(2), 603–640.

Constantinides, George M. and Darrell Duffie (1996). "Asset Pricing with Heterogeneous Consumers." *Journal of Political Economy*, 104(2), 219–240.

Costinot, Arnaud and Iván Werning (2018). "Robots, Trade, and Luddism: A Sufficient Statistic Approach to Optimal Technology Regulation." NBER Working Paper No. 27622.

Ferriere, Axelle and Gaston Navarro (2020). "The Heterogeneous Effects of Government Spending: It's All About Taxes." Mimeo, Paris School of Economics.

Findeisen, Sebastian and Dominik Sachs (2016). "Education and Optimal Dynamic Taxation: The Role of Income-Contingent Student Loans." *Journal of Public Economics*, 138, 1–21.

Fleck, Johannes, Jonathan Heathcote, Kjetil Storesletten, and Giovanni L. Violante (2020). "Tax progressivity at state and local level." Mimeo, University of Oslo.

Groneck, Max and Johanna Wallenius (2020). "It Sucks to Be Single! Marital Status and Redistribution of Social Security." Mimeo, Stockholm School of Economics.

Guerreiro, Joao, Sergio Rebelo, and Pedro Teles (2017). "Should Robots be Taxed?" NBER Working Paper No. 23806.

Guner, Nezih, Remzi Kaygusuz, and Gustavo Ventura (2014). "Income Taxation of U.S. Households: Facts and Parametric Estimates." *Review of Economic Dynamics*, 17(4), 559–581.

Guvenen, Fatih, Greg Kaplan, Jae Song, and Justin Weidner (2017). "Lifetime Incomes in the United States over Six Decades." NBER Working Paper No. 23371.

Guvenen, Fatih, Burhanettin Kuruscu, and Serdar Ozkan (2014). "Taxation of Human Capital and Wage Inequality: A Cross-Country Analysis." *Review of Economic Studies*, 81(2), 818–850.

Hassler, John, Per Krusell, Kjetil Storesletten, and Fabrizio Zilibotti (2008). "On the Optimal Timing of Capital Taxes." *Journal of Monetary Economics*, 55(4), 692–709.

Heathcote, Jonathan, Fabrizio Perri, and Giovanni L. Violante (2010a). "Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States, 1967-2006." *Review of Economic Dynamics*, 13(1), 15–51.

Heathcote, Jonathan, Fabrizio Perri, and Giovanni L. Violante (2020a). "The Rise of US Earnings Inequality: Does the Cycle Drive the Trend? ." *Review of Economic Dynamics*, 37, 181–204.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante (2010b). "The Macroeconomic Implications of Rising Wage Inequality in the United States." *Journal of Political Economy*, 118(4), 681–722.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante (2013). "From Wages to Welfare: Decomposing Gains and Losses from Rising Inequality." In *Advances in Economics and Econometrics: Tenth World Congress*, vol. 2, edited by Daron Acemoglu, Manuel Arellano, and Eddie Dekel, pp. 235–280. Cambridge University Press.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante (2014). "Consumption and Labor Supply with Partial Insurance: An Analytical Framework." *American Economic Review*, 104(7), 2075–2126.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante (2017). "Optimal Tax Progressivity: An Analytical Framework." *Quarterly Journal of Economics*, 132(4), 1693–1754.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante (2020b). "Optimal Progressivity with Age-Dependent Taxation." *Journal of Public Economics*, 189, 104074.

Heathcote, Jonathan and Hitoshi Tsujiyama (2019). "Optimal Income Taxation: Mirrlees Meets Ramsey." Mimeo, Göthe University Frankfurt.

Hubmer, Joachim, Per Krusell, and Anthony A. Smith Jr. (forthcoming). "Sources of U.S. Wealth Inequality: Past, Present, and Future." In *NBER Macroeconomics Annual 2020*, vol. 35, edited by Martin Eichenbaum and Erik Hurst. University of Chicago Press.

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

Jakobsson, Ulf (1976). "On the Measurement of the Degree of Progression." *Journal of Public Economics*, 5(1-2), 161–168.

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

Kapička, Marek (2020). "Quantifying the Welfare Gains from History Dependent Income Taxation." Mimeo.

Katz, Lawrence F. and Kevin M. Murphy (1992). "Changes in Relative Wages, 1963-1987: Supply and Demand Factors." *Quarterly Journal of Economics*, 107(1), 35–78.

Kindermann, Fabian and Dirk Krueger (2014). "High Marginal Tax Rates on the Top 1%? Lessons from a Life Cycle Model with Idiosyncratic Income Risk." NBER Working Paper No. 20601.

Krueger, Dirk and Alexander 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, Dirk and Alexander Ludwig (2016). "On the Optimal Provision of Social Insurance: Progressive Taxation versus Education Subsidies in General Equilibrium." *Journal of Monetary Economics*, 77, 72–98.

Krueger, Dirk, Fabrizio Perri, Luigi Pistaferri, and Giovanni L. Violante (2010). "Cross Sectional Facts for Macroeconomists." *Review of Economic Dynamics*, 13(1), 1–14.

Krusell, Per, Lee E. Ohanian, José-Víctor Ríos-Rull, and Giovanni L. Violante (2000). "Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis." *Econometrica*, 68(5), 1029–1053.

Lockwood, Benjamin B. and Matthew Weinzierl (2015). "*De Gustibus non est Taxandum*: Heterogeneity in Preferences and Optimal Redistribution." *Journal of Public Economics*, 124, 74–80.

Lockwood, Benjamin B. and Matthew Weinzierl (2016). "Positive and Normative Judgments Implicit in U.S. Tax Policy, and the Costs of Unequal Growth and Recessions." *Journal of Monetary Economics*, 77, 30–47.

Mertens, Karel and José Luis Montiel Olea (2018). "Marginal Tax Rates and Income: New Time Series Evidence." *Quarterly Journal of Economics*, 133(4), 1803–1884.

Moffitt, Robert A (2020). "Reconciling Trends in US Male Earnings Volatility: Results from a Four Data Set Project." NBER Working Paper No. 27664.

Moll, Benjamin, Lukasz Rachel, and Pascual Restrepo (2019). "Uneven Growth: Automation's Impact on Income and Wealth Inequality." Mimeo, Princeton University.

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

Pavoni, Nicola, Ofer Setty, and Giovanni L. Violante (2016). "The Design of 'Soft' Welfare-to-Work Programs." *Review of Economic Dynamics*, 20, 160–180.

Pavoni, Nicola and Giovanni L. Violante (2007). "Optimal Welfare-to-Work Programs." *Review of Economic Studies*, 74(1), 283–318.

Perese, Kevin (2017). "CBO's New Framework for Analyzing the Effects of Means-Tested Transfers and Federal Taxes on the Distribution of Household Income." Congressional Budget Office Working Paper 2017-09.

Piacquadio, Paolo Giovanni (2017). "A Fairness Justification of Utilitarianism." *Econometrica*, 85(4), 1261–1276.

Piketty, Thomas and Emmanuel Saez (2007). "How Progressive is the U.S. Federal Tax System? A Historical and International Perspective." *Journal of Economic Perspectives*, 21(1), 3–24.

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

Saez, Emmanuel and Gabriel Zucman (2019). *The Triumph of Injustice: How the Rich Dodge Taxes and How to Make Them Pay*. W.W. Norton & Company.

Saez, Emmanuel and Gabriel Zucman (forthcoming). "The Rise of Income and Wealth Inequality in America: Evidence from Distributional Macroeconomic Accounts." *Journal of Economic Perspectives*.

Scheuer, Florian and Iván Werning (2017). "The Taxation of Superstars." *Quarterly Journal of Economics*, 132(1), 211–270.

Slemrod, Joel and Jon Bakija (2017). *Taxing Ourselves: A Citizen's Guide to the Debate over Taxes*. 5th Edition, MIT Press.

Stantcheva, Stefanie (2017). "Optimal Taxation and Human Capital Policies over the Life Cycle." *Journal of Political Economy*, 125(6), 1931–1990.

Stantcheva, Stefanie (2020). "Dynamic Taxation." *Annual Review of Economics*, 12, 801–831.

Stiglitz, Joseph E. (1985). "Equilibrium Wage Distributions." *Economic Journal*, 95(379), 595–618.

Thuemmel, Uwe (2019). "Optimal Taxation of Robots." CESifo Working Paper No. 7317.

Wu, Chunzan (forthcoming). "More Unequal Income but Less Progressive Taxation." *Journal of Monetary Economics*.