Redistribution in Growing Economies

Axelle Ferriere
PSE & CNRS

Philipp Grübener
Goethe University

Dominik Sachs
University of St. Gallen*

February 13, 2022

Abstract

We analyze the dynamics of the equity-efficiency trade-off along the growth path. To do so, we incorporate the optimal income taxation problem into a state-of-the-art multi-sector structural change general equilibrium model with non-homothetic preferences. We identify two key opposing forces. First, long-run productivity growth allows households to shift their consumption expenditures away from necessities. This implies a reduction in the dispersion of marginal utilities, and therefore calls for a welfare state that declines along the growth path. Yet, economic growth is also systematically associated with an increase in the skill premium, which raises inequality and the desire to redistribute. We quantitatively analyze these opposing forces for two countries: the U.S. from 1950 to 2010, and China from 1989 to 2009. Optimal redistribution decreases at early stages of development, as the role of non-homotheticities prevails. At later stages of development the rising income inequality dominates and the welfare state should become more generous.

Keywords: Fiscal Policy, Growth, Non-Homothetic Preferences, Redistribution
JEL Codes: E62, H21, O11, O23, O41

*axelle.ferriere@psemail.eu, gruebener@econ.uni-frankfurt.de, dominik.sachs@unisg.ch. This paper uses data from the US Census and the American Community Survey provided by IPUMS USA (Ruggles, Flood, Goeken, Grover, Meyer, Pacas, and Sobek 2020).

1 Introduction

Fiscal redistribution is a central tool for governments to reduce poverty and inequality. A large literature has analyzed the optimal design of the welfare state, which is shaped by the classic trade-off between efficiency and redistribution: higher taxes allow for a more generous redistribution, but disincentivize labor supply. This literature has primarily focused on steady-states and business-cycle fluctuations in one-sector economies, trying to understand how the welfare state should look like in stationary economies and how it should adjust during expansions or recessions.

We adopt a different perspective and consider long-run productivity growth. Growth has potentially large and conflicting effects on poverty and inequality. On the one hand, as an economy becomes richer, absolute poverty tends to decrease;[1] necessities account for a decreasing share of aggregate consumption, as emphasized in the structural change literature. On the other hand, a growing economy typically reallocates towards high-skill intensive industries, which increases the demand for high-skilled workers. The skill premium rises, and thus inequality tends to increase.[2] Finally, growth may also change labor elasticities over time, altering the efficiency costs of taxation. Overall, long-run productivity growth generates non-trivial dynamics for the costs and needs for redistribution over time.

This paper analyzes how the welfare state should adjust over time in growing economies. To that end, we combine the workhorse models of public finance and macro development. We build a heterogeneous-agent variant of a state-of-the-art structural change general equilibrium model with non-homothetic CES preferences (Buera, Kaboski, Rogerson, and Vizcaino 2021; Comin, Lashkari, and Mestieri 2021). We incorporate the optimal income taxation problem into this framework to analyze the dynamics of the equity-efficiency trade-off due to aggregate growth and changes in the composition of consumption.

First, we study how the optimal fiscal plan changes with growth in a simplified partial equilibrium set-up. We characterize the fully optimal nonlinear tax schedule in the spirit of Mirrlees (1971) and show that, absent changes in prices and wages, the main channel at play with non-homothetic preferences is a reduction in the dispersion of marginal utilities. As a result, optimal redistribution decreases over time. We also document that the optimal tax schedule is well approximated by the parametric function developed in Ferriere, Grübener, Navarro, and Vardishvili (2021), which motivates us to analyze optimal nonlinear taxes using this parametric tax function in the general equilibrium environment.

We then turn to the general equilibrium model, to account for the fact that sectoral

[1] See Ferreira and Ravallion (2011).

[2] This mechanism, which is referred to as "skill-biased structural change" by Buera, Kaboski, Rogerson, and Vizcaino (2021), is further amplified by skill-biased technical change.

reallocation along the growth path and skill-biased technological change endogenously increase the skill premium. We calibrate our model to two countries: the U.S. from 1950 to 2010, and China from 1989 to 2009. We find that optimal redistribution decreases at early stages of development even when inequality rises, as the role of non-homotheticities prevails. On the other hand, at later stages of development the rising skill premium and income inequality dominate, so that the optimal tax-and-transfer system becomes more redistributive.

We build a rich structural change model with three sectors, agriculture, manufacturing goods, and services, and households heterogeneous in their skill and productivity. The key ingredients of the model are (i) non-homothetic constant-elasticity-of-substitution preferences à la Comin, Lashkari, and Mestieri (2021) with sector-specific income elasticities; and (ii) differential skill intensities across sectors: Agriculture uses high-skilled labor relatively little, whereas services represent the most high-skill intensive sector. In poor countries, households spend a large share of their income on agricultural necessity goods. With aggregate growth and non-homothetic preferences, households endogenously shift their consumption away from agricultural goods, toward goods and eventually services. This pattern creates the typical hump-shaped profile for the manufacturing share in GDP and an ever increasing share of services in aggregate output. In addition to this demand-driven structural change, the model allows for heterogeneous growth rates of productivity across sectors and skill-biased technical change. In general equilibrium, the increasing demand for high-skilled workers will generate an endogenous increase in the skill premium.

We start with a simplified partial equilibrium setup to analyze the effect of growth on optimal redistribution in the presence of non-homothetic preferences. In the canonical optimal taxation problem with homothetic preferences aggregate growth is irrelevant for optimal redistribution; the only relevant statistic is relative inequality. With non-homothetic preferences, growth reduces dispersion in marginal utilities, and thus reduces the need for redistribution over time. In a low-productivity economy, poor households spend most of their income on agricultural necessity goods. As the economy grows, all households eventually decrease their share of spending on agricultural goods, reflecting the fact that primary needs are satisfied. Marginal utilities decrease, especially for the poorer. With non-homothetic preferences, growth also affects labor supply elasticities. However, we show that this effect is quantitatively smaller. We calibrate the partial equilibrium model to the United States in 1950. We pursue an inverse optimum approach and choose welfare weights such that the observed 1950 tax-and-transfer system is optimal given the chosen welfare weights. Then, we increase all incomes by a factor corresponding to GDP per capita growth until 2010, keeping relative inequality and relative prices constant. We show that because of the economy getting richer and absolute poverty concerns becoming less severe marginal tax rates fall across the entire income distribution. This translates into smaller transfers paid out to households, so that the welfare state is much less generous.

We then return to the general equilibrium model to analyze the dynamics of the welfare state when also accounting for price and wage changes. We calibrate the model to match salient features of structural change in the U.S. from 1950 to 2010 and in China from 1989 to 2009. In both countries over the respective time period the skill premium rose significantly. The model accounts for that with a combination of skill-biased structural change and skill-biased technical change.

In our calibrated model optimal redistribution should become more generous over time in the United States. However, this optimal evolution of the welfare state is non-monotonic. From 1950 to 1980, the welfare state should become less generous, as the rise in inequality through an increasing skill premium is relatively small and the effect of non-homotheticities is relatively large. From 1980 to 2010 there is a much larger increase in inequality and non-homotheticities play a smaller role because the economy is already larger to start with. Therefore, the rising inequality effect dominates and the government should redistribute more. We decompose the changes further and show that the welfare state should become less generous over time because of aggregate growth and because relative prices shift such that the consumption baskets of the poor become relatively cheaper. On the other hand, the welfare state should become more generous because the skill premium and the share of high-skilled individuals rises.

The effects of non-homotheticities are even more important for China. From 1989 to 2009 there is a large rise in the skill premium in China. On the other hand, there is also very strong growth in GDP per capita. The agriculture share was initially quite high and decreased substantially over that time period. Even with the large rise in inequality the welfare state should not become much more redistributive; over the first half of that period it should have even become much less generous because of the strong effect of non-homotheticities.

Related Literature. This paper relates to two large but separate literatures. A large literature studies optimal taxation in stationary environments. One strand of this literature following Mirrlees (1971) analyzes the unrestricted fully non-linear optimal tax schedule in usually relatively simple models (Diamond 1998; Saez 2001). Much of this strand of the literature focuses on static environments, while Farhi and Werning (2013) and Golosov, Troshkin, and Tsyvinski (2016) extend the approach to dynamic settings. Another strand of the optimal taxation literature following Ramsey (1927) exogenously restricts the tax instruments available to the government, but usually analyzes richer models (Bhandari, Evans, Golosov, and Sargent 2017; Heathcote, Storesletten, and Violante 2017; Werning 2007). Relative to the optimal taxation literature, we embed the optimal tax problem into a state-of-the-art model of structural transformation and aggregate growth instead of considering a stationary environment.

The second large literature this paper relates to is the literature on structural transformation. This shift of the economy out of agriculture to manufacturing and increasingly services has long been considered to be one of the key features of growing economies (Kuznets 1973). The literature usually considers two main drivers of the structural transformation. The first driver are income effects through non-homothetic preferences, as emphasized by Kongsamut, Rebelo, and Xie (2001). As households become richer, they demand more and more manufactured goods and services such that the economy shifts out of agriculture towards these sectors. The second key driver is differential productivity growth across sectors (Ngai and Pissarides 2007). Both of these forces are considered to be important drivers of the structural transformation and are therefore both incorporated into state-of-the-art models of structural change (Duarte and Restuccia 2010; Herrendorf, Rogerson, and Valentinyi 2014), so we also allow for both of them in our framework. While much of this literature focuses on representative agent models, we build on Buera, Kaboski, Rogerson, and Vizcaino (2021) and Fang and Herrendorf (2021), who incorporate some limited household heterogeneity. Most of the literature uses Stone-Geary preferences to capture non-homotheticities (Geary 1950). However, a number of recent papers have shown that other non-homothetic preferences fit some important features of the data better (Boppart 2014; Święcki 2017). In that respect, we use the non-homothetic CES preferences proposed for a model of structural change by Comin, Lashkari, and Mestieri (2021).

Besides these two large literatures, this paper shares the interest in designing the welfare state in growing economies with Song, Storesletten, Wang, and Zilibotti (2015). Their focus is on the design of the Chinese pension systems.

Roadmap. The paper proceeds as follows. Section 2 summarizes some empirical regularities through which growth might matter for optimal redistribution. In Section 3 we introduce the partial equilibrium setup, in which we incorporate aggregate growth and non-homothetic preferences into the optimal income taxation problem. In Section 4 we extend the framework to general equilibrium to also capture the dynamics of inequality. We calibrate the model to the United States and China and perform the optimal tax analysis in general equilibrium. Section 5 concludes.

2 Motivating Facts

Before going to the model, we describe a number of key observations about economic growth. We focus on observations that might matter for how much a government might want to redistribute.

Observation 1: Absolute poverty. Economic growth can be a powerful mechanism to alleviate absolute poverty. Ferreira and Ravallion (2011) show that it is a robust feature of long-run growth that it lifts people out of absolute poverty. As a particularly impressive example, consider the case of China. According to data provided by the World Bank, in 1990 around two thirds of the Chinese population had less than \$1.90 (at 2011 international prices) available per day. This share fell to almost zero in 2016.[3] This may be important for a government's taste for redistribution, as redistribution might provide the largest benefits if it is directed to the very poor.

Observation 2: Structural change. One of the most prominent features associated with aggregate growth is the structural transformation (Jones 2016). The classic split of the economy is into three sectors, which are agriculture, goods/manufacturing, and services. It is a very robust regularity across countries that the nominal agriculture share is constantly declining, the manufacturing share is hump-shaped, and the services share is rising. In combination with the next two observations this may have important implications for redistribution.

Observation 3: Relative prices. A related empirical regularity is the evolution of relative prices. It is a robust feature of development that the relative price of services is rising with growth (Buera and Kaboski 2012; Duarte and Restuccia 2017). This may have important implications for redistribution, as households at different positions in the income distribution consume very different consumption baskets. Price changes of different consumption baskets associated with growth may affect the need for redistribution.

Observation 4: Skill intensity. The next related observation is that the service sector tends to be the most high-skill intensive in the sense that it uses a lot of high-skilled labor in production. The shift towards this sector may therefore drive up the demand and price of high-skilled labor, with important implications for inequality, as pointed out by Buera and Kaboski (2012) and Buera, Kaboski, Rogerson, and Vizcaino (2021).

Observation 5: Inequality. A final robust feature observed for a variety of countries is the rise of the skill premium. Wage inequality between college educated and non-college educated workers is increasing. This is well known for the U.S. (Katz and Murphy 1992), but also true for other countries such as China (Fang and Herrendorf 2021). Technical change is closely related to this phenomenon through the shift towards high-skill intensive sectors described above or skill-biased technical change.

[3]The data is from the World Development Indicators.

3 Mirrleesian Setup in Partial Equilibrium

In this part of the paper we want to isolate and quantify the effect of growth on optimal redistribution when preferences are non-homothetic. For this purpose, we incorporate non-homothetic CES preferences à la Comin, Lashkari, and Mestieri (2021) into the workhorse model of optimal taxation following Mirrlees (1971), Diamond (1998), and Saez (2001).

3.1 Households

There is a continuum of heterogeneous households, who are characterized by their labor productivity θ. They choose labor supply n, so that their pre-tax labor income is given by $y = \theta n$. They consume goods from three sectors: agriculture, goods, and services. Households' preferences are defined by the following utility function over a consumption aggregator $C = \mathcal{C}\left(C_A, C_G, C_S\right)$, defined below, and labor supply n:

$$U(C,n) = \frac{C^{1-\gamma}}{1-\gamma} - B\frac{n^{1+\frac{1}{\varepsilon}}}{1+\frac{1}{\varepsilon}}, \tag{1}$$

where γ is the coefficient of relative risk aversion, ε is the Frisch elasticity of labor supply, and B is an additional labor disutility parameter.

Households are assumed to be hand-to-mouth. Therefore, they choose consumption and labor subject to their static budget constraint

$$p_A C_A + p_G C_G + p_S C_S = y - \mathcal{T}\left(y\right), \tag{2}$$

where p_j denotes the price of commodity j and $\mathcal{T}\left(y\right)$ is the nonlinear income tax schedule faced by the households.

For the consumption aggregator, we assume non-homothetic CES preferences following Comin, Lashkari, and Mestieri (2021).[4] It is standard in the literature on structural transformation to assume non-homothetic preferences. Indeed, non-homothetic preferences are frequently identified as one of the most important drivers of structural change (Kongsamut, Rebelo, and Xie 2001). The most common specification of non-homotheticities are preferences of the Stone-Geary type (Herrendorf, Rogerson, and Valentinyi 2014). Non-homothetic CES preferences share key properties with Stone-Geary preferences. Appropriately parameterized, agricultural goods are necessities, whereas services are luxuries. Stone-Geary preferences, however, have some disadvantages, which non-homothetic CES preferences overcome. First, the non-homotheticities vanish asymptotically meaning that as countries grow richer preferences behave as if they were homoth-

[4] To be precise, we use the formulation from Comin, Lashkari, and Mestieri (2017).

etic. Second, they imply marginal propensities to consume out of a change in permanent income that are constant across income levels. Third, they imply a functional relationship between income and price elasticities of demand. All these features are rejected by the data. None of the criticisms applies to non-homothetic CES preferences.

Utility is defined through the following constraint:

$$\sum_{j}^{J} \left(\Omega_j C^{\varepsilon_j}\right)^{\frac{1}{\sigma}} C_j^{\frac{\sigma-1}{\sigma}} = 1. \tag{3}$$

The expenditure function implied by these preferences is

$$E\left(C; \mathbf{p}\right) = \left[\sum_{j}^{J} \Omega_j C^{\varepsilon_j} p_j^{1-\sigma}\right]^{\frac{1}{1-\sigma}} \tag{4}$$

and the Hicksian demand function can be written as

$$C_j = \Omega_j \left(\frac{p_j}{E}\right)^{-\sigma} C^{\varepsilon_j}. \tag{5}$$

With these preferences, σ is the elasticity of substitution between goods. ε_j governs the income elasticity of demand. C is a consumption index, whereas C_j denotes the consumption of individual commodities.[5]

At first sight, when looking at the definition in equation 3, these preferences may seem hard to work with. They yield, however, a quite tractable demand system. This can be seen from the Hicksian demand function in equation 5. It can be easily seen from this equation that a good's share in expenditure will go up in income if ε_j is large and that the expenditure share will go down if ε_j is small. Hence, for a good to be a luxury, it has to have a large ε_j, whereas a low ε_j characterizes a necessity. Comin, Lashkari, and Mestieri (2021) estimate the preferences using micro data for countries at different stages of development and find stable parameters with the ε_j being largest for services and lowest for agriculture, with goods having an income elasticity between the two other sectors.

3.2 Government

The government chooses a fully unrestricted nonlinear tax schedule as in Mirrlees (1971), Diamond (1998), and Saez (2001). To define the government problem, we first write the

[5]We derive the expenditure function and the Hicksian demand function in Appendix A.

household maximization problem:

$$
\begin{aligned}
V(\theta;\mathcal{T}(\cdot)) &\equiv \max_{C,n} \ U(C,n) \ \text{ s.t.} \\
p_A C_A + p_G C_G + p_S C_S &= n\theta - \mathcal{T}(n\theta).
\end{aligned}
\tag{6}
$$

Households choose consumption C, defined by the non-homothetic CES consumption aggregator, and labor n to maximize their utility subject to their static budget constraint: Spending on the three commodities has to equal after-tax income. We denote by $V(\theta;\mathcal{T}(\cdot))$ the household value obtained by the optimal choice given the nonlinear tax schedule and by $n(\theta;\mathcal{T}(\cdot))$ the labor policy function.

Let $\{w(\theta)\}$ denote Pareto weights and G an exogenous spending requirement. In partial equilibrium we do not have to specify which commodities the government consumes. Then, we can write the government's maximization problem as

$$
\begin{aligned}
&\max_{\mathcal{T}(\cdot)} \int_{\underline{\theta}}^{\overline{\theta}} V(\theta;\mathcal{T}(\cdot))w(\theta)f(\theta)d\theta \ \text{ s.t.} \\
&[\lambda]: \int_{\underline{\theta}}^{\overline{\theta}} \mathcal{T}(n(\theta;\mathcal{T}(\cdot))\theta)f(\theta)d\theta \geq G.
\end{aligned}
\tag{7}
$$

The government chooses a nonlinear tax schedule to maximize the integral over household utilities, weighted with the Pareto weights and the mass of households of a given productivity $f(\theta)$. The government is constrained by its budget constraint, stating that the revenues from the income tax have to cover the exogenous spending requirement. We denote by λ the multiplier on the government budget constraint, which is the marginal value of public funds.

3.3 Optimal Tax Formula

The optimal tax formula can be derived using standard techniques. Specifically, we use a perturbation approach following Saez (2001) to derive the optimal tax formula. We relegate the derivation to the appendix. Optimal marginal tax rates for a certain skill level θ^* are given by

$$
\forall\, \theta^*: \ \frac{\mathcal{T}'(y(\theta^*))}{1-\mathcal{T}'(y(\theta^*))} = \left(1+\frac{1}{\varepsilon}\right)\left(1 - \mathop{\mathbb{E}}_{\theta>\theta^*}\left[\frac{u'(\theta)}{E'(\theta)}\frac{w(\theta)}{\lambda} - \eta(\theta)\mathcal{T}'(y(\theta))\right]\right)\frac{1-F(\theta^*)}{f(\theta^*)\theta^*}.
$$

This formula describes the standard efficiency-redistribution trade-off. The first term, $\left(1+\frac{1}{\varepsilon}\right)$, captures efficiency concerns. This tax formula is an inverse elasticity rule: The larger labor supply elasticities, the stronger individuals respond to higher taxes, and the lower marginal tax rates should be optimally. The term $\frac{u'(\theta)}{E'(\theta)}\frac{w(\theta)}{\lambda}$ accounts for redistributive concerns. If individuals with a higher skill than θ^* have very high incomes, their marginal utilities will be low, making this term small. Given that the term is subtracted, this means that this will be a force towards higher marginal tax rates. With high income inequality, implying low marginal utilities at the top, it is optimal to have high marginal tax rates because this raises large amounts of revenue from the top of the distribution to be redistributed at the bottom. $\eta(\theta)$ captures income effects. When marginal taxes at θ go up, this lowers incomes of everybody earning more. With leisure being a normal good, this will increase their labor supply. Hence, income effects are a force for higher marginal tax rates. Finally, the shape of the skill distribution matters. With a lot of mass above a certain skill level, there are many people whose average taxes are raised by an increase of marginal taxes at θ^*. This is a reason for higher marginal tax rates. However, if there is a lot of mass at this skill level, then distortions will be more important, so that lower taxes will be optimal.

Our description of the forces determining optimal taxes so far would apply equally to the case with homothetic preferences. They key difference applies to the term accounting for redistributive concerns $\frac{u'(\theta)}{E'(\theta)}\frac{w(\theta)}{\lambda}$. With homothetic preferences, $E'(\theta)$ is always equal to one for every skill level and therefore also independent of economic growth. However, with non-homothetic CES preferences this term causes redistributive concerns to become smaller as an economy becomes richer. Keeping relative inequality constant, increasing the incomes of all agents reduces the benefits from distribution. This is how the shift away from absolute poverty and the consumption of necessities enters the optimal tax formula.

Note that also the income effect term is affected by the presence of non-homotheticities. We derive the formula for the income effect in the appendix. We show in the next part, however, that the key change with non-homothetic preferences is to the taste for redistribution.

3.4 Calibration

We now evaluate the quantitative significance of growth in combination with non-homothetic preferences by computing the optimal income tax schedules in an economy that is calibrated to the U.S. in 1950 and for a counterfactual economy, in which every individual is richer by a factor corresponding to U.S. real GDP per capita growth from 1950 to 2010. For this exercise, we keep relative inequality constant.

For the calibration, we assume that the income distribution follows a log-normal distribution with a Pareto tail. To discipline the income distribution we use data from the 1950 U.S. Census (Ruggles, Flood, Goeken, Grover, Meyer, Pacas, and Sobek 2020). As income measure we use wage and salary income. Because incomes are reported only in bins we refrain from estimating the parameters of the distributions directly. Rather, we set them such that the implied income distribution is consistent with the income

Table 1: 1950 U.S. Income Distribution

Decile	**Data**	**Model**
1	0.96%	2.47%
2	2.92%	4.24%
3	4.90%	5.07%
4	6.75%	7.13%
5	8.34%	7.41%
6	10.00%	9.67%
7	11.69%	10.46%
8	13.33%	13.25%
9	15.68%	15.44%
10	25.43%	24.85%

Notes: This table compares the income distributions for the U.S. in 1950 in the data and in the model.

distribution in the data. We assume that the income distribution for the majority of the population follows a log-normal distribution and only adjust the incomes of the top 5% such that they follow a Pareto distribution. We let the Pareto parameter decline linearly to ensure a smooth hazard ratio $\frac{1-H(y)}{h(y)}$, as in Sachs, Tsyvinski, and Werquin (2020). The income distributions in the data and in the model are shown in Table 1.

We fix a number of preferences parameters exogenously. We set the coefficient of relative risk aversion γ equal to two and the Frisch elasticity of labor supply ε equal to 0.2. We set the non-homothetic CES parameters $\sigma = 0.3$, $\varepsilon_A = 0.1$, $\varepsilon_G = 1.0$, and $\varepsilon_S = 1.8$, in line with the estimates of Comin, Lashkari, and Mestieri (2021). We use these parameter values also in the general equilibrium model, so these parameters are also summarized in Table 3. In partial equilibrium, we also have to set the prices of the three commodities exogenously. Prices and the Ω_j parameters of the non-homothetic CES preferences cannot be distinguished from each other. They are jointly set such that we match the nominal sector shares of the U.S. in 1950, with roughly 60% of the economy in the services sector, 34% in the goods sector, and 6% in the agricultural sector. Because we match these targets exactly in our general equilibrium model, we incorporate the prices and preference parameters from the GE calibration into the PE environment. This gives us a very close match for the sector shares in PE.

We calibrate taxes and spending to be consistent with taxes and spending of the U.S. federal government in 1950. We assume that the government raises revenues using a parametric tax function. The tax payment of an individual earning income y is given in equation 8.

$$\mathcal{T}(y) = \exp\left[\log(\lambda)\, y^{-\tau}\right] y - T. \tag{8}$$

Individuals pay taxes on their income, where tax progressivity is determined by the parameter τ and the level of tax rates is determined by the parameter λ. Also, households receive a lump-sum transfer T. The level of the lump sum transfer is set to match government spending on income security (roughly 1% of GDP in 1950). Exogenous government spending G is set to account for all other federal government spending (roughly 14% of GDP). The parameters of the tax function are set to match average marginal tax rates along the income distribution.[6]

Finally, we back out the underlying skill distribution from the income distribution given all other parameters using the households' first order conditions, as in Saez (2001). For the optimal taxation problem, we follow an inverse optimum approach. That is, we compute Pareto weights such that the calibrated 1950 tax schedule is optimal given these weights (Bourguignon and Spadaro 2012; Christiansen 1977; Hendren 2020; Lockwood and Weinzierl 2016).

For our counterfactual 2010 economy, we scale down all prices by the same magnitude, equivalent to a proportional increase in income for everybody. We choose the magnitude to be the increase in GDP per capita from 1950 to 2010. We recalibrate taxes and spending to be consistent with the distribution of average marginal tax rates and government spending in the U.S. in 2010. Economic growth in combination with non-homothetic preferences leads to a shift in the consumption allocation across the different goods: Richer households spend less on agriculture and goods and more on services. The agriculture share falls from 6% to 2.5%; the goods share declines from 34% to 25%; and the services share rises from 60% to 73%. Note that this underestimates the structural change that we observe in the United States in the data. The reason is that we are looking at nominal sector shares but keep relative prices constant. However, it is well known that the relative price of services is rising with development ("Baumol's cost disease", Baumol (1967)). In the general equilibrium model we capture this force and thereby are able to match nominal sector shares. In this partial equilibrium framework we only capture a real shift towards services consumption and away from necessities, so that we do not account for the entire increase in the nominal service share.

Note that while relative inequality in incomes does not change by construction, the underlying inequality in the skill distribution does change. We are going to revisit the importance of that change at the end of the results section.

3.5 Optimal Taxes in Partial Equilibrium

Figure 1 depicts the optimal marginal tax rates in the two economies. The 1950 optimal marginal tax rates schedule is the calibrated tax schedule by virtue of the inverse optimum approach. Marginal tax rates are increasing along the income distribution. In addition to financing the exogenous government spending, the government uses the tax revenues

[6] This calibration of taxes and spending is similar to the general equilibrium model. We discuss data and targets in more detail there.

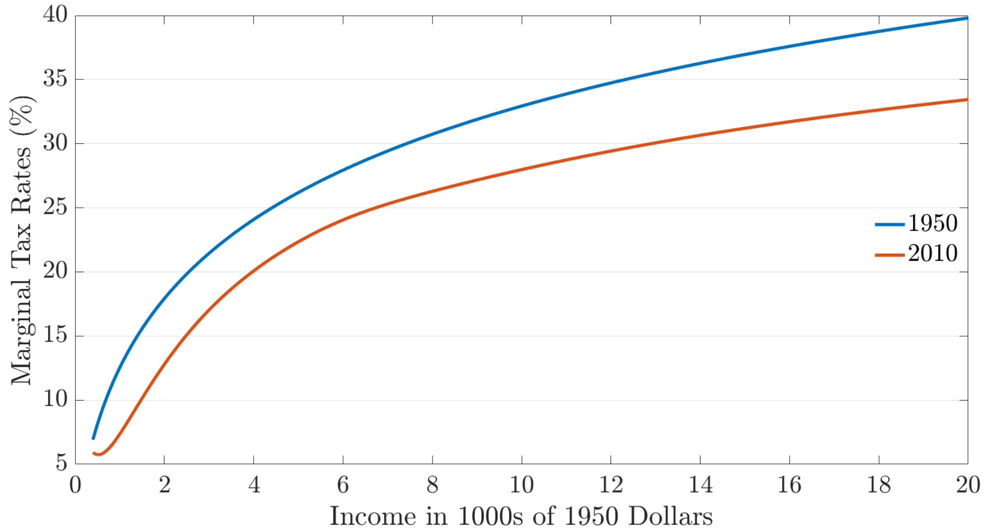

Figure 1: Optimal Marginal Tax Rates

Notes: This figure compares optimal marginal tax rates in the calibrated 1950 economy and the counterfactual 2010 economy, which is characterized by higher income levels.

to give a lump sum transfer of around 1.1% of GDP. This lowers average tax rates of the entire tax-and-transfer system to zero at the bottom, which is shown in Figure 2.

The two figures also show the optimal marginal and average tax schedules for the counterfactual 2010 economy. Optimal marginal rates are lower across the entire income distribution. Quantitatively, the drop is roughly 5% at most income levels, with a slightly smaller difference at the bottom of the distribution. These lower tax rates are not sufficient to finance the entire exogenous government spending, so that instead of a lump sum transfer there is a lump sum tax of 2% of GDP. This optimal system then implies very different average tax rates. The lump sum tax matters a lot for average tax rates at the bottom of the distribution. The poorest individuals now face much higher average tax rates. By contrast, average tax rates are much lower for high earners because the lump sum tax matters less for them and the marginal tax rates schedule is shifted downwards.

Hence, with non-homothetic preferences aggregate growth is associated with less redistribution. Absolute poverty is less of a concern and the consumption baskets also of the poorest contain more of the relative luxury good services and less of the necessity agriculture. We now show that it is indeed this declining taste for redistribution that accounts for most of the change to the optimal marginal tax rates and not the changing skill distribution or income effects.

Figure 3 shows again the marginal tax rates schedule for 1950 and 2010. Additionally, we consider two intermediate cases. First, we compute optimal rates for a scenario in which prices are at the 2010 level, i.e. aggregate growth has happened. However, we keep labor supply and the skill distribution fixed such that not just income inequality

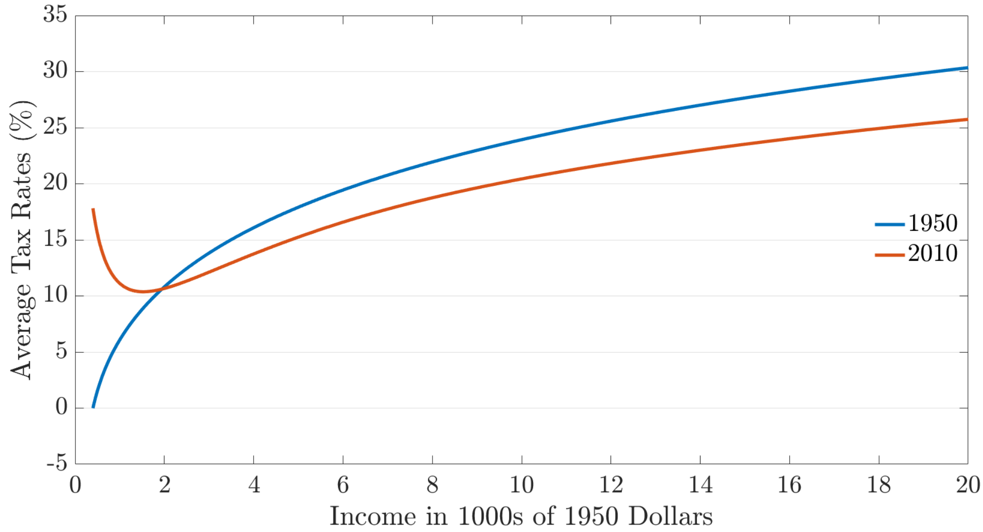

Figure 2: Optimal Average Tax Rates

Notes: This figure compares optimal average tax rates, implied by the entire tax-and-transfer system including the lump sum grant, in the calibrated 1950 economy and the counterfactual 2010 economy, which is characterized by higher income levels.

but also underlying skill inequality is the same as in 1950. Note that because of the non-homotheticities labor supply is higher across the entire income distribution in 1950 compared to 2010. This effect is strongest for the lowest incomes: Because they are very poor in absolute terms in 1950, their labor supply is very high. The poorest in 2010 are much richer in absolute terms, even if their relative income share is the same. Therefore, the reduction in labor supply is largest at the bottom of the income distribution. To keep the income distribution the same, as imposed by the calibration strategy, we need that the skill distribution also changes most at the bottom of the distribution. All skill levels are lower in 1950, but the change is largest at the bottom. Therefore, if we do a counterfactual where we impose inequality in skills to be as in 1950, with income inequality being the same anyways, but with lower prices because of aggregate growth as in 2010, we have optimally higher marginal tax rates compared to 2010, bringing the tax rates schedule closer to the 1950 schedule. This is shown in Figure 3 in the case "2010 with 1950 inequality".

We consider a second intermediate case in which we also fix the income effects at the 1950 level. Income effects are larger in 1950 than in 2010, as on average poorer individuals respond more strongly to receiving unearned income. As explained above, larger income effects are a force for higher marginal tax rates. Increasing marginal tax rates at some point in the distribution is more beneficial as individuals earning higher incomes will work more because of the loss through higher average taxes. Hence, the "2010 with 1950 inequality and income effects" schedule brings us even closer to the 1950 optimum.

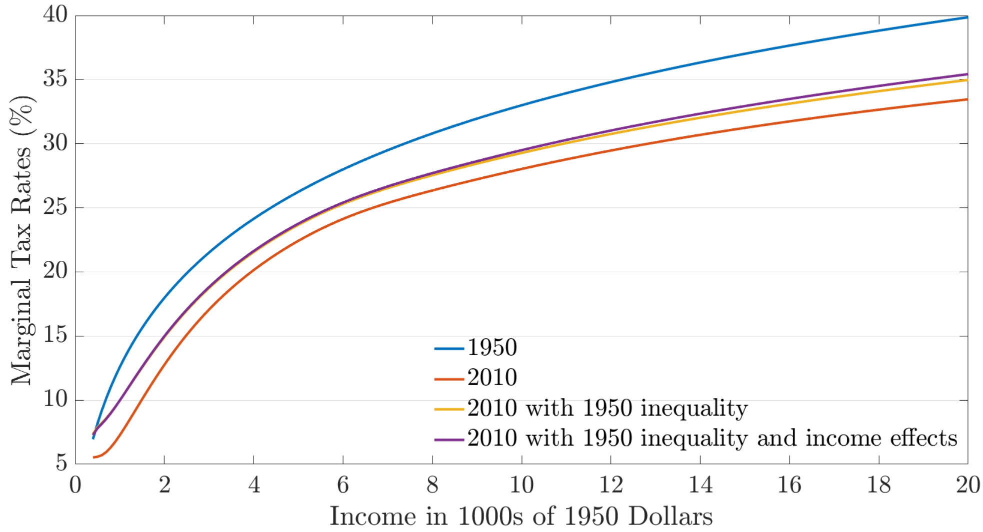

Figure 3: Optimal Marginal Tax Rates: Decomposition

Notes: This figure compares optimal marginal tax rates, implied by the entire tax-and-transfer system including the lump sum grant, in the calibrated 1950 economy and the counterfactual 2010 economy, which is characterized by higher income levels. It additionally compares two intermediate cases "2010 with 1950 inequality" and "2010 with 1950 inequality and income effects".

However, quantitatively these two changes account only for a small part of the change in the optimal tax-and-transfer systems between 1950 and 2010. Therefore, we conclude that the most important implication of non-homothetic preferences in combination with aggregate growth is the change to the taste for redistribution.

4 General Equilibrium

We now turn to the general equilibrium model, in which in addition to aggregate growth we account for changes to relative inequality and relative prices.

4.1 Households

In the general equilibrium model, household heterogeneity takes two dimensions. Households are characterized by their skill level, which can be high or low. Firms pay wages by skill level per efficiency unit of labor. Within skill, households differ in their productivity. Denote the wage a household receives with w and the productivity within skill with θ. Then, the household problem can be written as

$$\begin{aligned} &\max_{C,n} \frac{C^{1-\gamma}}{1-\gamma} - B\frac{n^{1+\varphi}}{1+\varphi} \\ &\text{s.t. } E\left(C\right) = w\theta n\left[1-\exp\left[\log\left(\lambda\right)\left(w\theta n\right)^{-\tau}\right]\right] + T. \end{aligned} \tag{9}$$

C is the non-homothetic CES consumption aggregator and n is labor supply. The budget constraint of the household equalizes consumption expenditure and after-tax-and-transfer income: λ is the level parameter of the parametric tax function, τ governs progressivity, and T is a lump sum transfer.

The first order condition of the household with respect to consumption reads

$$C^{-\gamma} = \chi E'(C), \tag{10}$$

where χ is the multiplier on the budget constraint. The first order condition with respect to labor is

$$\begin{aligned} Bn^{\varphi} &= \chi\Big\{w\theta\left[1-\exp\left[\log(\lambda)(w\theta n)^{-\tau}\right]\right] \\ &\quad + w\theta n\left(-\exp\left[\log(\lambda)(w\theta n)^{-\tau}\right]\right)\log(\lambda)(-\tau)(w\theta n)^{-\tau-1}w\theta\Big\} \\ &= \chi\Big\{w\theta\left[1-\exp\left[\log(\lambda)(w\theta n)^{-\tau}\right]\right] \\ &\quad + (w\theta)^{1-\tau}\exp\left[\log(\lambda)(w\theta n)^{-\tau}\right]\log(\lambda)\tau n^{-\tau}\Big\}. \end{aligned} \tag{11}$$

We can combine these two first order conditions to

$$0 = Bn^{\varphi} - \frac{C^{-\gamma}}{E'(C)}\left[w\theta\left[1-\exp\left[\log(\lambda)(w\theta n)^{-\tau}\right]\right] + (w\theta)^{1-\tau}\exp\left[\log(\lambda)(w\theta n)^{-\tau}\right]\log(\lambda)\tau n^{-\tau}\right]. \tag{12}$$

4.2 Production

The production side of the economy closely follows Buera, Kaboski, Rogerson, and Vizcaino (2021). The production function is given by

$$Y_{jt} = A_{jt}\left[\alpha_{jt}H_{jt}^{\frac{\rho-1}{\rho}} + (1-\alpha_{jt})L_{jt}^{\frac{\rho-1}{\rho}}\right]^{\frac{\rho}{\rho-1}}, \tag{13}$$

where the subscript j denotes the sector, with $j \in \{A, G, S\}$ for agriculture, goods, and manufacturing and t denotes the time period. Y_j is sectoral output, which is produced using high-skilled and low-skilled labor. H_j and L_j are efficiency units of high-skilled and low-skilled labor, respectively. The parameters A_{jt} denote skill-neutral technology parameters, and α_{jt} denote skill intensities. These parameters are sector-specific and time-varying. The elasticity of substitution in production ρ is constant across time and sectors.

Firms maximize profits under perfect competition. The firm problem is static, so we omit the time subscript for notational convenience. The firm problem then reads:

$$\max_{H_j,L_j} p_j A_j \left[\alpha_j H_j^{\frac{\rho-1}{\rho}} + (1-\alpha_j) L_j^{\frac{\rho-1}{\rho}}\right]^{\frac{\rho}{\rho-1}} - w_H H_j - L_j. \tag{14}$$

p_j denotes the price of commodity j. w_H is the wage paid per efficiency unit of high-skilled labor. The wage paid to low-skilled labor w_L is normalized to 1, so that w_H can be interpreted as the skill premium.

The solution of the firm problem implies that given the skill premium all prices are pinned down by parameters:

$$p_j = \frac{1}{A_j}\left[\frac{\alpha_j^\rho}{w_H^{\rho-1}} + (1-\alpha_j)^\rho\right]^{\frac{1}{1-\rho}}. \tag{15}$$

4.3 Government

The government raises income taxes to finance a lump sum transfer and exogenous government spending. The tax payment is given by the parametric tax function

$$\mathcal{T}(y) = \exp\left[\log(\lambda)\, y^{-\tau}\right] y - T, \tag{16}$$

where $\exp\left[\log(\lambda)\, y^{-\tau}\right]$ is the average tax rate applied to income $y = w\theta n$.

In addition to the lump sum transfer the government finances exogenous government spending in all sectors: G_a, G_g, G_s. The government budget constraint hence reads

$$\sum_{j=1}^{J} p_j G_j + T = \sum_{ik} \pi_{ik} w_i \theta_{ik} n_{ik} \exp\left[\log(\lambda)\,(w_i\theta_{ik}n_{ik})^{-\tau}\right], \tag{17}$$

where i denotes worker skill and k refers to within-skill productivity. π_{ik} is the mass of workers with skill i and within-skill productivity k.

4.4 Equilibrium

In equilibrium markets for high-skilled and low-skilled labor and markets for agriculture, goods, and services have to clear. From the firm problem, we can derive demand for high skilled labor as

$$H_j = \frac{Y_j}{A_j}\left(\frac{\alpha_j p_j A_j}{w_H}\right)^\rho. \tag{18}$$

Define also total consumption in each sector, consisting of private and government consumption, as

$$D_j = G_j + \sum_{ik} C_{jik}\pi_{ik}. \tag{19}$$

Imposing goods market clearing we can rewrite equation (18) as

$$H_j = \frac{D_j}{A_j} \left(\frac{\alpha_j p_j A_j}{w_H} \right)^{\rho}.$$

Then, labor market clearing in the market for high skilled labor requires

$$\sum_{Hk} \pi_{Hk} n_{Hk} \theta_{Hk} = \sum_{j=1}^{J} \frac{D_j}{A_j} \left(\frac{\alpha_j p_j A_j}{w_H} \right)^{\rho}. \tag{20}$$

This condition already imposes goods market clearing. Therefore, the only market left is the market for low-skilled labor, which clears by Walras' law if equation (20) holds.

The equilibrium definition is standard. The equilibrium consists of a set of prices and wages, given a tax-and-transfer system, such that households solve their utility maximization problem, firms solve their profit maximization problem, three goods and two labor markets clear, and the government budget is balanced. Before turning to the optimal tax problem in this economy, we discuss the calibration in the next section.

4.5 Calibration for the U.S.

Our first quantitative application of the model is to the United States. We calibrate the model to three years: 1950, 1980, and 2010. We choose parameters to be consistent with the two key determinants of optimal redistribution that we highlight: absolute poverty concerns and the level of the economy (aggregate growth, structural change) and relative inequality (changing relative prices, changing skill premium). We now describe our calibration strategy in detail, starting with a summary of the parameters to be chosen.

The first set of parameters we have to choose concerns demographics. In each period there is a share f_H of high-skilled individuals. Additionally, we have to parameterize the distribution of efficiency units within skill. We opt to keep within type inequality constant over time. While especially since the 1980s there also has been a rise of incomes of the very top earners relative to other highly educated individuals, we abstract from this change in order to not introduce more exogenous changes into the model. Another set of parameters that is kept constant over time is the set of all preference parameters. We need to choose the non-homothetic CES parameters $\sigma, \varepsilon_A, \varepsilon_G, \varepsilon_S, \Omega_A, \Omega_G, \Omega_S$ and the other preference parameters γ, φ, B. On the production side we need to set the elasticity of substitution in production ρ, which is constant over time, and the time-varying skill neutral technology parameters A_A, A_G, A_S as well as the time-varying skill intensities $\alpha_A, \alpha_G, \alpha_S$. Finally, we have to choose three parameters of the tax-and-transfer system λ, τ, T in every period.

We first fix a number of parameters exogenously. We set the non-homothetic CES

Table 2: Moments for the U.S. Calibration

Moment	1950	1980	2010
Skill premium	1.44	1.54	2.06
High-skill inc. share agriculture	5.22%	20.26%	23.62%
High-skill inc. share manufacturing	8.27%	19.96%	39.60%
High-skill inc. share services	16.46%	37.14%	57.67%
Service share	-	69.59%	-
Manufacturing share	-	28.04%	-
Real GDP per capita	0.50	1.00	1.70
Rel. price agriculture to manufacturing	1.88	1.00	0.55
Rel. price services to manufacturing	0.94	1.00	1.49
G/Y	14.00%	14.00%	14.00%
T/Y	1.12%	2.97%	3.61%
AMTR Top 10%	27.54%	46.03%	32.68%
AMTR Bottom 90%	14.94%	36.30%	27.95%

Notes: This table summarizes the moments for the U.S. calibration. Real GDP per capita and relative prices are normalized to 1 in 1980.

parameters in line with the estimates of Comin, Lashkari, and Mestieri (2021) to $\varepsilon_A = 0.1$, $\varepsilon_G = 1.0$, $\varepsilon_S = 1.8$, and $\sigma = 0.3$. These preference parameters imply that agricultural goods are necessities, while services are luxuries. We set the coefficient of relative risk aversion γ to two and the Frisch elasticity of labor supply to 0.2, in line with micro estimates. We also exogenously fix the elasticity of substitution in production ρ at 1.42, as in Buera, Kaboski, Rogerson, and Vizcaino (2018). For the within type heterogeneity we distinguish between ten groups per skill level, accounting for 10% of the total mass per skill each. We set the efficiency units for each of these deciles based on within skill wage differences from U.S. Census data. Finally, we impose a few normalizations. We set Ω_G to one. We also fix A_S to one in 1980 and pick the productivity levels in agriculture and goods such that all relative prices are one in 1980.

We continue with calibrating the 1980 economy as our base year. Assuming that the skill premium is matched, it is possible to express the three skill intensities as functions of the share of income going to high-skilled individuals in each sector. We compute this data moment from the 1950 Census 1% Sample, the 1980 Census 5% sample, and the 2010 American Community Survey (Ruggles, Flood, Goeken, Grover, Meyer, Pacas, and Sobek 2020). In the data, we consider an individual as high skilled if the person has at least a four year college degree. Everybody with less than a four year degree is considered low skilled. We restrict our sample to individuals who are between 25 and 60 years old. We use Census Bureau industrial classifications to group individuals into the three broad sectors of our model. As income measure we use total pre-tax wage and salary income. Income in the data is top-coded. We apply a simple adjustment to account for top-coding by multiplying all incomes at the top-coding threshold with 1.5. The high-skilled income shares by sector for the three years are reported in Table 2, which includes all the data moments targeted in our U.S. calibration. Two important patterns emerge from this data. First, services is generally the high-skill intensive sector, with a larger share of income going to high-skilled individuals than in the other two sectors. Hence, as an economy grows richer and households shift their consumption towards services, this will drive up the demand for high-skilled labor and thereby put upward pressure on the skill premium. This is what Buera, Kaboski, Rogerson, and Vizcaino (2021) refer to as skill-biased structural change. The second key pattern is that the high-skilled income share goes up over time in each sector. This translates into rising skill intensities, which is commonly referred to as skill-biased technical change.

The remaining parameters that are calibrated internally are the non-homothetic CES parameters Ω_A and Ω_S, the high-skilled population share f_H, the labor disutility parameter B, and all the government parameters $\lambda, \tau, T, G_A, G_G, G_S$. They are calibrated jointly; here we discuss which data moments are most closely related to each of these targets. Ω_A and Ω_S are set to match the 1980 sector shares of goods and services. We compute these data moments from National Income and Product Accounts (NIPA), based on value added by industry. To smooth fluctuations we take five year averages around the respective years. The high-skilled population share is closely related to the skill premium, which we also compute from the Census data. The skill premium is defined as the ratio of average weekly earnings of high-skilled individuals divided by average weekly earnings of low-skilled individuals. We compute weekly earnings as annual wage and salary income divided by the number of weeks in employment. The labor disutility parameter B is set such that average labor supply of the high-skilled is equal to 0.33.

This leaves the government parameters to be calibrated. We discipline government spending using historical tables provided by the Office of Management and Budget. First, we discipline the size of the lump-sum transfer by matching spending on income security programs. This has increased over time, from 1.12% in 1950 to 3.61% in 2010. Again, these numbers are averaged over five year windows. We consider every other government expenditure as part of the exogenous spending requirement. Total expenditures have risen quite significantly over time. However, one aspect to be taken into account that is not modeled is that a sizeable part of the additional spending has been debt financed. This is the case for 1980 and even more so for 2010. When accounting for debt the financing requirement through taxes was relatively stable across the three years. Therefore, we choose to keep a constant spending to output ratio in the calibration. This has the advantage that our results are not affected by changing spending requirements over time. This could have sizeable effects on desired tax progressivity (Ferriere, Grübener, Navarro, and Vardishvili 2021; Heathcote, Storesletten, and Violante 2017; Heathcote and Tsujiyama 2021), which we can abstract from by making this choice. In our three sector model we also have to take a stand on how government consumption is allocated across the three sectors. As our baseline assumption, we choose to set government spending as a proportional share of private consumption in each sector. Finally, to discipline the progressivity of the tax system we rely on average marginal tax rates estimated by Mertens and Montiel Olea (2018). The tax progressivity parameter τ is closely related to the difference between the average marginal tax rate (AMTR) faced by the top 10% of the income distribution compared to the AMTR of the bottom 90%. The remaining condition for the last government parameter λ is that the government budget has to clear.

Having calibrated the base year, we have to set the remaining parameters for the 1950 and 2010 economies. The skill intensities can be determined following the same approach as for the base year 1980 before calibrating the remaining parameters internally. Specifically, we need to determine skill-neutral technology parameters A_A, A_G, A_S, high-skilled population shares f_H, and all the government parameters. We discipline the skill-neutral technology parameters by matching growth rates in real GDP per capita and changes in relative sectoral prices. The data for sectoral prices is from NIPA, as for the sector shares. The skill-neutral technology parameters are rising over time, capturing aggregate growth. The relative price of services is falling, whereas the relative price of agriculture is rising, implying that productivity growth is lowest in services. Again, the high-skilled population share is closely related to the skill premium, which is rising over time. However, given that skill-biased technical and structural change cause a rise in the skill premium, the model requires the high-skilled population share to be rising over time, in line with the data as we discuss below. Lastly, the calibration for the government parameters is as in the base year 1980. Table 2 shows that the gap between the AMTR faced by the top 10% compared to the bottom 90% is roughly constant between 1950 and 1980, while it is significantly lower in 2010. This translates into similar progressivity parameters in 1950 and 1980, but a significantly lower one in 2010. This is in line with the estimated time series for progressivity based on the loglinear tax function in Ferriere and Navarro (2020) also based on the data by 2018. Heathcote, Storesletten, and Violante (2020), by contrast, estimate a relatively stable progressivity parameter for the loglinear tax function between 1980 and 2010. However, a key difference is that we model transfers separately. While progressivity of purely the income tax function decreases in our calibration, redistribution through transfers increases over time.

Table 4 shows the model fit for a number of untargeted moments. First, while we target the sector shares in 1980 in our calibration, we do not target the sector shares in 1950 and 2010. The model still matches these sector shares very well. It captures the 20 percentage point rise in the service share from 1950 to 2010 and the associated drops in the agriculture and goods shares. Capturing these structural changes well is key for

Table 3: Parameters for the US Calibration

Parameter	Interpretation	Value
Preferences		
σ	Elasticity of substitution	0.300
ε_A	Income elasticity agriculture	0.100
ε_G	Income elasticity goods	1.000
ε_S	Income elasticity services	1.800
Ω_A	Level parameter agriculture	0.008
Ω_G	Level parameter goods	1.000
Ω_S	Level parameter services	19.272
γ	Coefficient of risk aversion	2.000
φ	Labor supply elasticity	5.000
B	Labor disutility	4291.443
Production		
ρ	Elasticity of substitution	1.420
A_A	Neutral technology agriculture	0.157, 0.839, 3.411
A_G	Neutral technology goods	0.325, 0.836, 2.251
A_S	Neutral technology services	0.412, 1.000, 1.669
α_A	Skill intensity agriculture	0.126, 0.302, 0.351
α_G	Skill intensity goods	0.170, 0.299, 0.479
α_S	Skill intensity services	0.262, 0.440, 0.606
f_H	Population share high skilled	0.097, 0.238, 0.371
Government		
G_A	Government consumption agriculture	0.000, 0.001, 0.001
G_G	Government consumption goods	0.004, 0.006, 0.010
G_S	Government consumption services	0.007, 0.015, 0.026
T	Transfer	0.004, 0.011, 0.016
λ	Tax function level	0.210, 0.229, 0.203
τ	Tax function progressivity	0.260, 0.262, 0.163

Notes: This table summarizes the calibrated parameters for the U.S. For the time-varying parameters the values correspond to the calibration for the years 1950, 1980, and 2010.

Table 4: Untargeted Moments in the US calibration

Moment	Data	Model
Service share 1950	59.79%	61.05%
Manufacturing share 1950	33.31%	33.11%
Service share 2010	81.04%	79.27%
Manufacturing share 2010	17.93%	19.84%
High-skilled population share 1950	8.56%	9.73%
High-skilled population share 1980	21.59%	23.88%
High-skilled population share 2010	34.98%	37.12%

Notes: This table compares untargeted moments with the model implied values for the U.S. calibration.

our analysis of optimal taxes and transfers over time. The very good model fit for the sector shares gives us confidence that the non-homothetic CES preferences, parameterized according to the estimates of Comin, Lashkari, and Mestieri (2021), capture the degree of non-homotheticities very well.

Table 4 also shows the model implied high-skilled population shares over time. We need these as free parameters in the model calibration to match the skill premium exactly. The implied values are very close to the data. While there is a close relationship between the explicitly targeted high-skilled income shares by sector and the skill premium on the one hand and the high-skilled population share on the other hand, the fit does not have to be perfect because we impose the same skill premium across sectors in the model, which is not exactly the case in the data. Therefore, it is reassuring that the model implied population shares are very close to the data.

Finally, the model also captures the fact that labor supply is falling over time. Bick, Fuchs-Schündeln, and Lagakos (2018) show for a cross-section of countries that hours worked tend to be higher for countries at early stages of development compared to more developed countries. Boppart and Krusell (2020) document falling hours in the time series for the U.S. and a variety of other developed countries. We also see a fall in hours worked in the Census data.

4.6 Optimal Tax-and-Transfer System for the U.S.

The Ramsey problem in this economy is to choose the parameters of the tax-and-transfer system λ, τ, T to maximize welfare. As in the partial equilibrium setup, we follow an inverse optimum approach. To implement this, we define Pareto weights as

$$f(\theta w) = \mu + (\theta w)^{\nu}. \tag{21}$$

Pareto weights are a function of the product of the skill specific wage and the within skill efficiency units of labor. We set parameters μ and ν of the Pareto weight function such that the calibrated 1950 tax system is optimal at the observed skill premium. To make the 1950 tax-and-transfer system optimal the planner has to put a lower weight on low income and a higher weight on high income households than a utilitarian planner would do. For the years 1980 and 2010, we have to decide which Pareto weights to use. One option would be to apply the Pareto weight function using the parameters μ and ν required for 1950 to the new skill premium in 1980 and 2010. This has the disadvantage that given the estimated Pareto weights for 1950 we would increase the weights the planner puts on richer households, partially undoing the effect of the non-homotheticities on desired redistribution. Simply keeping the Pareto weights constant across groups even though the skill premium changed has the disadvantage that the planner may put different weights on individuals with different skill levels, who have the same income. Hence, as our benchmark we choose to keep μ constant across years but to adjust ν such that the ratio between the weight on the richest group and the weight on the poorest group remains the same.

In Figure 4 we show the optimal average tax rates implied by the entire tax-and-transfer system for income deciles for the three years.[7] The benchmark year is 1950, for which the optimum corresponds to the calibrated system. The overall tax-and-transfer system is progressive in the sense that average tax rates are increasing with income. In the lowest income decile the average tax rate is negative, implying that households receive a net transfer.

The optimal tax-and-transfer system in 1980 is less redistributive. The net transfer for the bottom income decile is essentially zero. Tax rates are slightly higher also for the other bottom income deciles. The top income decile, by contrast, faces a lower optimal tax rate. Which changes drive these results? The skill premium and the share of high-skilled individuals increase from 1950 to 1980. Standard optimal tax theory abstracting from level effects thus would imply that more redistribution is optimal. However, the non-homotheticity effect overcompensates the standard channel. Consumption of necessities in form of agricultural goods drops between 1950 and 1980 and consumption of the relative luxury good services rises. Therefore, the planner has a lower desire for redistribution, explaining why the optimal tax-and-transfer system becomes less redistributive.

Until 2010, however, this change is more than reversed. The rise in inequality is much stronger over this later time period, which strongly affects the desire to redistribute. Moving from 1980 to 2010, the effect of non-homotheticities is also weaker relative to the change between 1950 and 1980 because the economy is already starting from a higher level.

In Figure 5 we perform a decomposition of the change between 1950 and 2010. Over the entire time period the tax-and-transfer system becomes more progressive in average rates. We proceed in five steps. The starting point is to find the optimal tax-and-transfer system in 1950 in partial equilibrium such that we can then shut off and on different channels in the next steps. Thanks to the inverse optimum approach and relatively weak general equilibrium effects of taxes on the skill premium (see discussion below), the optimal partial equilibrium system that we find as starting point for this exercise is virtually indistinguishable from the optimal general equilibrium system shown in Figure 4.

In the second step we reduce all prices by the same factor to capture the amount of aggregate growth that took place between 1950 and 2010. Households are richer now, but they still face the same relative prices. Also, relative inequality in incomes is unchanged.

[7]We replicate this figure in Appendix B.4 for the case in which we simply keep Pareto weights the same across groups in later years. There is no meaningful difference in optimal tax rates across the two cases.

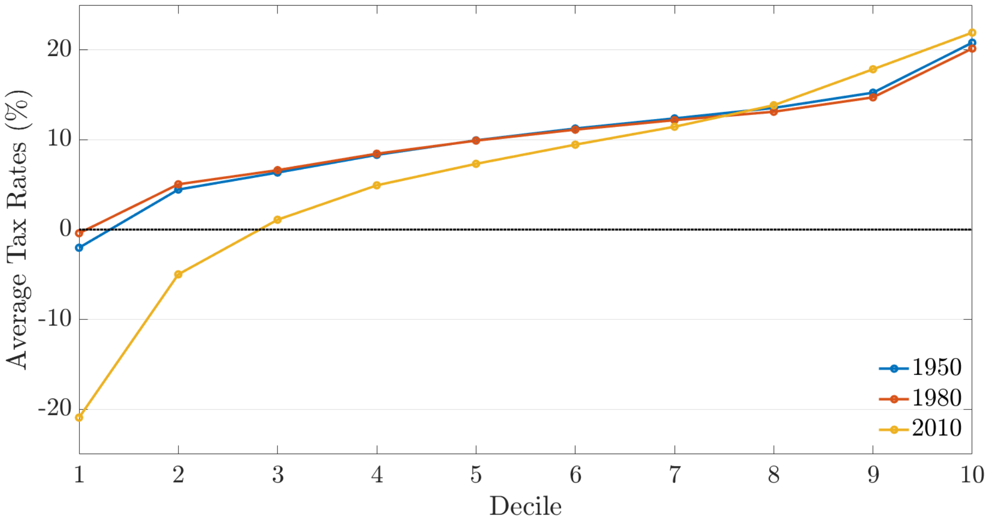

Figure 4: Optimal Average Tax Rates

Notes: This figure shows optimal average rates given by the entire tax-and-transfer system by income decile for the U.S. in 1950, 1980, and 2010.

This isolates the non-homotheticity effect. This uniform price change has a sizeable effect on optimal redistribution. Because the economy is much richer after this price change consumption shifts towards luxuries and the planner's desire to redistribute declines. Optimal average tax rates are only mildly progressive with much higher tax rates in the bottom income deciles.

The third step of the decomposition is to also change relative prices. Because agricultural goods get cheaper over time and services get more expensive this is another force for less desired redistribution. The goods that are disproportionately consumed by the poor are getting cheaper compared to the previous step, whereas the rich face higher prices of their consumption baskets. Hence, the planner does not have to implement as much redistribution.

Next, we adjust the skill premium to its 2010 level. This implies a higher level of inequality because the skill premium rises from 1.44 to 2.06. As expected, this increases the progressivity of the overall tax-and-transfer system. Tax rates for the bottom deciles fall, while tax rates at the top rise.

The effect of changing the skill premium is of similar size as the effect of adjusting population shares from the 1950 to the 2010 levels, which is the last step of the decomposition. This also significantly increases the optimal net transfers received by the poor and increases tax rates at the top. A larger share of high skilled individuals raises income inequality and the number of individuals from which high taxes can be raised, explaining the large changes to the tax system.

With this last step we are almost back at the 2010 optimal tax-and-transfer system,

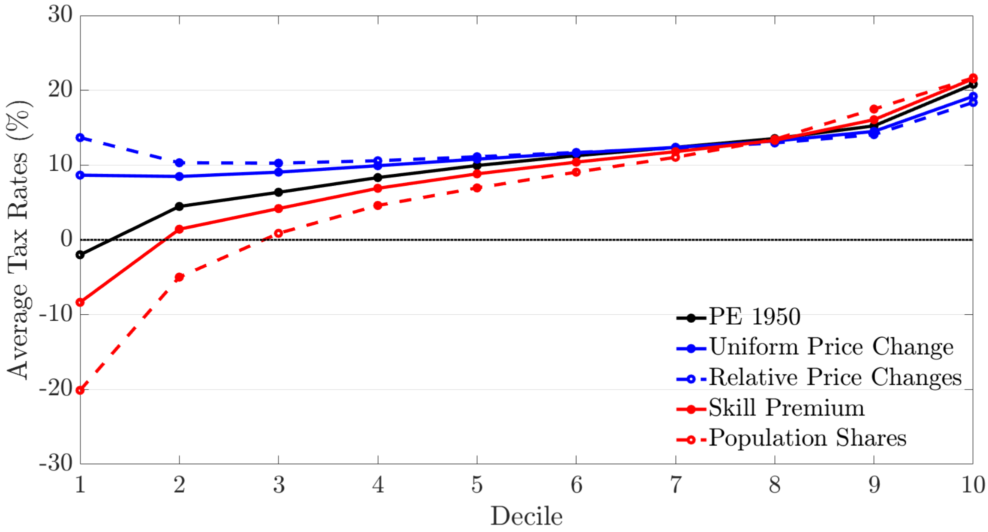

Figure 5: Optimal Average Tax Rates: Decomposition

Notes: This figure decomposes the change in optimal average tax rates between 1950 and 2010.

even though we are ignoring general equilibrium effects of taxes in this decomposition. This suggests that general equilibrium effects of tax changes are relatively weak in this model. This is the case because there is a lot of overlap in the income distributions of low and high skilled individuals. Therefore, changing the tax systems does not induce large changes in the relative labor supply of low skilled versus high skilled, so that the skill premium moves little. It requires much larger changes to the tax system to observe significant changes to the skill premium.

4.7 Calibration for China

We now turn to an alternative calibration of the model to the Chinese economy. This is of interest because we have seen that the role of non-homotheticities is larger at earlier stages of development. The current calibration is relatively rough because some of the required data is less easily available for China than for the United States. However, the calibration captures key features of structural change and the dynamics of inequality in China. Also because of limited data availability we restrict ourselves to the years 1989, 1999, 2009.

Table 5 shows the data moments we are targeting for China. For the aggregate growth part of the data moments, we use data from the World Development Indicators for the sectoral contributions to total value added. The service sector expanded from 33% in 1989 to 44% in 2009. Over the same time period also the goods sector grew slightly from 43% to 46%. Correspondingly, the agricultural sector shrank from 25% to 10%. This shows the potential for an even larger effect of non-homotheticities, as households spend

Table 5: Moments for the China Calibration

Moment	1989	1999	2009
Skill premium	1.00	1.10	1.58
High-skill inc. share agriculture	2.00%	4.00%	8.00%
High-skill inc. share manufacturing	2.00%	6.00%	14.00%
High-skill inc. share services	11.00%	18.00%	34.00%
Service share	32.89%	38.57%	44.41%
Manufacturing share	42.50%	45.36%	45.96%
Real GDP per capita	0.41	1.00	2.45
Rel. price agriculture to manufacturing	1.10	1.00	0.90
Rel. price services to manufacturing	0.90	1.00	1.10
G/Y	13.00%	13.00%	13.00%
T/Y	1.00%	1.00%	1.00%
AMTR Top 10%	14.00%	14.00%	14.00%
AMTR Bottom 90%	14.00%	14.00%	14.00%

Notes: This table summarizes the moments for the China calibration. Real GDP per capita and relative prices are normalized to 1 in 1999.

a large share of their incomes on necessities initially.

Relatedly, even though we are considering a shorter time period than for the U.S., aggregate growth was much larger for China. We take data for the growth rate of real GDP per capita from Fang and Herrendorf (2021). GDP per capita rose by a factor of six. We also use data from this paper to infer targets for the high-skilled income shares by sector, which determines the skill intensities in the model. Because their sector definitions are different from ours, we cannot simply take their data. Hence, the targets reported in Table 5 are approximations, but should be roughly in line with the true values. As for the U.S., services are the high-skill intensive sector and skill intensities are going up over time. We also use the evidence from Fang and Herrendorf (2021) to discipline the skill premium. In 1989, there was no skill premium, but it increased to 1.58 in 2009, with the larger part of this increase happening in the second half of the sample.

Finally, when calibrating the government parameters we have to take a stand on the size of the welfare state, total government spending, and the progressivity of taxes. While China has a fairly progressive income tax that has become more important over time (Li and Ma 2017; Piketty and Qian 2009), it raises the majority of its tax revenue through corporate taxes and value added taxes. As value added taxes are generally considered to be less progressive than income tax, we decide to target a flat tax rate in all periods. The total amount of revenue raised is roughly constant over time relative to GDP, so we also keep this constant at 14% over time. The Chinese welfare state has expanded, but a large share of this goes to the pension system, which we do not model. Hence, we calibrate a constant split between government consumption and transfers, with G/Y of

13% and T/Y of 1%.

Our calibration strategy to match these targets closely follows what we have done for the U.S. However, we choose slightly different parameters for the non-homothetic CES utility function, within the range of estimates by Comin, Lashkari, and Mestieri (2021). We could alternatively keep them at their U.S. levels, but adjusting them slightly helps improving the match for all the sector shares. Also, we choose the evolution of sectoral relative prices such that we match the sector shares in all periods well. As for the U.S., we match the middle year sector shares exactly, but here we also target the sector shares in the other years explicitly.

4.8 Optimal Tax-and-Transfer System for China

The optimal tax-and-transfer system for China is shown in Figure 6. As for the U.S., we follow an inverse optimum approach. We pick as our base year 1999, so the optimal tax-and-transfer system for the year 1999 corresponds to our calibration. It is mildly progressive in overall average tax rates of the entire tax-and-transfer system because there is a small lump sum transfer combined with a flat tax.

Moving back in time to 1989 is associated with a slightly lower skill premium. Also, to match the skill premium at calibrated skill intensities our model requires the share of highly skilled individuals to go up over time (in line with the data). For these reasons, standard tax theory would predict that less redistribution is desirable in 1989. However, there is also a large increase in average income and a significant shift away from the consumption of necessities from 1989 to 1999. This latter effect dominates. Instead of paying net taxes as in the 1999 (inverse) optimum, the lowest income decile would receive a net transfer. Average tax rates are also lower for the next income deciles and higher at the top.

The optimal tax-and-transfer-system is also more redistributive in 2009 compared to 1999. The economy is again much richer on average and particularly the bottom income groups have moved further away from absolute poverty. They spend a smaller share of their income on agricultural goods. However, there is also a larger share of high-skilled who earn a much higher skill premium. Thus, the planner prefers to impose a higher tax rate at the top. This allows to give a larger transfer providing more redistribution towards the bottom of the distribution. When comparing the optimal 1989 and 2009 systems, the 2009 system is only slightly more redistributive even though inequality is much larger. The effect of individuals growing out of absolute poverty with aggregate growth and non-homothetic preferences almost cancels out the rise in inequality.

As our benchmark, we again adjusted Pareto weights for the other years as we did in the U.S. case, keeping the ratio of the Pareto weight on the highest earning group to that of the lowest earning group constant. In Figure 8 in the appendix we repeat the exercise

Table 6: Parameters for the China Calibration

Parameter	Interpretation	Value
Preferences		
σ	Elasticity of substitution	0.500
ε_A	Income elasticity agriculture	0.100
ε_G	Income elasticity goods	1.000
ε_S	Income elasticity services	1.300
Ω_A	Level parameter agriculture	0.081
Ω_G	Level parameter goods	1.000
Ω_S	Level parameter services	1.365
γ	Coefficient of risk aversion	2.000
φ	Labor supply elasticity	5.000
B	Labor disutility	2504.842
Production		
ρ	Elasticity of substitution	1.420
A_A	Neutral technology agriculture	0.227, 0.744, 2.514
A_G	Neutral technology goods	0.250, 0.794, 2.614
A_S	Neutral technology services	0.359, 1.000, 3.092
α_A	Skill intensity agriculture	0.061, 0.099, 0.170
α_G	Skill intensity goods	0.061, 0.129, 0.242
α_S	Skill intensity services	0.187, 0.261, 0.418
f_H	Population share high skilled	0.048, 0.095, 0.160
Government		
G_A	Government consumption agriculture	0.002, 0.004, 0.007
G_G	Government consumption goods	0.042, 0.011, 0.027
G_S	Government consumption services	0.033, 0.009, 0.025
T	Transfer	0.004, 0.003, 0.003
λ	Tax function level	0.140, 0.140, 0.140
τ	Tax function progressivity	0.000, 0.000, 0.000

Notes: This table summarizes the calibrated parameters for China. For the time-varying parameters the values correspond to the calibration for the years 1989, 1999, and 2009.

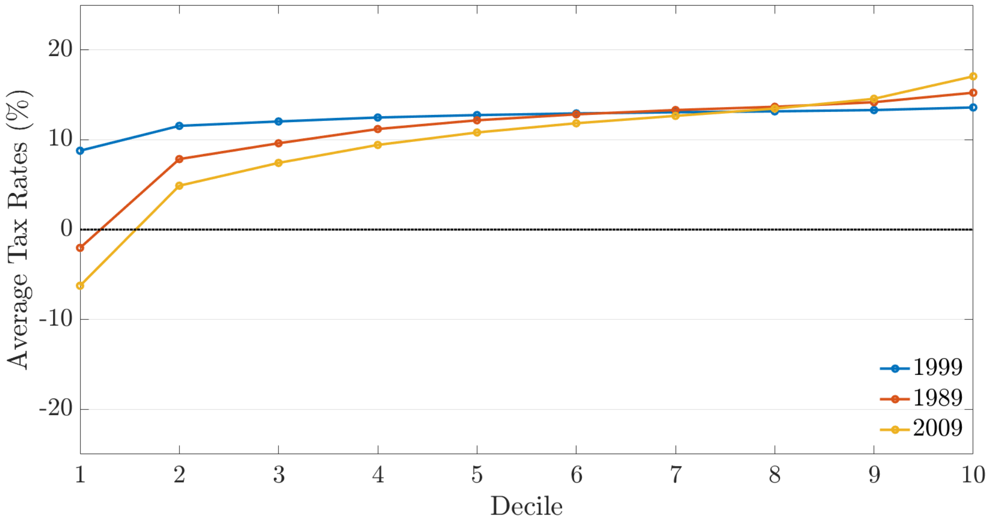

Figure 6: Optimal Average Tax Rates

Notes: This figure shows optimal average rates given by the entire tax-and-transfer system by income decile for China in 1989, 1999, and 2009.

for Pareto weights that are simply kept constant for groups characterized by a skill level and within-skill productivity. This makes a larger difference for the results than in the U.S. case. However, the main message is unaffected: Aggregate growth in combination with non-homothetic preferences lowers the benefits of redistribution. Quantitatively, the effect is even stronger, with the welfare state optimally becoming less generous from 1989 to 2009.

5 Conclusion

In this paper we incorporate the optimal income taxation problem into a state-of-the-art model of structural change. This allows us to show that aggregate growth matters for optimal redistribution, beyond standard effects through relative inequality. With non-homothetic preferences, the taste for redistribution declines with aggregate growth as households move away from absolute poverty and shift their consumption away from necessities towards luxuries. However, aggregate growth and structural change have additional effects on the optimal design of the welfare state through changing prices of different commodities and changing inequality.

In our calibrated economies, growth has non-monotonic effects on the optimal size of the welfare state. At early stages of development the effect of non-homotheticicities is dominant, as many people have to spend a large share of their incomes on necessities. At later stages of development this effect weakens and relative inequality concerns become dominant.

These patterns naturally raise the question whether using public debt to finance the welfare state at early stages of development could be beneficial. We explore this possibility in ongoing work.

References

Baumol, William J. (1967). "Macroeconomics of unbalanced growth: the anatomy of urban crisis." *The American Economic Review* 57.3, pp. 415–426.

Bhandari, Anmol, David Evans, Mikhail Golosov, and Thomas J. Sargent (2017). "Public debt in economies with heterogeneous agents." *Journal of Monetary Economics* 91, pp. 39–51.

Bick, Alexander, Nicola Fuchs-Schündeln, and David Lagakos (2018). "How do hours worked vary with income? Cross-country evidence and implications." *The American Economic Review* 108.1, pp. 170–99.

Boppart, Timo (2014). "Structural change and the Kaldor facts in a growth model with relative price effects and non-Gorman preferences." *Econometrica* 82.6, pp. 2167–2196.

Boppart, Timo and Per Krusell (2020). "Labor supply in the past, present, and future: a balanced-growth perspective." *Journal of Political Economy* 128.1, pp. 118–157.

Bourguignon, François and Amedeo Spadaro (2012). "Tax–benefit revealed social preferences." *The Journal of Economic Inequality* 10.1, pp. 75–108.

Buera, Francisco J. and Joseph P. Kaboski (2012). "The rise of the service economy." *The American Economic Review* 102.6, pp. 2540–69.

Buera, Francisco J., Joseph P. Kaboski, Richard Rogerson, and Juan I. Vizcaino (2018). "Skill-Biased Structural Change." *Working Paper.*

— (2021). "Skill-biased structural change." *Forthcoming in the Review of Economic Studies.*

Christiansen, V. (1977). "The theoretical basis for deriving distributive weights to be used in costbenefit analysis." *Memorandum from the Institute of Economics, University of Oslo* 25.

Comin, Diego, Danial Lashkari, and Martí Mestieri (2017). "Structural change with long-run income and price effects." *Working Paper.*

— (2021). "Structural change with long-run income and price effects." *Econometrica* 89.1, pp. 311–374.

Diamond, Peter A. (1998). "Optimal income taxation: an example with a U-shaped pattern of optimal marginal tax rates." *The American Economic Review*, pp. 83–95.

Duarte, Margarida and Diego Restuccia (2010). "The role of the structural transformation in aggregate productivity." *The Quarterly Journal of Economics* 125.1, pp. 129–173.

— (2017). "Relative prices and sectoral productivity." *Journal of the European Economic Association.*

Fang, Lei and Berthold Herrendorf (2021). "High-skilled services and development in China." *Journal of Development Economics* 151, p. 102671.

Farhi, Emmanuel and Iván Werning (2013). "Insurance and taxation over the life cycle." *The Review of Economic Studies* 80.2, pp. 596–635.

Ferreira, Francisco H. G. and Martin Ravallion (2011). "Poverty and Inequality: The Global Context." *The Oxford Handbook of Economic Inequality.* Ed. by Brian Nolan, Wiemer Salverda, and Timothy M. Smeeding. Oxford University Press.

Ferriere, Axelle, Philipp Grübener, Gaston Navarro, and Oliko Vardishvili (2021). "Larger Transfers Financed with More Progressive Taxes? On the Optimal Design of Taxes and Transfers." *Working Paper.*

Ferriere, Axelle and Gaston Navarro (2020). "The Heterogeneous Effects of Government Spending: It's All About Taxes." *Working Paper.*

Geary, Roy C. (1950). "A note on "A constant-utility index of the cost of living"." *The Review of Economic Studies* 18.1, pp. 65–66.

Golosov, Mikhail, Maxim Troshkin, and Aleh Tsyvinski (2016). "Redistribution and social insurance." *The American Economic Review* 106.2, pp. 359–86.

Hanoch, Giora (1975). "Production and demand models with direct or indirect implicit additivity." *Econometrica* 43.3, pp. 395–419.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante (2017). "Optimal tax progressivity: An analytical framework." *The Quarterly Journal of Economics* 132.4, pp. 1693–1754.

— (2020). "Presidential Address 2019: How Should Tax Progressivity Respond to Rising Income Inequality?" *Journal of the European Economic Association* 18.6, pp. 2715–2754.

Heathcote, Jonathan and Hitoshi Tsujiyama (2021). "Optimal income taxation: Mirrlees meets Ramsey." *Journal of Political Economy* 129.11, pp. 3141–3184.

Hendren, Nathaniel (2020). "Measuring economic efficiency using inverse-optimum weights." *Journal of Public Economics* 187, p. 104198.

Herrendorf, Berthold, Richard Rogerson, and Akos Valentinyi (2014). "Growth and structural transformation." *Handbook of Economic Growth.* Vol. 2. Elsevier, pp. 855–941.

Jones, Charles I. (2016). "The facts of economic growth." *Handbook of Macroeconomics*. Vol. 2. Elsevier, pp. 3–69.

Katz, Lawrence F. and Kevin M. Murphy (1992). "Changes in relative wages, 1963–1987: supply and demand factors." *The Quarterly Journal of Economics* 107.1, pp. 35–78.

Kongsamut, Piyabha, Sergio Rebelo, and Danyang Xie (2001). "Beyond balanced growth." *The Review of Economic Studies* 68.4, pp. 869–882.

Kuznets, Simon (1973). "Modern economic growth: findings and reflections." *The American Economic Review* 63.3, pp. 247–258.

Li, Rong and Guangrong Ma (2017). "Personal-Income-Tax Reforms and Effective-Tax Functions in China." *Finanz-Archiv: Zeitschrift für das Gesamte Finanzwesen* 73.3, p. 317.

Lockwood, Benjamin B. and Matthew Weinzierl (2016). "Positive and normative judgments implicit in US tax policy, and the costs of unequal growth and recessions." *Journal of Monetary Economics* 77, pp. 30–47.

Mertens, Karel and José Luis Montiel Olea (2018). "Marginal tax rates and income: New time series evidence." *The Quarterly Journal of Economics* 133.4, pp. 1803–1884.

Mirrlees, James A. (1971). "An exploration in the theory of optimum income taxation." *The Review of Economic Studies* 38.2, pp. 175–208.

Ngai, L. Rachel and Christopher A. Pissarides (2007). "Structural change in a multisector model of growth." *The American Economic Review* 97.1, pp. 429–443.

Piketty, Thomas and Nancy Qian (2009). "Income inequality and progressive income taxation in China and India, 1986-2015." *American Economic Journal: Applied Economics* 1.2, pp. 53–63.

Ramsey, Frank P. (1927). "A Contribution to the Theory of Taxation." *The Economic Journal* 37.145, pp. 47–61.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek (2020). "IPUMS USA: Version 10.0 [dataset]. Minneapolis, MN: IPUMS, 2020." *Dataset*.

Sachs, Dominik, Aleh Tsyvinski, and Nicolas Werquin (2020). "Nonlinear tax incidence and optimal taxation in general equilibrium." *Econometrica* 88.2, pp. 469–493.

Saez, Emmanuel (2001). "Using elasticities to derive optimal income tax rates." *The Review of Economic Studies* 68.1, pp. 205–229.

Song, Zheng, Kjetil Storesletten, Yikai Wang, and Fabrizio Zilibotti (2015). "Sharing high growth across generations: pensions and demographic transition in China." *American Economic Journal: Macroeconomics* 7.2, pp. 1–39.

Święcki, Tomasz (2017). "Determinants of structural change." *Review of Economic Dynamics* 24, pp. 95–131.

Werning, Ivan (2007). "Optimal fiscal policy with redistribution." *The Quarterly Journal of Economics* 122.3, pp. 925–967.

A Partial Equilibrium Model Appendix

A.1 Non-Homothetic CES

We specify preferences over the three commodities using a non-homothetic CES utility function as introduced by Hanoch (1975) and recently popularized in the structural change literature by Comin, Lashkari, and Mestieri (2021). These preferences do not admit expressing Marshallian demand functions in closed form, so we use the Hicksian demand function (5) and the expenditure function (4). These can be derived from the expenditure minimization problem of a household subject to constraint (3)

$$\min \sum_{j=1}^{J} p_j C_j - \chi \left[\sum_{j=1}^{J} \left(\Omega_j C^{\varepsilon_j}\right)^{\frac{1}{\sigma}} C_j^{\frac{\sigma-1}{\sigma}} - 1 \right], \tag{22}$$

where χ is the Lagrange multiplier. The first order conditions with respect to C_j read

$$p_j - \chi \left(\Omega_j C^{\varepsilon_j}\right)^{\frac{1}{\sigma}} \frac{\sigma - 1}{\sigma} C_j^{-\frac{1}{\sigma}} \quad \forall\, j. \tag{23}$$

Dividing the first order conditions for two commodities we obtain

$$\frac{p_j}{p_i} = \frac{\left(\Omega_j C^{\varepsilon_j}\right)^{\frac{1}{\sigma}}}{\left(\Omega_i C^{\varepsilon_i}\right)^{\frac{1}{\sigma}}} \frac{C_j}{C_i}. \tag{24}$$

Using this we can express any commodity j as a function of commodity i:

$$C_j = \left(\frac{p_i}{p_j}\right)^{\sigma} \frac{\Omega_j C^{\varepsilon_j}}{\Omega_i C^{\varepsilon_i}} C_i. \tag{25}$$

We can plug this back into the constraint (3) for every good j such that we are left with one equation for a generic commodity i. Solving this equation yields the Hicksian demand function. Multiplying the expression with the respective price and summing over all commodities yields the expenditure function.

Below we are going to need the first and second derivatives of the expenditure function, which are given by

$$E'(C) = \frac{1}{1-\sigma}\left[\sum_{j=1}^{I}\Omega_j\varepsilon_j C^{\varepsilon_j-1}p_j^{1-\sigma}\right]E(C)^{\sigma}, \tag{26}$$

$$E''(C) = \frac{\sigma}{1-\sigma}\left[\sum_{j=1}^{I}\Omega_j\varepsilon_j C^{\varepsilon_j-1}p_j^{1-\sigma}\right]E(C)^{\sigma-1}E'(C) + \frac{E(C)^{\sigma}}{1-\sigma}\left[\sum_{j=1}^{J}\Omega_j\varepsilon_j(\varepsilon_j-1)C^{\varepsilon_j-2}p_j^{1-\sigma}\right]. \tag{27}$$

A.2 Derivation of Optimal Tax Formula

We derive the optimal tax formula using standard perturbation techniques following Saez (2001). The idea is that at the optimal tax schedule a small perturbation of the tax system may not affect welfare. To derive the optimal tax formula, we therefore consider a small increase of the marginal tax rate, $d\mathcal{T}'$ in the interval $[y^* - dy, y^*]$.

The first effect of this perturbation of the tax schedule is that individuals earning incomes larger than y^* pay $d\mathcal{T}'dy$ more taxes. This reduces their welfare, which must be weighted with their endogenous marginal social welfare weight, but gives the government additional funds. Hence, the mechanical effect can be written as

$$dW^M(\theta^*) = d\mathcal{T}'dy\int_{\theta^*}^{\bar{\theta}}\left(1 - \frac{\frac{u'(\theta)}{E'(\theta)}w(\theta)}{\lambda}\right)dF(\theta). \tag{28}$$

The second effect is the substitution effect. Individuals whose income falls in the interval where the marginal tax rate changes adjust their labor supply by

$$\frac{\partial y(\theta^*)}{\partial\mathcal{T}'}d\mathcal{T}' = -\varepsilon_{y,1-\mathcal{T}'}(\theta^*)\frac{y(\theta^*)}{1-\mathcal{T}'(y(\theta^*))}d\mathcal{T}'. \tag{29}$$

This adjustment has to be weighted with the mass of individuals who are affected:

$$h(y(\theta^*))dy = f(\theta^*)\frac{1}{\varepsilon_{y,\theta}(\theta^*)}\frac{\theta^*}{y(\theta^*)}dy. \tag{30}$$

The impact of this change on individuals' welfare is of second order by the envelope theorem. However, the labor supply change has a first order impact on the government budget. Hence, the substitution effect can be written as follows:

$$dW^S(\theta^*) = -\mathcal{T}'(y(\theta^*))\varepsilon_{y,1-\mathcal{T}'}(\theta^*)\frac{y(\theta^*)}{1-\mathcal{T}'(y(\theta^*))}d\mathcal{T}'f(\theta^*)\frac{1}{\varepsilon_{y,\theta}(\theta^*)}\frac{\theta^*}{y(\theta^*)}dy. \tag{31}$$

The third effect is an income effect. Given that we use preferences with income effects, labor supply of those earnings more than $y(\theta^*)$ adjust their labor supply even though their marginal tax rates do not change. We denote the income effect as $\eta(\theta) = \frac{\partial y(\theta)}{\partial T}$. The income effect can be written as

$$dW^I(\theta^*) = d\mathcal{T}' dy \int_{\theta^*}^{\bar{\theta}} \eta(\theta)\, \mathcal{T}'(y(\theta))\, dF(\theta). \tag{32}$$

At the optimum it has to be the case that

$$dW^M + dW^S + dW^I = 0. \tag{33}$$

Solving this equation yields the optimal tax formula

$$\frac{\mathcal{T}'(y(\theta^*))}{1-\mathcal{T}'(y(\theta^*))} = \frac{\varepsilon_{y,\theta}(\theta^*)}{\varepsilon_{y,1-\mathcal{T}'}(\theta^*)} \frac{\int_{\theta^*}^{\bar{\theta}} \left[1 - \frac{\frac{u'(\theta)}{E'(\theta)} w(\theta)}{\lambda} + \eta(\theta)\, \mathcal{T}'(y(\theta))\right] dF(\theta)}{f(\theta^*)\, \theta^*}. \tag{34}$$

From the first order condition for the optimal choice of the lump sum element we can obtain an expression for λ:

$$\int_{\underline{\theta}}^{\bar{\theta}} \left(1 - \frac{\frac{u'(\theta)}{E'(\theta)} w(\theta)}{\lambda} + \eta(\theta)\mathcal{T}'(y(\theta))\right) dF(\theta) = 0 \leftrightarrow \lambda = \frac{\int_{\underline{\theta}}^{\bar{\theta}} \frac{u'(\theta)}{E'(\theta)} w((\theta)) dF(\theta)}{\int_{\underline{\theta}}^{\bar{\theta}} \left(1 + \eta(\theta)\mathcal{T}'(y(\theta))\right) dF(\theta)}. \tag{35}$$

To apply the formula, we still need expressions for $\varepsilon_{y,\theta}(\theta)$, $\varepsilon_{y,1-\mathcal{T}'}(\theta)$, and $\eta(\theta)$. To derive these, consider the household problem

$$\max_{C,y}\; u(C) - v\left(\frac{y}{\theta}\right) \;\; s.t. \; E(C) = y - \mathcal{T}(y) + T. \tag{36}$$

The first order condition of this problem reads as

$$v'\left(\frac{y}{\theta}\right) - \frac{u'(C)}{E'(C)}(1 - \mathcal{T}'(y))\theta = 0. \tag{37}$$

Applying the implicit function theorem to the first order condition, we obtain

$$\frac{\partial y}{\partial(1-\mathcal{T}')} = -\frac{-\frac{\theta u'(C)}{E'(C)} + \frac{(E''(C)u'(C) - u''(C)E'(C))(1-\mathcal{T}'))\theta}{(E'(C))^2} \frac{\partial C}{\partial(1-\mathcal{T}')}}{v''\left(\frac{y}{\theta}\right)\frac{1}{\theta} + \mathcal{T}''(y)\,\theta \frac{u'(C)}{E'(C)} + \frac{(E''(C)u'(C) - u''(C)E'(C))(1-\mathcal{T}'))\theta}{(E'(C))^2} \frac{\partial C}{\partial y}}. \tag{38}$$

From the budget constraint we get

$$\frac{\partial C}{\partial(1-\mathcal{T}')} = 0 \tag{39}$$

$$\frac{\partial C}{\partial y} = \frac{1-\mathcal{T}'(y)}{E'(C)}. \tag{40}$$

Hence, it follows for the elasticity that

$$\varepsilon_{y,1-\mathcal{T}'}(\theta) = \frac{\frac{\theta u'(C)}{E'(C)}(1-\mathcal{T}'(y))}{\left[v''\left(\frac{y}{\theta}\right)\frac{1}{\theta} + \mathcal{T}''(y)\,\theta\frac{u'(C)}{E'(C)} + \frac{(E''(C)u'(C)-u''(C)E'(C))(1-\mathcal{T}')\theta}{(E'(C))^2}\frac{1-\mathcal{T}'(y)}{E'(C)}\right]y} \tag{41}$$

Similarly, we can compute

$$\frac{\partial y}{\partial \theta} = -\frac{-v''\left(\frac{y}{\theta}\right)\frac{y}{\theta^2} - \frac{u'(C)}{E'(C)}(1-\mathcal{T}') + \frac{(E''(C)u'(C)-u''(C)E'(C))(1-\mathcal{T}')\theta}{(E'(C))^2}\frac{\partial C}{\partial \theta}}{v''\left(\frac{y}{\theta}\right)\frac{1}{\theta} + \mathcal{T}''(y)\,\theta\frac{u'(C)}{E'(C)} + \frac{(E''(C)u'(C)-u''(C)E'(C))(1-\mathcal{T}')\theta}{(E'(C))^2}\frac{\partial C}{\partial y}} \tag{42}$$

From the budget constraint we get that

$$\frac{\partial C}{\partial \theta} = 0, \tag{43}$$

so that we obtain

$$\varepsilon_{y,\theta} = \frac{v''\left(\frac{y}{\theta}\right)\frac{y}{\theta} + \theta\frac{u'(C)}{E'(C)}(1-\mathcal{T}')}{\left[v''\left(\frac{y}{\theta}\right)\frac{1}{\theta} + \mathcal{T}''(y)\,\theta\frac{u'(C)}{E'(C)} + \frac{(E''(C)u'(C)-u''(C)E'(C))(1-\mathcal{T}')\theta}{(E'(C))^2}\frac{1-\mathcal{T}'(y)}{E'(C)}\right]y}. \tag{44}$$

Combining these results, we obtain

$$\begin{aligned}\frac{\varepsilon_{y,\theta}(\theta)}{\varepsilon_{y,1-\mathcal{T}'}(\theta)} &= \frac{v''\left(\frac{y}{\theta}\right)\frac{y}{\theta} + \theta\frac{u'(C)}{E'(C)}(1-\mathcal{T}')}{\frac{\theta u'(C)}{E'(C)}(1-\mathcal{T}'(y))} \\ &= \frac{v''\left(\frac{y}{\theta}\right)\frac{y}{\theta} + v'\left(\frac{y}{\theta}\right)}{v'\left(\frac{y}{\theta}\right)} \\ &= 1 + \frac{1}{\varepsilon}\end{aligned} \tag{45}$$

Finally, we can compute the income effect as

$$\eta(\theta) = -\frac{\frac{(E''(C)u'(C)-u''(C)E'(C))(1-\mathcal{T}')\theta}{(E'(C))^2}\frac{\partial C}{\partial T}}{v''\left(\frac{y}{\theta}\right)\frac{1}{\theta} + \mathcal{T}''(y)\,\theta\frac{u'(C)}{E'(C)} + \frac{(E''(C)u'(C)-u''(C)E'(C))(1-\mathcal{T}')\theta}{(E'(C))^2}\frac{\partial C}{\partial y}} \tag{46}$$

with

$$\frac{\partial C}{\partial b} = \frac{1}{E'(C)}. \tag{47}$$

Rearranging this equation yields

$$\eta\left(\theta\right)=-\frac{\frac{\left(E''(C)u'(C)-u''(C)E'(C)\right)\theta}{\left(E'(C)\right)^2}}{\frac{u'(C)}{\varepsilon\frac{y}{\theta}}+\frac{\mathcal{T}''(y)}{1-\mathcal{T}'(y)}\theta u'(C)+\frac{\left(E''(C)u'(C)-u''(C)E'(C)\right)(1-\mathcal{T}')\theta}{\left(E'(C)\right)^2}}. \tag{48}$$

B General Equilibrium Model Appendix

B.1 Firm Problem

In this section we derive equation (15) from the firm problem (14). The first order conditions of the problem are

$$p_jA_j\alpha_jH_j^{\frac{\rho-1}{\rho}-1}\left[\alpha_jH_j^{\frac{\rho-1}{\rho}}+\left(1-\alpha_j\right)L_j^{\frac{\rho-1}{\rho}}\right]^{\frac{\rho}{\rho-1}-1}=w_H, \tag{49}$$

$$p_jA_j\left(1-\alpha_j\right)L_j^{\frac{\rho-1}{\rho}-1}\left[\alpha_jH_j^{\frac{\rho-1}{\rho}}+\left(1-\alpha_j\right)L_j^{\frac{\rho-1}{\rho}}\right]^{\frac{\rho}{\rho-1}-1}=1. \tag{50}$$

Dividing equation (49) by equation (50) yields

$$w_H=\frac{\alpha_j}{1-\alpha_j}\left(\frac{H_j}{L_j}\right)^{-\frac{1}{\rho}}.$$

This can be solved for the demand for low-skilled labor as a function of the skill premium and high-skilled labor demand:

$$L_j=H_jw_H^{\rho}\left(\frac{1-\alpha_j}{\alpha_j}\right)^{\rho}. \tag{51}$$

We can plug (51) into (49) to obtain

$$
\begin{aligned}
w_H &= p_j A_j \alpha_j H_j^{\frac{\rho-1}{\rho}-1} \left[\alpha_j H_j^{\frac{\rho-1}{\rho}} + (1-\alpha_j) \left(H_j w_H^{\rho} \left(\frac{1-\alpha_j}{\alpha_j} \right)^{\rho} \right)^{\frac{\rho-1}{\rho}} \right]^{\frac{\rho}{\rho-1}-1} \\
\Rightarrow w_H &= p_j A_j \alpha_j H_j^{\frac{-1}{\rho}} \left[H_j^{\frac{\rho-1}{\rho}} \left[\alpha_j + (1-\alpha_j) w_H^{\rho-1} \left(\frac{1-\alpha_j}{\alpha_j} \right)^{\rho-1} \right] \right]^{\frac{1}{\rho-1}} \\
\Rightarrow w_H &= p_j A_j \alpha_j H_j^{\frac{-1}{\rho}} H_j^{\frac{1}{\rho}} \left[\alpha_j + (1-\alpha_j) w_H^{\rho-1} \left(\frac{1-\alpha_j}{\alpha_j} \right)^{\rho-1} \right]^{\frac{1}{\rho-1}} \\
\Rightarrow w_H^{\rho-1} &= p_j^{\rho-1} A_j^{\rho-1} \alpha_j^{\rho-1} \left[\alpha_j + (1-\alpha_j) w_H^{\rho-1} \left(\frac{1-\alpha_j}{\alpha_j} \right)^{\rho-1} \right] \\
\Rightarrow w_H^{\rho-1} &= p_j^{\rho-1} A_j^{\rho-1} \alpha_j^{\rho} + p_j^{\rho-1} A_j^{\rho-1} \alpha_j^{\rho-1} (1-\alpha_j) w_H^{\rho-1} \left(\frac{1-\alpha_j}{\alpha_j} \right)^{\rho-1} \\
\Rightarrow w_H^{\rho-1} &= p_j^{\rho-1} A_j^{\rho-1} \alpha_j^{\rho} + p_j^{\rho-1} A_j^{\rho-1} (1-\alpha_j)^{\rho} w_H^{\rho-1} \\
\Rightarrow 1 &= p_j^{\rho-1} \frac{A_j^{\rho-1}}{w_H^{\rho-1}} \alpha_j^{\rho} + p_j^{\rho-1} A_j^{\rho-1} (1-\alpha_j)^{\rho} \\
\Rightarrow 1 &= p_j^{\rho-1} A_j^{\rho-1} \left[\frac{\alpha_j^{\rho}}{w_H^{\rho-1}} + (1-\alpha_j)^{\rho} \right] \\
\Rightarrow 1 &= p_j A_j \left[\frac{\alpha_j^{\rho}}{w_H^{\rho-1}} + (1-\alpha_j)^{\rho} \right]^{\frac{1}{\rho-1}} .
\end{aligned}
$$

Solving this equation for p_j delivers equation (15) in the main text.

B.2 Equilibrium

To solve for an equilibrium in the market for high-skilled labor, we first compute demand for high-skilled labor. For that purpose, we use equations (13), (51), and (15). First, plug (51) into (13):

$$
\begin{aligned}
Y_j &= A_j \left[\alpha_j H_j^{\frac{\rho-1}{\rho}} + (1-\alpha_j) \left(H_j w_H^{\rho} \left(\frac{1-\alpha_j}{\alpha_j} \right)^{\rho} \right)^{\frac{\rho-1}{\rho}} \right]^{\frac{\rho}{\rho-1}} \\
\Rightarrow Y_j &= A_j \left[\alpha_j H_j^{\frac{\rho-1}{\rho}} + (1-\alpha_j) H_j^{\frac{\rho-1}{\rho}} \left(\frac{\alpha_j}{1-\alpha_j} \frac{1}{w_H} \right)^{1-\rho} \right]^{\frac{\rho}{\rho-1}} \\
\Rightarrow Y_j &= A_j H_j \left[\alpha_j + (1-\alpha_j) \left(\frac{\alpha_j}{1-\alpha_j} \frac{1}{w_H} \right)^{1-\rho} \right]^{\frac{\rho}{\rho-1}} \\
\Rightarrow Y_j &= A_j H_j \left[\alpha_j + (1-\alpha_j)^{\rho} \alpha_j^{1-\rho} w_H^{\rho-1} \right]^{\frac{\rho}{\rho-1}} \\
\Rightarrow \left(\frac{Y_j}{A_j H_j} \right)^{\frac{\rho-1}{\rho}} &= \alpha_j + (1-\alpha_j)^{\rho} \alpha_j^{1-\rho} w_H^{\rho-1} \\
\Rightarrow \frac{\alpha_j^{\rho}}{w_H^{\rho-1}} + (1-\alpha_j)^{\rho} &= \left(\frac{Y_j}{A_j H_j} \right)^{\frac{\rho-1}{\rho}} \frac{1}{\alpha_j^{1-\rho}} \frac{1}{w_H^{\rho-1}}.
\end{aligned}
$$

Next, we can rewrite (15) to obtain

$$
\frac{\alpha_j^{\rho}}{w_H^{\rho-1}} + (1-\alpha_j)^{\rho} = (p_j A_j)^{1-\rho}.
$$

Combining the last two equations yields

$$
\begin{aligned}
\left(\frac{Y_j}{A_j H_j} \right)^{\frac{\rho-1}{\rho}} \frac{1}{\alpha_j^{1-\rho}} \frac{1}{w_H^{\rho-1}} &= (p_j A_j)^{1-\rho} \\
\Rightarrow \left(\frac{Y_j}{A_j H_j} \right)^{\frac{\rho-1}{\rho}} &= \left(\frac{\alpha_j p_j A_j}{w_H} \right)^{1-\rho} \\
\frac{Y_j}{A_j H_j} &= \left(\frac{\alpha_j p_j A_j}{w_H} \right)^{-\rho}.
\end{aligned}
$$

Rearranging this gives equation (18).

B.3 Computation

We can solve for the equilibrium as follows.

1. Guess the wage premium w_H.
2. Compute the prices implied by the guess for the wage premium using equation (15).
3. Guess a lump-sum transfer.
4. Solve household problem for all worker types.

(a) Guess a consumption aggregator C.

(b) Compute the expenditure function and its derivative, as given by equations (4) and (26).

(c) Guess labor supply n.

(d) Check whether the first order condition (12) holds. If yes, move on; if no, update guess for n.

(e) Check whether the budget constraint (9) holds. If yes, move on; if no, update guess for C.

5. Check whether the government budget clears (17). If yes, move on; if no, update guess for T.

6. Check whether the high-skilled labor market (20) clears. If yes, we have found the equilibrium; if no, update guess for w_H.

B.4 Alternative Pareto Weights

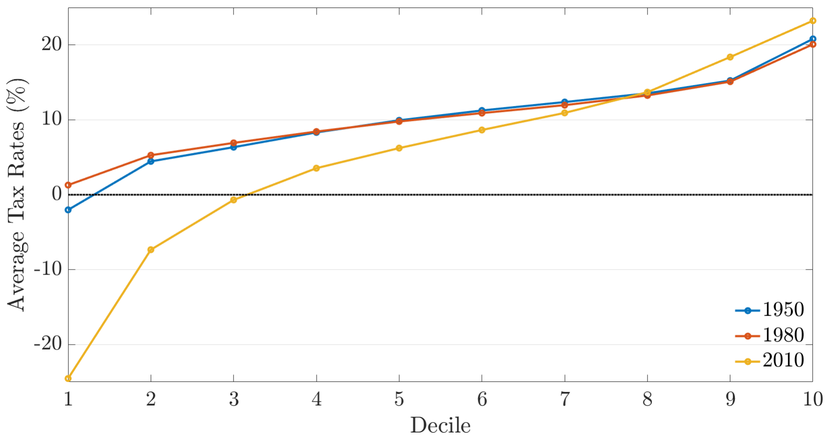

Figure 7: Optimal Average Tax Rates: U.S.

Notes: This figure shows optimal average rates given by the entire tax-and-transfer system by income decile for the U.S. in 1950, 1980, and 2010. Pareto weights are kept constant across years as described in the main text.

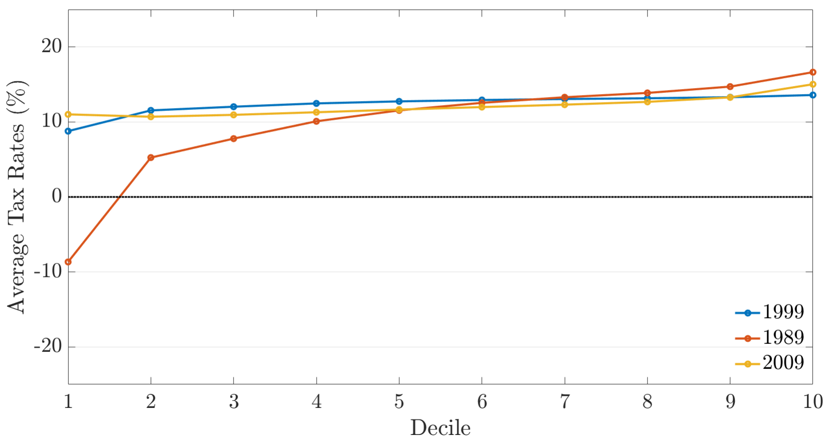

Figure 8: Optimal Average Tax Rates: China

Notes: This figure shows optimal average rates given by the entire tax-and-transfer system by income decile for China in 1989, 1999, and 2009. Pareto weights are kept constant across years as described in the main text.