Weighing Down the Top - Evaluating two Techniques to Estimate Tax Progressivity

Preliminary Version - Do not cite or circulate *

Johannes König†

October 21, 2019

Abstract

Progressivity is a central feature of the tax system, which has many implications for the economy's functioning. Applied, structural work demands a simple way to capture it. A power function approximation of the tax system delivers one parameter that captures the residual income elasticity. Because this approximation is fairly accurate, tractable, and interpretable it is immensely popular in applied, structural research. I show that the most popular way to estimate this parameter, i.e. in a log-log OLS specification, produces worse fit than a nonlinear least squares specification. The log-log OLS specification implicitly put weights on observations with high gross incomes that are drastically lower than observations at the bottom, leading to a higher estimated progressivity and a worse fit.

*I thank Richard Blundell, Patrick Burauel, Giacomo Corneo, Giulio Fella, Michael Graber, Robin Jessen, Maria Metzing, Carsten Schröder and Alexandros Theloudis as well as participants at the IIPF 2019 for valuable discussions and comments. The usual disclaimer applies.

†DIW Berlin/SOEP, Mohrenstr. 58, 10117 Berlin, Germany (e-mail: jkoenig@diw.de).

1 Introduction

The progressivity of the tax system is a central object of public and political debate when questions of vertical equity are raised. Further, it is a central object of the economic environment, which determines the agents' incentives and economic outcomes. Thus, measuring progressivity in a comprehensive and coherent way is highly desirable as it not only serves the public interest but is also a crucial ingredient in economic modeling.

In life-cycle models of individual behavior, accounting for progressive taxation is essential. Over their life-cycle individuals experience unanticipated changes in the remuneration of their work, which individuals would want to insure, but cannot since insurance markets are incomplete. The government can offer insurance of such risks by making the tax system progressive (insurance effect), but progressive taxation also induces a disincentive effect on, for example, labor supply or saving (see for example Varian (1980); Eaton and Rosen (1980)). Several papers, among them Blundell et al. (2016a); Heathcote et al. (2017a); Kaplan (2012), take these two effects into account in structural models of individuals' life-cycles. They introduce progressivity of the tax system by choosing a power function to approximate the relationship between gross and net income (see Feldstein, 1969). The exponent of this power function gives the degree of progressivity (elasticity of residual income due to Musgrave and Thin (1948)) of the tax system. The function is attractive because of three features: 1) *fit* to the empirical relationship , 2) *tractability* in (structural) economic models, and 3) *interpretability*, because it directly states progressivity.

In this paper I evaluate two techniques to estimate the power function: 1) OLS after a log-transformation of gross and net incomes (Log OLS) and 2) nonlinear least squares (NLS). On the face of it, both approaches should deliver mathematically equivalent results. I, however, demonstrate that Log OLS implicitly applies a weighting scheme, which reduces the weight of errors at the upper end of the income distribution. This results in the estimated progressivity being larger. This is intuitive: the higher one's gross income, the more closely will the effective tax system resemble a linear tax system because past the final bracket progression ceases and the system becomes linear. Therefore, the influence of high income observations will be to lower estimated progressivity. If their influence on the estimation is diminished, estimated progressivity will increase. Estimation using NLS does not apply this particular weighting, rather it weights all observations equally. As a result, NLS delivers a better fit than Log OLS in terms of root mean square error (RMSE) and mean absolute error (MAE).

The exponent of the power function giving the global residual income elasticity is a coarse measure of progressivity. However, the measure is still of high importance because of its utility in economic modeling. In particular, the measure is used in optimal progressivity exercises in macroeconomic and life-cycle models. A very relevant example of such an optimal progressivity model is Heathcote et al. (2017a). The authors specify a life-cycle model of consumption, saving, labor supply, and human capital investment with the tax and transfer system represented by the power function. The authors use Log OLS to estimate the exponent of the power function (the progressivity parameter) and find that it is lower than in the utilitarian optimum they derive. Using their data, I estimate the progressivity parameter using NLS and find that progressivity parameter is much larger than the estimate based on Log OLS.

The rest of the paper is structured as follows. In Section 2.2, I give a short account of the recent findings on progressivity and document the use of the power function approximation in the literature focusing on the branch that sees progressivity as an insurance device. In Section 3, I formally introduce the power function approximation and its relationship to the residual income elasticity. Section 4 introduces the two datasources. Section 5 presents the estimation results, while section 6 illustrates the importance of the findings. Section 7 concludes.

2 Tax Progressivity - Measurement and Modeling

In this section I will give a brief introduction to the recent debate on the measurement of progressivity in general and then give an account of the rationale behind including progressivity in structural models, which turns on the insurance-incentive trade-off.

2.1 Measurement

The debate on the degree of progressivity of the tax system has been reinvigorated by a several of papers authored by Thomas Piketty, Emmanual Saez, and Gabriel Zucman. A first account of the changing nature of progressivity over the last decades in the United States is given in Piketty and Saez (2007). The paper used administrative data on US taxpayers to trace the evolution of average tax rates along the distribution of income from 1960 to 2004. It focused on the top-tail of the distribution, showing that there had been precipitous drop in the tax rates faced by top income earners. For example, while the top 0.01% of earners faced an average federal tax rate of 71.6% in 1960, they faced a rate of 34.7% in 2004.

In Piketty et al. (2017) the authors trace the evolution of the distribution of national income over the period 1913 to 2014. Their findings are that incomes for the bottom 50% have stagnated, while incomes at the top have grown strongly. They show that this phenomenon coincides with the decline in progressivity of the tax system, raising suspicions of a causal relationship. In fact, while tax rates have tended to decline for the top 1%, tax rates have increased for the bottom half. However, many trends like "...widespread deregulation, weakened unions, and an erosion of the federal minimum wage"(Piketty et al., 2017, p.604) ran concurrent.

It should be noted that the findings by Piketty et al. (2017) are controversial. Auten and Splinter (2018) over roughly the same period considered in Piketty et al. (2017) find that taxes of the top 1% increased, while they fell for the bottom 90%. I will not attempt to settle this ongoing debate. Rather, I find it important to contextualize the findings of this study in the wider context of the literature on progressivity measurement.

Table 1. Some Literature Using the Power Function Approximation

Citation	Area	Subject
Gouveia and Strauss (1994)	Taxation	estimation of effective individual tax function
Li and Sarte (2004)	Growth	calibrated growth model investigating the effect of progressivity on growth
Conesa and Krueger (2006)	GE model	deriving optimal tax progressivity in a general equilibrium model
Kaplan (2012)	LC consumer model	life-cycle model describing income and wealth inequality
Blundell et al. (2016a)	LC labor supply model	life-cycle model of family labor supply
Heathcote et al. (2017a)	LC labor supply and human capital model	derives optimal tax progressivity in life-cycle model with labor supply and human capital investment
Boar and Midrigan (2019)	GE model	model of firm mark-ups and inequality

2.2 Modeling

While the above papers make important contributions to the economic literature and public debate by informing us about the connection between inequality and progressivity, the contributions remain descriptive. Embedding progressivity within an economic model that is capable of description and policy recommendations is a different matter. Often, to make a model operational one has to sacrifice accuracy in representing a relationship for tractability. This is certainly the case when progressivity is included in economic models. Because the power function approximation of the tax system is tractable in most models, represents the relationship fairly relative to its complexity, and is easily interpretable, it is widely used. A small and incomplete collection of previous papers using the approximation is provided in Table 1. As the table shows, the approximation has been used both in the macroeconomic and microeconomic literature. It can serve to simply describe the relationship between gross and net incomes, but it can also be employed for optimal tax exercises.

In life-cycle models, progressivity of the tax system has a special role, as it provides insurance of shocks to income.

Progressive Taxation as Insurance In a life-cycle setting, progressive taxation acts as an insurance device against otherwise uninsurable shocks. The idea that taxation can act as an insurance mechanism dates back to the 1980s with foundational contributions by Varian (1980) and Eaton and Rosen (1980). While Varian (1980) considers a dynamic model, in which an individual may self-insure against income risk via saving, Eaton and Rosen (1980) explicitly model labor supply with uncertainty in wages but neglect dynamics. Both Eaton and Rosen (1980) and Varian (1980) come to the conclusion that taxation can be desirable if an individual faces shocks and is risk averse. Varian (1980) considers an optimal nonlinear tax schedule with the finding that a more progressive tax schedule is optimal, when the Arrow-Pratt measure of absolute risk aversion is increasing. Under the imposition of a utility function that features constant relative risk aversion (CRRA) marginal tax rates increase in income, given that the coefficient of relative risk aversion is larger than 1.

Low and Maldoom (2004) make the connection between the two papers and examine optimal taxation when labor income is risky and both labor supply and savings are choice variables of the individual. The fundamental trade-off is between the incentive effect on labor supply stemming from income uncertainty and the benefit of social insurance that derives from lowering the variance of net income. They determine that the trade-off is parametrized by the ratio of prudence to risk aversion.

Several recent empirical studies like Blundell et al. (2016a), Blundell et al. (2015), Heathcote et al. (2014), and Heathcote et al. (2017a) seek to estimate the degree of insurance stemming from sources like savings, (family) labor supply and taxes over the life-cycle. Blundell et al. (2016a) find that insurance via progressive taxation makes up a sizable contribution ($\sim$11%) to insurance of permanent wage shocks, but other forces, most prominently (family) labor supply, dominate.

For Norwegian data analyzed in Blundell et al. (2015) the riskiness of earnings is strongly attenuated by the tax and transfer system, especially for those with lower education, who experience roughly 20% less impact of a permanent shock of one standard deviation on annual disposable compared to annual market income. Heathcote et al. (2014) calculate the overall insurance provided by the aforementioned mechanisms. They do, however, find that own labor supply dampens the effect persistent shocks have on consumption by roughly 15%. Heathcote et al. (2017a) go a step beyond the description of insurance mechanisms and provide a closed-form expression for social welfare, which crucially depends on the riskiness of earnings. Finally, they characterize the progressivity of the optimal income tax, given by the size of the power function exponent, according to a utilitarian objective. They find that progressivity is too high in the status-quo compared to the optimum.

3 Modeling the Retention Function

This section introduces the power function approximation of the retention function. The retention function $\mathcal{T}(\cdot)$ takes as inputs gross income as well as characteristics of the tax unit and returns net income.[1] The function $\mathcal{T}(\cdot)$ is nonlinear, not continuous, and therefore non-differentiable. It is not tractable, depending on the type of life-cycle model, to choose a non-continuous, non-differentiable, or highly complex retention function.[2] The power

[1] In this paper I will neither make an explicit distinction between pre-government and gross income nor between post-government and net income; I will define the terms in section 4.

[2] While life-cycle models solved by numerical dynamic optimization can accomodate a non-continuous, non-differentiable retention function, it is often computationally too expensive to do so. When model solution proceeds via the first-order approach, while also approximating the life-time budget constraint, it turns out to be very advantageous to choose a power-function approximation, as the resulting structural

function approximation of $\mathcal{T}(\cdot)$ is attractive because it is simple, differentiable, and easy to interpret.

3.1 The Power Function Approximation

Choosing a power function to approximate the retention function was popularized by Feldstein (1969). The relationship between gross ($y_{i,t}$) and net income ($y_{i,t}^{\text{net}}$) is given by

$$y_{i,t}^{\text{net}} = \mathcal{T}(y_{i,t}) \approx \chi y_{i,t}^{1-\tau}. \tag{1}$$

Or, after a log-transformation,

$$\ln y_{i,t}^{\text{net}} \approx \ln \chi + (1-\tau) \ln y_{i,t}. \tag{2}$$

The parameter τ determines the curvature of the retention function $\mathcal{T}$ and thus the progressivity of the tax system. At $\tau = 0$, the tax system would be linear and χ would give the (average) tax rate. When τ is larger than zero, the system is progressive, and analogously it is regressive if τ is smaller than zero. In equation (1), the parameters χ and τ are neither individual- nor time-specific. Accordingly such an approximation will miss out on much of the variation that is driven by i) the differences in the assessment criteria that apply to the particular tax unit, e.g. whether the tax unit consists of a couple filing jointly or a single, and ii) the differences in the relevant parameters of the tax code over time, e.g. the top tax rate. Further, the function is constrained to start at the origin. Thus, any transfer recipients with zero gross income cannot be represented well by this function.

3.2 Measuring Progressivity

The choice of the power function is also motivated to a large extent by its interpretability. The parameter τ is crucial metric discussed in the economics of taxation: the residual income elasticity , i.e. the elasticity of net income with respect to gross income (Musgrave and Thin, 1948). The residual income elasticity can be expressed as $\frac{\partial(y-T(y))}{\partial y}\frac{y}{y-T(y)} = \frac{1-\partial T(y)/\partial y}{1-T(y)/y}$, where $T(\cdot)$ gives the tax liability. Jakobsson (1976) showed that when one compares two tax schedules T_1 and T_2, one is more progressive than the other, if

$$\frac{1-\partial T_1(y)/\partial y}{1-T_1(y)/y} < \frac{1-\partial T_2(y)/\partial y}{1-T_2(y)/y} \quad \forall y. \tag{3}$$

One implication of Jakobsson's theorem is that for a progressive tax schedule $\partial T(y)/\partial y > T(y)/y \ \forall y$.[3] In the case of the power function approximation,

$$\frac{\partial\left(y-T(y)\right)}{\partial y}\frac{y}{y-T(y)} = \frac{\partial \mathcal{T}(y)}{\partial y}\frac{y}{\mathcal{T}(y)} = 1-\tau.$$

equations will be linear in log-space and the model becomes tractable (see Blundell et al. (2016a) and Blundell et al. (2016b)).

[3] Accordingly, when we approximate progressive tax schedules with continuous and differentiable functions, this implies that the approximation function has to be strictly convex. For intuition why this must be the case, construct the limiting case where $T(\cdot)$ is linear and therefore $\partial T(y)/\partial y = T(y)/y \ \forall y$.

Thus, the residual income elasticity can be directly determined from the exponent of the power function.

4 Data

In my empirical exercises I employ two datasets: 1) the replication Panel Study of Income Dynamics (PSID) dataset by Heathcote et al. (2017a) and 2) my construction of the Congressional Budget Office (CBO) dataset on net and gross incomes following Heathcote et al. (2017a). For the first dataset the authors use the tax calculator TAXSIM (Feenberg and Coutts, 1993) provided by the National Bureau of Economic Research (NBER).

PSID Data As detailed in Heathcote et al. (2017a,b), the authors use four waves of the PSID from 2000 to 2006. Note that the data is collected biennially. Pregovernment/gross income includes all types of labor and capital income as well as private transfers. Taxable income is derived by subtracting deductions from gross income. The deductions are medical expenses, mortgage interest, state taxes paid, and charitable giving. Finally they add 50% (the employer's share) of the FICA tax to taxable income. Postgovernment/net income is gross income minus taxes plus transfers. The taxes are federal, state, and FICA tax, while the transfers are the main public cash transfers (TANF, SSI, and so on). The authors impute the marginal Social Security benefits earned that year because they have subtracted FICA tax. They do not make such an imputation for medicare benefits because they are conditional on age and not working. They restrict their sample to households with heads between the ages of 25 and 60 and those where at least one person in the household earns more than the part time-equivalent of the minimum wage.

CBO Data To perform a robustness check, Heathcote et al. (2017a) use CBO tables on the distribution of household income (CBO, 2019). The CBO publishes tables of post-goverment income, taxes, and transfers along the distribution of pre-goverment income. The tables report means within quintiles and for the percentile groups 81-90, 91-95, 96-99, and the top 1% where the ordering variable are measure of pre-government income. The CBO defines two measures of pre-government income: 1) market income, which includes labor, business, capital, and retirement income from non-governmental sources and 2) income before transfers and taxes, which is market income plus social security benefits. Heathcote et al. (2017a) are not specific on the measure they use, so I use both. Like Heathcote et al. (2017a) I subtract Medicare from post-government income. I use all data from 2000 to 2006. I generate a weight variable to account for the different shares of the population each observation in the dataset represents.

5 Results

This section is grouped into two subsections. First, I replicate the results in Heathcote et al. (2017a) using the provided PSID data. I illustrate the difference between estimates derived using OLS after log-transformation, conforming to equation (2) and nonlinear least

Table 2. Progressivity Estimates and GOF in Heathcote et al. 2017 Dataset

	full sample		excl. outlier	
	Log OLS	NLS	Log OLS	NLS
$1-\tau$	0.819	0.936	0.819	0.902
	(0.00506)	(0.0223)	(0.00516)	(0.00532)
RMSE	17878	8933	11357	6894
MAE	4457	4566	4338	3989
Obs.	12875	12875	12874	12874

Note: Own calculation based on Heathcote et al. (2017a) replication dataset. Bootstrapped standard errors in parentheses based on 500 replicates.

squares, conforming to equation (1). Second, I perform the robustness check conducted by Heathcote et al. (2017a) using the CBO data again with both estimation techniques.

5.1 PSID

I use the PSID data to estimate the power function approximation using Log OLS and NLS. The estimates of $1-\tau$, bootstrapped standard errors, as well as fit statistics are presented in table 2. In columns 1 and 2 I compare the models using the full dataset, while I compare the models when excluding one very high gross income outlier (>5 m) in columns 3 and 4. For columns 1 and 2 the estimate of $1-\tau$ using Log OLS is significantly smaller (0.819) than the one obtained by nonlinear least squares (0.936) implying a more progressive tax system.

Goodness of fit is assessed by computing error from observed data and model predictions, i.e. $e_{i,t} = y_{i,t}^{\text{net}} - \hat{y}_{i,t}^{\text{net}}$, where $\hat{y}_{i,t}^{\text{net}}$ is the respective prediction of net income.[4]. I then compute either the root mean square error (RMSE) or mean absolute error (MAE) from these errors. The RMSE indicates that the model estimated using NLS fits roughly *twice* as well as Log OLS. The MAE on the other hand is smaller for the log-transformed model, but not as significantly as for the RMSE (roughly 2% smaller). Thus, Log OLS appears to fit well in the middle of the distribution, but is not very sensitive to high-income outliers. This is a consequence of the log-transformation: outliers in the top tail of the income distribution are less influential after log transformation than in levels.

Excluding outliers is a double-edged sword. On the one hand it reduces the sensitivity of the progressivity estimate and excludes observations that are possibly mismeasured, on the other hand it excludes possibly valid observations and mechanically increases progressivity because for the top tail of the income distribution the tax system becomes linear past the final tax bracket.

To illustrate the influence of extreme outliers, I exclude one very high income outlier in columns 3 and 4 of table 2. I also do this because this outlier appears in only one year and is extremely far removed from the rest of the distribution. The progressivity estimate of Log OLS does not budge (0.819), while the progressivity estimate of NLS drops (0.902)

[4]In the case of Log OLS I adjust for Jensen's inequality in the prediction. See for example Rainey (2017)

with a much smaller standard error. Goodness of fit improves in both specifications. For Log OLS the RMSE falls by about 37% and the MAE drops marginally by about 2.7%, while for the NLS specification the RMSE drops by about 23% and the MAE drops by about 13%. The fit for the NLS specification improves strongly especially in terms of the MAE. Without the outlier, both RMSE and MAE are now lower in the NLS specification, but the significant difference in the estimate of $1 - \tau$ remains. The same holds for the large difference in the RMSE between log-transformed and NLS specifications: the RMSE of the Log OLS specification is still roughly 65% larger. The MAE is roughly 8% smaller for the NLS specification.

To gain a visual impression of the difference between the two procedures, I plot the absolute errors ($|e_{i,t}|$) produced by either procedure along the distribution of gross income in figure 1. The dark gray dots indicate errors produced by the Log OLS procedure, the light gray dots indicate errors for the NLS procedure. Both in the complete sample (panel a) and when excluding the outlier (panel b), the dark gray dots lie above the light gray dots for large values of gross income, for small values of gross income the reverse appears to be true, however, the difference is small. The figure points out the fact that the Log OLS procedure yields large errors at the top of the income distribution. However, the figure also illustrates that the fit at the bottom is not very different between Log OLS and NLS.

Implicit Weights As figure 1 and table 2 show, the procedures differ radically in terms of the weight they implicitly attach to a given observation along the gross income distribution. I can calculate these weights by making a comparison of the individual squared errors. For each observation (i, t) the following relationship holds:

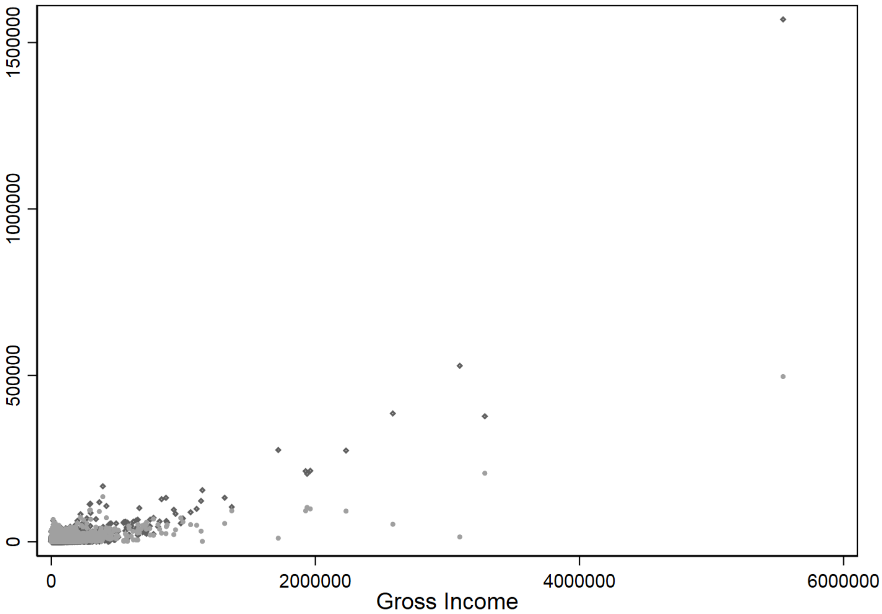
$$\begin{aligned} e^2_{i,t,\text{Log OLS}} &= w_{i,t} e^2_{i,t,\text{NLS}} \\ w_{i,t} &= \frac{e^2_{i,t,\text{Log OLS}}}{e^2_{i,t,\text{NLS}}}, \end{aligned} \tag{4}$$

where $e_{i,t,\text{NLS}}$ is the error due to the NLS procedure and $e_{i,t,\text{Log OLS}}$ is the error due to the Log OLS procedure. $w_{i,t}$ is a weight per observation that adjusts the squared errors produced by the NLS procedure so that NLS would produce the errors of the Log OLS procedure. Thus $w_{i,t}$ gives the weight that reproduces the results of Log OLS when estimating NLS.

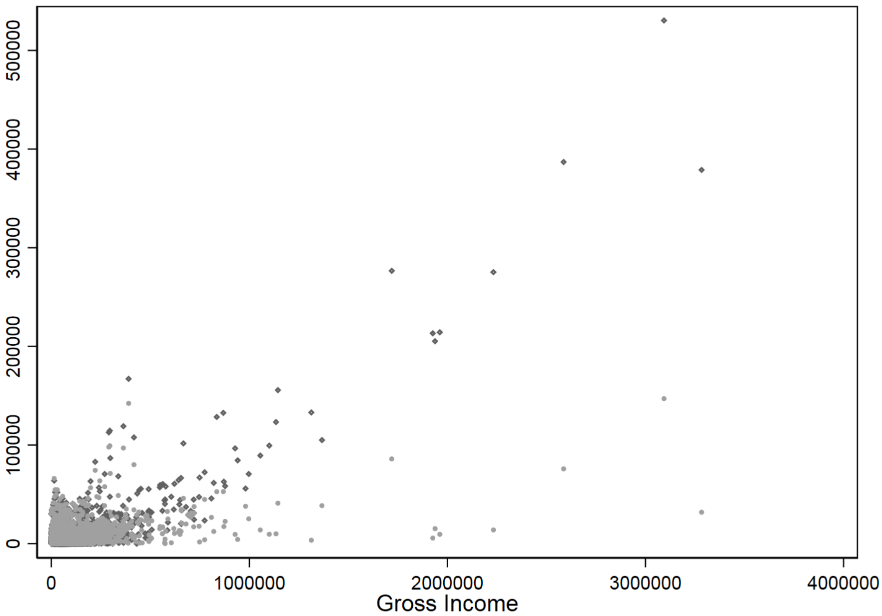
I show the weights $w_{i,t}$ along the distribution of gross income in figure 2. Panel (a) of the figure shows results for all observations. It becomes apparent that if one wanted to achieve the same results with NLS that one obtained with Log OLS, one would need to give extraordinary weight (greater than 200 m) to observations at low gross incomes. The qualitative pattern holds even if we drop some of the highest $w_{i,t}$ and only focus on those below 100,000. These results are shown in panel (b) of the figure.

Progressivity Estimates over Time Finally, I want to show the same comparison of procedures over time. As changes to the tax system and changes of the distribution

Figure 1. Absolute Errors along the Distribution of Gross Income



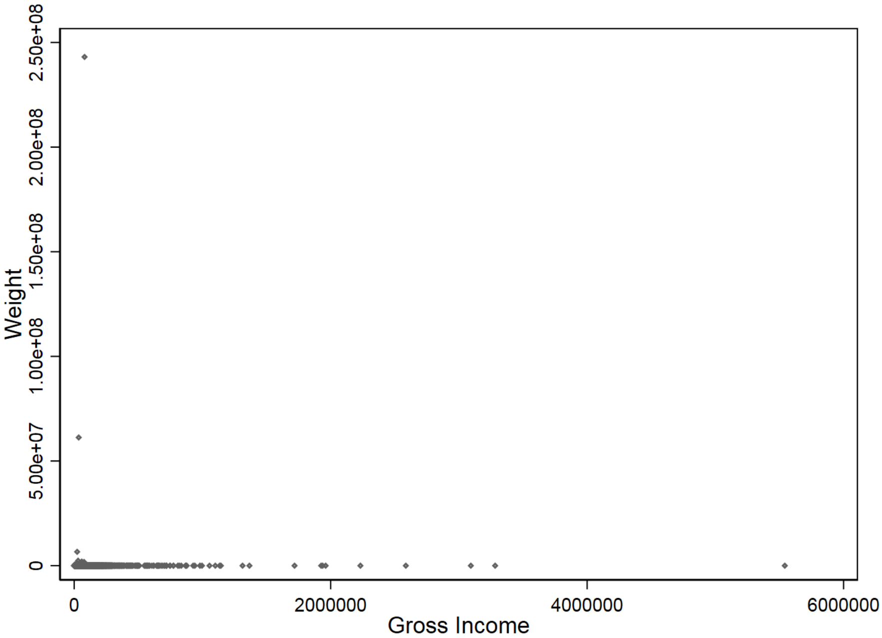
(a) Complete Sample



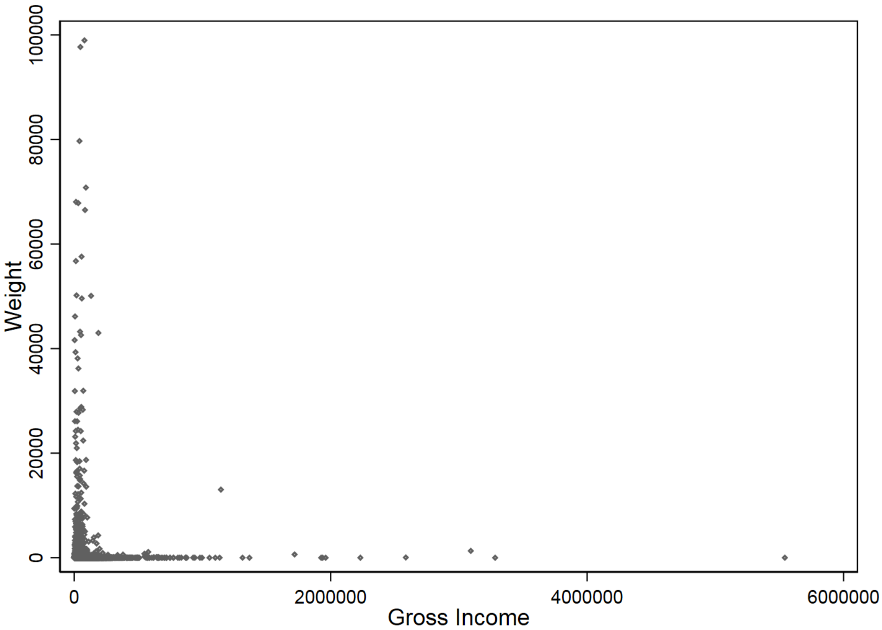
(b) Excluding Outlier

Note: Own calculation based on Heathcote et al. (2017a) replication dataset. I plot the absolute errors, i.e. the absolute difference between observed and predicted net income for the Log OLS (drak gray) and the NLS (light gray).

Figure 2. Weights w_i along the Distribution of Gross Income



(a) Complete Sample



(b) Excluding $w_{i,t}$>100,000

Note: Own calculation based on the Heathcote et al. (2017a) replication dataset. I plot the weights $w_{i,t}$ in equation 4.

of gross incomes will impact the estimated progressivity, I want to inspect how close the estimates cluster over the years to gauge whether pooling the data, as is frequently done in the literature, is advisable and to study the evolution of progressivity. Table 6 shows the results. First, the result that Log OLS always produces lower estimates of $1-\tau$ than NLS holds in all years. Second, results for both procedures cluster fairly consistently: Log OLS delivers results around 0.81 and NLS results around 0.9. Third, in all years NLS delivers a better fit in terms of the RMSE and in most years, the years without the outlier, it also produces better fit in terms of the MAE. In appendix A I show the analogous results when excluding outliers. The results show even more consistency regarding NLS; $1-\tau$ is roughly at 0.9 each year.

Table 3. Progressivity Estimates and GOF in Heathcote et al. 2017 Dataset by Year

	2000		2002		2004		2006	
	Log OLS	NLS	Log OLS	NLS	Log OLS	NLS	Log OLS	NLS
$1-\tau$	0.840	0.895	0.811	0.895	0.811	0.960	0.818	0.910
	(0.00648)	(0.00769)	(0.0131)	(0.00603)	(0.00851)	(0.0293)	(0.00928)	(0.0147)
RMSE	6952	5081	11855	6370	30864	9813	12559	7902
MAE	3141	3048	4075	3663	5183	5431	4907	4572
Obs.	3198	3198	3204	3204	3266	3266	3207	3207

Note: Own calculation based on Heathcote et al. (2017a) replication dataset. Bootstrapped standard errors in parentheses based on 500 replicates.

In terms of trends in progressivity, both procedures tell two different tales. The estimates from Log OLS suggest that progressivity increased over the observation period. Progressivity was lower in 2000, rose in 2002, and stayed roughly flat from then on. NLS estimates suggest that progressivity stayed flat throughout the period, especially when excluding the high income outlier.

Discussion The above results show that the log-transformed model does not give a lot of weight to observations in the top tail of the income distribution while giving a lot of weight to observations in lower tail of the income distribution. Log OLS therefore neglects a crucial part of the distribution that is taxed almost according to a flat schedule rather than a progressive one. As Scheuer and Slemrod (2019, p. 5) point out: "For those with income far beyond the threshold, the average tax rate,..., is well approximated by the top rate." The NLS specification is more sensitive to these observations in the top tail. Because the NLS specification is sensitive to these observations, it delivers a better fit in terms of the RMSE and the MAE.

5.2 CBO Tables

Heathcote et al. (2017a) cross-check their estimates of the progressivity parameter using data from the Congressional Budget Office (CBO). They give several reasons to perform this check. Most prominent and most pertinent to the result above, they acknowledge that the PSID undersamples the very rich, which would be precisely the group that would lower

Table 4. Progressivity Estimates using CBO Dataset

	income before transfers and taxes		market income	
	Log OLS	NLS	Log OLS	NLS
1-τ	0.790	0.948	0.746	0.927
Obs.	56	56	56	56

Note: Own calculation based on CBO (2019).

the progressivity estimates. Other reasons are that taxes are imputed through TAXSIM with the PSID and that the PSID only covers part of all in-kind transfers.

Table 4 contains the estimates of the progressivity parameter for a Log OLS specification and a NLS specification with either measure of pre-government income that the CBO reports. I do not report standard errors because they cannot help in the interpretation of the results with this dataset.

The Log OLS estimate I obtain for $1-\tau$ using income before transfers and taxes is fairly close to the one reported in Heathcote et al. (2017a), namely 0.79 so that τ must be equal to 0.21. Unsurprisingly, because the very rich are more well-represented in the CBO data, the progressivity estimate using NLS is a bit higher than the analogue in the PSID at 0.948, but it is still fairly close to the estimate including all observations in table 2. The estimates are different when I consider market income. In this case the implied estimate of $1-\tau$ is 0.746 using Log OLS. The NLS estimate implies more progressivity compared to income before transfers and taxes: $1-\tau$ drops to 0.927.

Discussion The estimates derived from CBO data reveal that a more accurate representation of the very rich leads to an increase of the estimated progressivity parameter $1-\tau$ for the NLS specification. Depending on whether one chooses income before transfers and taxes or market income, estimates of the progressivity parameter vary quite a bit. The estimate obtained with market income is quite different from the estimate using income before transfers. The reason is clear, market income does not include social security income making the relationship to post-government income, where the transfers are included, more progressive. The Log OLS model is clearly more sensitive to this change. With both pre-government incomes considered here, Log OLS delivers progressivity estimates that are close to PSID estimates but suggest slightly more progressivity. This is counterintuitive: the CBO data represents the top tail of the income distribution more accurately, which should lower the progressivity estimate. In contrast, this is precisely what is observed with NLS. NLS estimates of $1-\tau$ are somewhat higher than the analogous PSID estimates, suggesting less progressivity.

6 Impact

Above I have examined the fit of the two procedures to data. While the differences in estimates are certainly statistically significant, I have not discussed whether they would be relevant for the applied researcher. In this section I will pursue two exercises to show that the differences in estimates are economically relevant. First, I will predict net income from a synthetic dataset of gross incomes either following the relationship implied by Log OLS or NLS. Second, I will use the two sets of estimates of τ to derive hypothetical tax-adjusted Frisch elasticities.

6.1 Net Income Prediction

I adopt the following procedure to predict net incomes: I choose the set of estimates from the PSID using the sample without the outlier. Using these estimates, I predict net incomes from a small synthetic dataset of gross income, that I chose, which is shown in table 7 in the appendix. The synthetic distribution contains pre-government incomes from 0 to 3 Million Dollars. From this exercise I can produce a set of graphs that trace the net-of-tax function along gross income. I show in figure 3 these functions both for the full distribution of gross incomes (upper panel) and for gross incomes up to 300,000 Dollars (lower panel).

The upper panel of figure 3 shows that using the NLS estimates, which features the higher $1-\tau$ of 0.9, large differences in net incomes emerge at around 500,000 Dollars of gross income. From then on, as implied by the difference in the exponent, the differences get larger and larger. The figure also illustrates the findings from figure 1: while fit at the bottom appears virtually the same for both functions, fit at the top is very different.

To check whether the fit actually is different at the bottom of the distribution, I present the bottom panel of the figure. Here one can see that the Log OLS prediction actually starts out more progressive than the NLS prediction. This holds up until about 100,000 Dollars of income, when the NLS prediction crosses the Log OLS prediction line. But as is evident from 1 and from the bottom panel, the differences in predictions are small and cannot outweigh the progressively worse fit that Log OLS shows at the top end of the distribution.

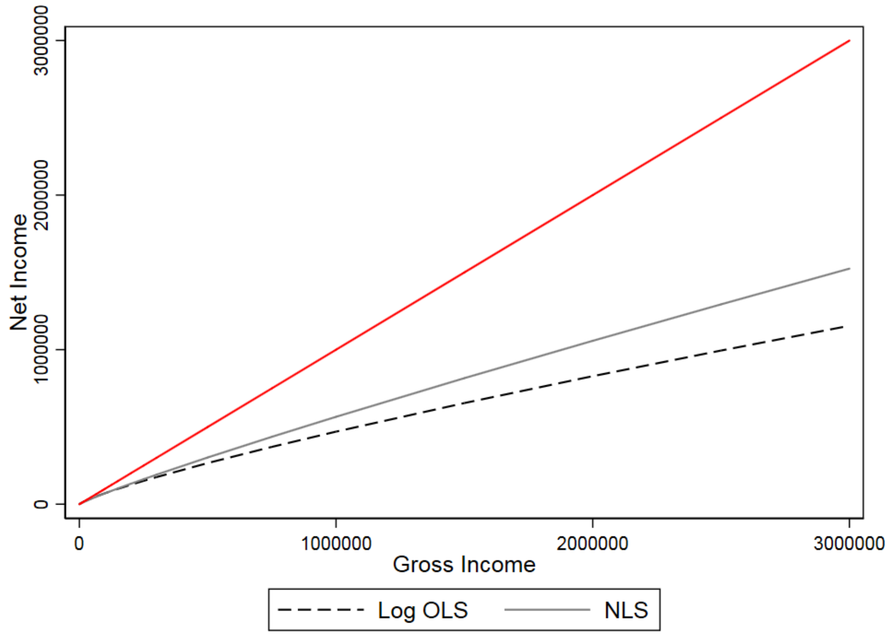
6.2 Frisch Elasticity Estimates

As evident from the above discussion of taxation as insurance in section 2.2, the progressivity parameter plays a major role in life-cycle models of labor supply as it influences how shocks – from the wage process or some other source – transmit to income, consumption and ultimately life-time utility. Further, it influences the labor supply reaction of individuals by attenuating the incentive to respond to shocks. Consider the Frisch elasticity of labor supply, i.e. the elasticity giving determining the reaction to a transitory wage shock. Estimates of the Frisch elasticity can be obtained from the estimation of an intertemporal labor supply equation. An example of such an equation, which can be derived from the model given in appendix C, is

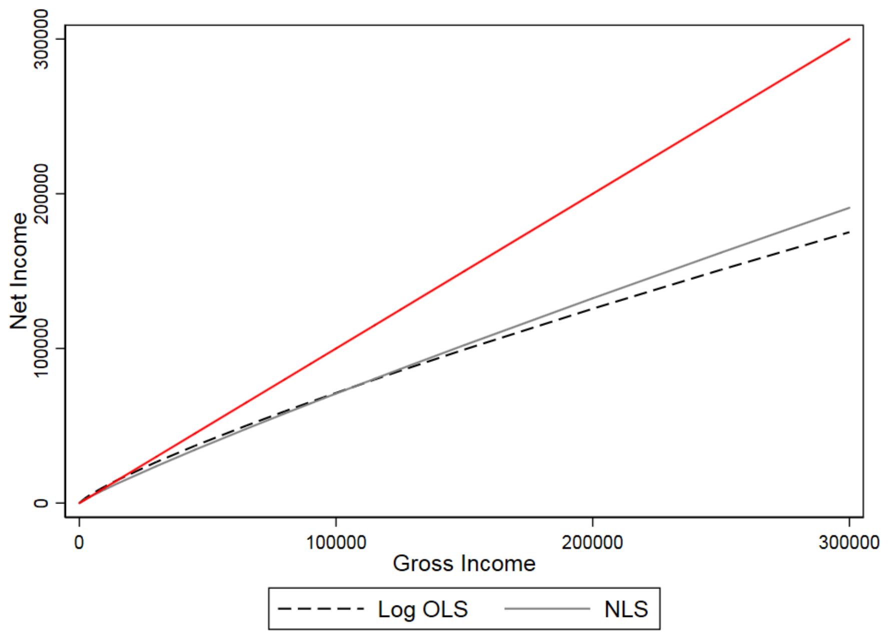
$$\Delta \ln h_t \approx \frac{1}{\gamma + \tau} \left[cons_t + (1-\tau) \Delta \ln w_t - \varsigma \Delta \Xi_t + \Delta \upsilon_t + \eta_t \right], \tag{5}$$

where Δ is the first difference operator, h are hours worked, $1/\gamma$ is the Frisch elasticity, $cons_t$ contains time-specific constants, w is the wage, Ξ_t are variables determining taste for

Figure 3. Predicted Net Incomes



(a) Full Distribution



(b) Pre-Gov. Income up to 300,000

Note: Own calculation based on the synthetic dataset in table 7. The red line is the 45 degree line.

work, υ are shocks to the taste for work and η is the approximation error of the marginal utility of wealth. Finally, $1-\tau$ is the exponent of the power function approximation of the tax system. The equation can be derived from the first-order conditions of the model in appendix C and then approximating the intertemporal Euler equation with a First-Order Taylor expansion. This approach is common in the literature and thus the above equation is instructive. The reaction in $\Delta \ln h_t$ to a change in $\Delta \ln w_t$ is $\frac{1-\tau}{\gamma+\tau}$. In the case without progressive taxation, when τ equals zero, the reaction would be $\frac{1}{\gamma}$, that is the ordinary Frisch elasticity. However, when equation (5) is estimated, the econometrician estimates the tax-adjusted Frisch elasticity. To calculate the unadjusted Frisch elasticity, an estimate of τ is needed. In table 5 I give an illustration of what the unadjusted Frisch elasticity would look like, for different tax-adjusted Frisch elasticity estimates and two different estimates of τ. The range of estimates of the tax-adjusted Frisch elasticities loosely follows the range of estimates presented in Keane (2011). The two estimates of τ correspond to the one estimated by Log OLS (0.18) and the one estimated by NLS (0.1)

Table 5. Unadjusted Frisch Elasticites

	τ	
	0.18	0.1
tax-adj. Frisch	unadj. Frisch	unadj. Frisch
0.5	0.685	0.588
0.4	0.535	0.465
0.3	0.392	0.345
0.2	0.255	0.227

Note: Own calculation.

The table shows that at low tax-adjusted Frisch elasticities, the diifference between an adjustment with either 0.18 or 0.1 is quite small, but as the estimates grow, so does the difference caused by the adjustment. At a tax-adjusted Frisch of 0.5, the difference between adjustment with τ=0.18 or τ=0.1 is about 0.1.

7 Conclusion

In the course of this paper I have examined two procedures to estimate the global progressivity parameter of the U.S. tax system. The Log OLS procedure is the most commonly applied and should, at least on paper, deliver the same estimates as the NLS procedure. When I apply the two procedures to PSID data and CBO data the following facts emerge: 1) Log OLS estimates of the progressivity parameter $1-\tau$ are always lower than the estimates obtained by NLS. This holds true, even if I remove a high income outlier and if I estimate year-by-year. 2) The Log OLS predictions of net income always produce larger root mean square errors and, most of the time, larger mean absolute errors when using PSID data. 3) In CBO data the top tail of the income distribution is represented more accurately. Despite this, Log OLS estimates of progressivity are larger. In contrast to this and in line with expectation, progressivity estimates using NLS are smaller.

The mechanism underlying these facts is the difference in the implicit weighting of observations along the pre-government income distribution that both procedures imply. As shown in figure 2, Log OLS heavily down-weights observations at the top of the distribution compared to NLS. Naturally, it follows that Log OLS implies more progressivity because high income observations, that face an almost linear tax system and would therefore drive progressivity down, are close to being ignored by the procedure.

As I have pointed out in the introduction, the use of the power function approximation and Log OLS as well as the use of these estimates in economic modeling is widespread. My recommendation to the researchers either estimating Log OLS or relying on estimates derived from Log OLS is to switch NLS. NLS delivers estimates that better approximate the tax system, as it outperforms by Log OLS on fit in both RMSE and MAE. Further, NLS does not exhibit the unintuitive behavior on CBO data. NLS reacts to the inclusion of more high income observations with an estimate that implies less progressivity compared to estimates using PSID. Log OLS indicates slightly more progressivity compared to estimates using PSID. The distinction is not trivial as the exercises in section 6 show. The different estimates of progressivity imply very different tax burdens and imply very different estimates of the unadjusted Frisch elasticity of labor supply.

A Progressivity over Time when Excluding Outliers

Table 6. Progressivity Estimates and GOF in Heathcote et al. 2017 Dataset by Year Excluding Outliers

	2000		2002		2004		2006	
	Log OLS	NLS	Log OLS	NLS	Log OLS	NLS	Log OLS	NLS
$1-\tau$	0.840	0.895	0.811	0.895	0.810	0.905	0.818	0.910
	(0.00650)	(0.00740)	(0.0131)	(0.00603)	(0.00851)	(0.0293)	(0.00928)	(0.0147)
RMSE	6952	5081	11855	6370	12914	6701	12559	7902
MAE	3141	3048	4075	3663	4702	4199	4907	4572
Obs.	3198	3198	3204	3204	3265	3265	3207	3207

Note: Own calculation based on Heathcote et al. (2017a) replication dataset. Bootstrapped standard errors in parentheses based on 500 replicates. RMSE and MAE calculated based on the difference between prediction in levels and net income. I have adjusted for Jensen's inequality in the log model prediction. I have excluded gross incomes larger than 5 million, which is one observation.

B Synthetic Data

Table 7. Synthetic Pre-Government Income and Predicted Net Incomes

Pre-Gov. Income	Post-Gov. Income	
	Log OLS	NLS
0	0	0
2000	2893	2077
4000	5104	3882
8000	9004	7256
10000	10809	8875
15000	15066	12795
30000	26578	23913
45000	37045	34475
65000	50063	48040
80000	59342	57938
90000	65352	64434
95000	68310	67655
100000	71241	70860
115000	79880	80383
125000	85525	86663
150000	99297	102159
200000	125676	132435
250000	150874	161972
300000	175170	190933
500000	266158	302721
750000	370977	436436
1000000	469531	565778
1500000	654443	815688
2000000	828302	1057426
2500000	994378	1293263
3000000	1154505	1524502

C Life-Cycle Model with Labor Supply

A model giving rise to a labor supply equation equaling (5) is defined here. The consumer chooses consumption and work hours to maximize life-time utility.

$$\max_{c_t,h_t} E_{t_0}\left[\sum_{t=t_0}^{T}\rho^{t-t_0}\mathrm{v}(c_t,h_t,b_t)\right], \tag{6}$$

where v is the in-period utility function taking consumption c_t, hours of work h_t, and taste-shifters b_t as arguments. ρ is the discount factor. I specify the taste shifter $b_t = \exp(\varsigma\Xi_t - \upsilon_t)$. $\varsigma\Xi_t$ is a linear combination of a set of personal characteristics. υ_t accounts for the non-systematic variation of the taste shifter, which is assumed to be normally distributed and uncorrelated over time. The functional form of the in-period utility function is given by

$$\mathrm{v}(c_t,h_t,b_t) = \frac{c_t^{1-\vartheta}}{1-\vartheta} - b_t\frac{h_t^{1+\gamma}}{1+\gamma}, \qquad \vartheta \geq 0, \gamma \geq 0, \tag{7}$$

where $\frac{1}{\vartheta}$ pins down the intertemporal elasticity of substitution with respect to consumption, while $\frac{1}{\gamma}$ gives the Frisch-elasticity of labor supply. Thus, in-period utility is additively-separable and conforms to constant relative risk aversion (CRRA).

The intertemporal budget constraint is

$$\frac{a_{t+1}}{(1+r_t)} = a_t + \chi\left(w_t h_t\right)^{1-\tau} - c_t, \tag{8}$$

where a_t represents assets, r_t the real interest rate,[5] Net income from labor supply is given by $\chi\left(w_t h_t\right)^{1-\tau}$.

References

Auten, G. and D. Splinter
2018. Income inequality in the united states: Using tax data to measure long-term trends. *Draft subject to change. http://davidsplinter. com/AutenSplinter-Tax_ Data_ and_ Inequality. pdf.*

Blundell, R., M. Graber, and M. Mogstad
2015. Labor income dynamics and the insurance from taxes, transfers, and the family. *Journal of Public Economics*, 127:58 – 73. The Nordic Model.

Blundell, R., L. Pistaferri, and I. Saporta-Eksten
2016a. Consumption inequality and family labor supply. *American Economic Review*, 106(2):387–435.

Blundell, R., L. Pistaferri, and I. Saporta-Eksten
2016b. Consumption inequality and family labor supply: Online appendix.

Boar, C. and V. Midrigan
2019. Markups and inequality. Working Paper 25952, National Bureau of Economic Research.

CBO
2019. The distribution of household income, 2016. Technical report.

Conesa, J. C. and D. Krueger
2006. On the optimal progressivity of the income tax code. *Journal of Monetary Economics*, 53(7):1425 – 1450.

[5] Accordingly, the model features an incomplete capital market.

Eaton, J. and H. S. Rosen
1980. Labor supply, uncertainty, and efficient taxation. *Journal of Public Economics*, 14(3):365–374.

Feenberg, D. and E. Coutts
1993. An introduction to the taxsim model. *Journal of Policy Analysis and management*, 12(1):189–194.

Feldstein, M. S.
1969. The effects of taxation on risk taking. *Journal of Political Economy*, 77(5):755–764.

Gouveia, M. and R. P. Strauss
1994. Effective federal individual income tax functions: An exploratory empirical analysis. *National Tax Journal*, Pp. 317–339.

Heathcote, J., K. Storesletten, and G. L. Violante
2014. Consumption and labor supply with partial insurance: An analytical framework. *American Economic Review*, 104(7):2075–2126.

Heathcote, J., K. Storesletten, and G. L. Violante
2017a. Optimal tax progressivity: An analytical framework. *The Quarterly Journal of Economics*, 132(4):1693–1754.

Heathcote, J., K. Storesletten, and G. L. Violante
2017b. Optimal tax progressivity: An analytical framework: Online appendix.

Jakobsson, U.
1976. On the measurement of the degree of progression. *Journal of Public Economics*, 5(1):161 – 168.

Kaplan, G.
2012. Inequality and the life cycle. *Quantitative Economics*, 3(3):471–525.

Keane, M. P.
2011. Labor supply and taxes: A survey. *Journal of Economic Literature*, 49(4):961–1075.

Li, W. and P.-D. Sarte
2004. Progressive taxation and long-run growth. *The American Economic Review*, 94(5):1705–1716.

Low, H. and D. Maldoom
2004. Optimal taxation, prudence and risk-sharing. *Journal of Public Economics*, 88(3):443 – 464.

Musgrave, R. A. and T. Thin
1948. Income tax progression, 1929-48. *Journal of Political Economy*, 56(6):498–514.

Piketty, T. and E. Saez
2007. How progressive is the us federal tax system? a historical and international perspective. *Journal of Economic perspectives*, 21(1):3–24.

Piketty, T., E. Saez, and G. Zucman
2017. Distributional national accounts: methods and estimates for the united states. *The Quarterly Journal of Economics*, 133(2):553–609.

Rainey, C.
2017. Transformation-induced bias: Unbiased coefficients do not imply unbiased quantities of interest. *Political Analysis*, 25(3).

Scheuer, F. and J. Slemrod
2019. Taxation and the superrich. Technical report, National Bureau of Economic Research.

Varian, H. R.
1980. Redistributive taxation as social insurance. *Journal of Public Economics*, 14(1):49–68.