LIS
Working Paper Series

No. 823

Optimal Labor Income Taxation – The Role of the Skill Distribution

Dingquan Miao

January 2022

Luxembourg Income Study (LIS), asbl

Optimal Labor Income Taxation - The Role of the Skill Distribution*

Dingquan Miao†

Abstract

I analyze the role of the distribution of skills in shaping optimal nonlinear income tax schedules. I use theoretical skill distributions as well as empirical skill distributions for 14 OECD countries. I find that a more dispersed log-normal skill distribution implies a more progressive optimal tax schedule. Optimal tax rates should be lower throughout if a greater number of unskilled agents cluster at the bottom, and the scheme is more progressive if a greater number of agents locate at the top. I also highlight how the impact of the skill distribution is affected by the form of the social welfare function and the utility function. The findings using empirical skill distributions suggest that the results are sensitive to the type of statistical estimator used to estimate the skill distribution.

Keywords: Income taxation, Simulations, Skill distributions.

JEL Classification: H21; J24.

*I am most grateful to my supervisors Spencer Bastani and Thomas Giebe. I thank Jukka Pirttilä, Sebastian Köhne, Hans Grönqvist, Kaisa Kotakorpi, Tomas Sjögren, Anna Sjögren, Magnus Carlsson, Tobias König, and seminar participants at the Linnaeus University, and the SNS job market 2021 for comments.

†Department of Economics and Statistics, Linnæus University. Email: dingquan.miao@lnu.se

1 Introduction

A well-known fact from optimal taxation theory is that optimal tax schemes depend on three components: the skill distribution of the population, the individual preferences over consumption and labor supply, and the social planner's taste for redistribution (Mirrlees, 1971). Previous research applied various combinations of model specifications in order to compute optimal marginal tax rates (see Table 1). Very few of those focus on the role of skill distributions in shaping optimal tax schemes (Kanbur and Tuomala, 1994; Mankiw et al., 2009). Compared to the social welfare functions and the individual utility functions, the skill distributions are less subjective and easier to quantify. What patterns of optimal tax schemes can be inferred from different skill distributions? What drives the pattern of optimal marginal tax rates, the subjective model specifications, or the underlying skill distributions? In this paper, I analyze the role of the skill distribution in shaping optimal tax schemes.

In the first part of the paper, I conduct extensive simulation studies covering five different types of theoretical skill distributions. I carefully simulate optimal tax schedules with many commonly used specifications. The skill distributions differ in three ways: whether there is a mass point of unskilled workers at the bottom, whether the distribution has a thick tail at the top and the general dispersion of the distribution. I illustrate these differences with four groups of skill distributions: a log-normal distribution, a log-normal distribution with a mass point at the bottom, a log-normal distribution with a Pareto tail (henceforth, log-normal-Pareto distribution), and a log-normal distribution with a mass point at the bottom and a Pareto tail.

In the second part of the paper, I use empirical skill distributions and calculate optimal income tax schedules for 14 OECD countries, using wages as proxies for skills.[1] The 14 countries are: Austria, Czech Republic, Estonia, Finland, Germany, Greece, Ireland, Japan, Luxembourg, Netherlands, Sweden, Switzerland, UK, and the US. For Sweden, I use large-scale survey data collected by Statistics Sweden. For the 13 other countries, I use data from the Luxembourg Income Study.[2]

[1] It is common to use wage as a proxy for skill in the literature on optimal income taxation.

[2] I utilize the large-scale wage survey data from Statistics Sweden, as the data provides more observations than the Luxembourg Income Study Database (see Section 4.1).

There are three findings from the first part of the study. First, the impact of the skill distribution on optimal marginal tax rates depends crucially on the choice of social welfare function in the cases where I attach a mass point at the bottom or a Pareto tail (or both) to the log-normal distribution. Moreover, whether there are income effects on labor supply is irrelevant to the pattern of optimal tax schemes if the government is extremely inequality averse. Second, the larger the number of unskilled workers at the bottom, the lower are the optimal tax rates for this group, a result which is valid across a range of different model specifications. Third, the thicker the Pareto tail, the more progressive (or less regressive if applying a max-min type social objective function) are the optimal tax schemes. Attaching a Pareto tail affects not only the optimal tax schemes beyond the cutoff points (i.e., the points where the Pareto tail is attached), but also at lower skill levels. The earlier the Pareto tail enters the log-normal skill distribution, the less progressive the optimal tax schemes are before the cutoff point.

There are three findings from the second part of the study. First, the patterns of optimal tax rates are flatter for Luxembourg, Switzerland, and the US. This finding can be attributed to a higher dispersion of the fitted log-normal distribution for these countries as compared to other countries in the sample. Second, attaching a Pareto tail alters the pattern of the optimal tax rates after the cutoff point, and a discontinuity in the optimal marginal tax rate arises at the cutoff point. The earlier the Pareto tail enters the fitted wage distributions, the more progressive are the optimal tax schemes. Third, based on the evidence from Sweden, it appears that the Pareto tail fits the wage distribution well for high-income earners, and optimal marginal tax rates of fitted log-normal-Pareto and non-parametric wage distributions become more similar for high-income earners.

This paper contributes to the literature in several ways. First, to the best of my knowledge, this is the first paper to systematically simulate and analyze the role of the skill distribution in shaping optimal tax schemes. I provide graphical results that span the entire skill distribution, revealing the general pattern of optimal marginal tax rates. Second, this paper exploits empirically calibrated models on wage distributions across different OECD countries. I find the simulated optimal tax schemes are sensitive to how the empirical wage distributions are estimated, e.g., whether parametric or non-parametric estimators are used. I suggest future simulation exercises should be careful about the estimated skill distributions.

This paper builds on Mirrlees' optimal nonlinear income tax model and provides additional numerical results for the pattern of optimal tax schemes. In Table 1, I provide a summary of previous numerical studies of optimal tax rates. For example, some show that optimal marginal tax rates decrease in skill for the vast majority of the population (Mirrlees, 1971; Tuomala, 1984). Some show an increasing pattern of optimal marginal tax rates along the income distribution (Kanbur and Tuomala, 1994; Tuomala, 2010). More recent studies present a U-shaped pattern of optimal marginal tax rates (Saez, 2001; Mankiw et al., 2009; Bastani, 2015). Some of these studies are related to this paper but focus on different model specifications (Tuomala, 1984; Dahan and Strawczynski, 2000; Tuomala, 2010). Other studies contribute to the theory or simulation methods (Saez, 2001; Bastani, 2015). Overall, very few studies analyze the role of the skill distribution on the shape of optimal tax schemes. An exception is Kanbur and Tuomala (1994) who show that the choice of the variance of the skill distribution can determine whether the pattern of optimal marginal tax rates is decreasing or increasing. Another exception is Mankiw et al. (2009), who indicate that the shape of optimal marginal income tax rates can be sensitive to the thickness of the right tail of the skill distribution, e.g., a U-shaped pattern of optimal marginal tax rates emerges when using a Pareto distribution to fit higher wages whereas optimal marginal tax rates are monotonically decreasing otherwise.

The rest of the paper proceeds as follows. Section 2 introduces the discrete optimal income tax model. Section 3 simulates and analyzes the optimal tax schemes for different theoretical skill distributions. Section 4 provides a simulation analysis based on empirical calibrations for 14 OECD countries. Finally, Section 5 concludes.

2 The discrete optimal income tax model

This paper uses a discrete type version of the Mirrlees (1971) optimal nonlinear income tax model. The problem of the government is to maximize social welfare which aggregates individual utilities, subject to the social planner's budget constraint, and a set of incentive compatibility constraints. This

Table 1: *Previous simulation studies of optimal income taxation*

Author (date) (1)	Skill distribution (2)	Utility function (3)	Social welfare function (4)	Shape of the optimal MTR (5)
Mirrlees (1971)	Log-normal $(\mu, \sigma) = (0.4, 0.39)$ $(\mu, \sigma) = (0.61, 1)$	Strictly concave in consumption	Strictly concave	Optimal MTR decreases for the majority of population
Tuomala (1984)	Log-normal $(\mu, \sigma) = (-1, 0.39)$ $(\mu, \sigma) = (-1, 1)$	Strictly concave in consumption	Utilitarian/ Strictly concave/ Max-min	Optimal MTR decreases with income
Kanbur and Tuomala (1994)	Log-normal $(\mu, \sigma) = (0.4, 0.39)$	Strictly concave in consumption	Utilitarian/ Strictly concave/ Max-min	Optimal MTR decreases with income
	Log-normal $(\mu, \sigma) = (0.4, 0.7)$	Strictly concave in consumption	Max-min	Optimal MTR decreases with income
	Log-normal $(\mu, \sigma) = (0.4, 1)$	Strictly concave in consumption	Utilitarian	Optimal MTR increases for the majority of population
Diamond (1998)	Pareto $\alpha = (0.5, 1.5, 5)$	Linear in consumption	Strictly concave	A U-Shaped optimal MTR scheme
Dahan and Strawczynski (2000)	Pareto $\alpha = 1.5$	Strictly concave in consumption	Strictly concave	Optimal MTR decreases at higher level of incomes
	Log-normal $(\mu, \sigma) = (-1, 0.39)$	A quasi-linear preference (linearity in consumption)	Strictly concave	Optimal MTR increases at higher level of incomes
Saez (2001)	Empirical distribution	Strictly concave in consumption	Utilitarian/ Max-min	A U-Shaped optimal MTR scheme
		A quasi-linear preference (linearity in consumption)	Strictly concave/ Max-min	
Mankiw et al. (2009)	Log-normal with a mass point at the bottom $(\mu, \sigma) = (2.757, 0.5611)$, $b = 0.05$ $(\mu, \sigma) = (2.721, 0.4880)$, $b = 0.05$	Strictly concave in consumption	Utilitarian	Optimal MTR decreases with income
	Log-normal with a mass point at the bottom and a Pareto tail $(\mu, \sigma) = (2.757, 0.5611)$, $b = 0.05$, $\alpha = 2$ $(\mu, \sigma) = (2.721, 0.4880)$, $b = 0.05$, $\alpha = 2$	Strictly concave in consumption	Utilitarian	A U-Shaped optimal MTR scheme
Tuomala (2010)	Log-normal $(\mu, \sigma) = (0.4, 0.39)$ $(\mu, \sigma) = (0.4, 0.5)$ $(\mu, \sigma) = (0.4, 0.7)$	Strictly concave in consumption (with bounded consumption)	Strictly concave	Optimal MTR increases with income
Bastani (2015)	Log-normal with a mass point at the bottom and a Pareto tail (Discrete types of agents) $(\mu, \sigma) = (2.757, 0.5611)$, $b = 0.05$, $\alpha = 2$	Strictly concave in consumption	Utilitarian	A U-Shaped optimal MTR scheme

The parameters μ and σ are the mean and standard deviation of the natural logarithm of a log-normal distribution. The notation b is the mass point at the bottom of the distributions, and α is the shape parameter of a Pareto distribution.

analysis builds on a labor supply model which assumes that workers only differ in skill levels, which are here proxied by market wage rates. (Stiglitz, 1982; Diamond, 1998; Mankiw et al., 2009).

I assume that there is a set of distinct types of agents, $S = \{1, 2, \ldots, N\}$. The skill for type $i \in S$ is n_i, and $n_i < n_j$ if $i < j$. As is customary in the literature, the tax schedule is defined as a set of consumption-income pairs (c_i, y_i), $i = 1, ..., N$, where the tax schedule is implicitly defined as $T(y_i) = y_i - c_i$. A type i agent earns before-tax income y_i and consumes after-tax income c_i and gains the payoff $u(c_i, y_i, n_i)$.

It is convenient to define the utility function as $u^i(c_i, y_i) := u(c_i, y_i, n_i)$. The utility function is separable and given by $u^i(c_i, y_i) = x(c_i) - v(y_i/n_i)$, with $x_c > 0$ and $v_{y/n} > 0$. I will consider both the quasi-linear case which arises when $x(c) = c$ and the case where the utility from consumption is strictly concave, namely, $x_{cc} < 0$.[3] The quasi-linear case implies that there is no income effect on labor supply. When the utility from consumption is strictly concave, the value of redistribution typically increases due to diminishing marginal utility of consumption. Agents decide on their labor supply given the tax scheme set by the government. Individual labor supply is expressed in terms of hours of work, and is given by $l_i(n_i) = y_i/n_i$. Thus, choosing a pre-tax income level y_i is equivalent to choosing labor supply.

The government aims to maximize social welfare represented by the social welfare function

$$W(u_1, u_2, \ldots, u_N) = \sum_i G(u^i(c_i, y_i))\pi_i, \tag{1}$$

where π_i is the fraction of type i agents. The social welfare function reflects the government's taste for redistribution, and $G(u^i)$ is an increasing and concave function of u^i. The government maximizes social welfare W subject to the revenue constraint

$$\sum_i T(y_i)\pi_i \leq E, \tag{2}$$

where E is the required revenue for public expenditure, and $T(y_i) = y_i - c_i$ is the tax revenue collected from type i agent. The exogenous required revenue

[3] Boadway et al. (2000) studies the properties of the optimal income tax when the utility function is linear in leisure. Tuomala (2010) simulates the marginal tax rate structure by using a quadratic utility of consumption (i.e., an upper bound for consumption).

depends on the total public expenditure and the non-tax revenue collected by the government. If $E = 0$, the income tax is only used for redistribution. Governments cannot observe individuals' productivities and can only use income as the tax base, $T(y_i)$. In this second best problem, the following incentive compatibility constraints are required

$$\forall i, j \in S : u^i(c_i, y_i) \geq u^i(c_j, y_j). \tag{3}$$

These constraints prevent individuals from choosing others' consumption-income pairs, such that no one deviates from his/her true productivity. People with higher incomes pay higher taxes from a redistributive point of view. However, these taxes cannot be too high as this would drive high-income earners away. The redistribution of income is therefore restricted by the incentive compatibility constraints.

Following Hellwig (2007), I consider the weakly relaxed income tax problem where the constraints in (3) are replaced by the necessary and sufficient conditions:

$$\forall i \in \{1, 2, \ldots, N-1\} : u^{i+1}(c_{i+1}, y_{i+1}) \geq u^{i+1}(c_i, y_i), \text{ and} \tag{4}$$

$$\forall i \in \{1, 2, \ldots, N-1\} : c_{i+1} \geq c_i. \tag{5}$$

Maximizing social welfare (1) subject to the revenue constraint (2) and the downward incentive compatibility constraints (4) and the consumption monotonicity constraints (5), we can obtain an implicit formula for optimal marginal tax rates

$$T'(y_i) = \underbrace{\frac{\lambda_i u_{c_i}^{i+1}}{\phi \pi_i}}_{\text{A}} \underbrace{\left(\theta(c_i, y_i, n_i) - \theta(c_i, y_i, n_{i+1})\right)}_{\text{B}}. \tag{6}$$

where λ_i is the multiplier on the ith downward incentive compatibility constraint in (4) and ϕ is the multiplier on the government budget constraint. The term $\theta(c_i, y_i, n_i)$ is defined as the marginal rate of substitution of labor supply for consumption for an agent of skill n_i at the point (y_i, c_i) of an indifference curve. The marginal utility of consumption for skill type $i + 1$ choosing the consumption-income bundle intended for type i is $u_{c_i}^{i+1}$, and

$u^{i+1}_{c_i} = u^{i}_{c_i}$ when the utility function is separable. It is assumed that the consumption monotonicity constraints (5) are not binding. The derivation of the optimal tax formula is sketched in Appendix A. The equation (6) indicates that optimal tax rate for type i agent increases in the marginal utility of consumption and decreases in the population mass of that agent. In addition, the greater the difference in MRS and therefore in ability, the higher the tax rate for the agent. Three elements on the right-hand side of equation (6) determine optimal tax rates: social welfare weights (A), population mass of skill distribution (A), and marginal rate of substitution of labor supply for consumption (B).

3 Part I: Numerical simulations for different skill distributions

3.1 Methodology

The optimal tax schemes depend on three elements: the shape of the skill distribution, individuals' utility over consumption and labor supply, and the attitudes of the government towards inequality. My simulation exercise considers several of the most commonly used model specifications to study the role of the skill distribution in shaping optimal tax schedules. Table 2 describes the parameters in the simulation. As already mentioned, I consider two types of utility functions. In the first, utility is strictly concave in consumption:

$$u(c,l) = \log c - \frac{\eta}{1+1/e} l^{1+1/e}. \tag{7}$$

The second utility function is quasi-linear in consumption and is given by:

$$u(c,l) = c - \frac{\eta}{1+1/e} l^{1+1/e}. \tag{8}$$

The compensated elasticity of labor supply with respect to the wage rate is in both cases $e = 0.3$, which is in line with the recommended elasticity in the labor supply literature (Chetty et al., 2011).

Table 2: Parameters in the simulation

Parameter	Value	Description	
	0	Utilitarian	
β	0.2	Intermediate	SWF parameter
	∞	Max-min	
η	80.32	Scale parameter for the calibration of labor supply	Utility function parameters
e	0.3	Compensated elasticity of labor supply	
μ	3	Location measure of log-normal distribution (log of median skill)	
σ	0.2, 0.4, 0.6, 0.8, 1	Dispersion measure of log-normal distribution	
b	5, 10, 15, 20	Mass point at the bottom (in percent)	Skill distribution parameters
α	1, 2, 3, 4	Shape parameter of Pareto distribution	
x_m	21.65, 25.3, 30.1, 37.4	Cutoff points of Pareto distribution	
N	1000, 10000	Number of skill types (for Part I and Part II respectively)	

The simulations employ three types of social welfare functions: a utilitarian social welfare function, an increasing and strictly concave function of utility, and a max-min social welfare function. These cases can be captured by weighting the individual utilities by the following function

$$G(u) = -\frac{1}{\beta}e^{-\beta u}, \tag{9}$$

where β reflects the degree of inequality aversion.[4] As β increases, the social welfare function becomes more concave and the government becomes more inequality averse. The extreme inequality aversion case, i.e., the max-min welfare function, is obtained when $\beta = \infty$. The utilitarian social welfare function is obtained when $\beta = 0$.

In my simulations, I use four different combinations of model specifications based on the type of utility function and the type of social welfare function. The case where the government has no incentive to redistribute income from high-skill to low-skill agents is excluded, that is, the combination of a utilitarian social objective and a quasi-linear utility function is not considered. I also exclude the case of a concave social welfare function and a concave utility function. This combination is, in fact, a concave transformation of a concave function (i.e., a concave utility function), and the choice

[4]This functional form is also used by Tuomala (1984, 2010).

of concavity is somewhat arbitrary. In Appendix B, I consider the case of a utilitarian social welfare function combined with different types of concave utility functions.

In this section, I consider four groups of skill distributions, referred to as Type I, Type II, Type III and Type IV summarized in Table 3.

Table 3: *Summary of simulation experiments*

Type of skill distribution		Considered parameter values
Type I:	Log-normal distributions (considering different dispersion parameters σ)	$\sigma = 0.2, 0.4, 0.6, 0.8, 1$
Type II:	Log-normal distributions (considering different mass points b at the bottom)	$b = 0\%, 5\%, 10\%, 15\%, 20\%$
Type III:	Log-normal-Pareto distributions (considering different shape parameters α)	$\alpha = 1, 2, 3, 4$
Type IV:	Log-normal-Pareto distributions (considering different cutoff points for the Pareto-distribution x_m)	$x_m = 21.65, 25.3, 30.1, 37.4$

All simulations employ the log-normal distribution parameters $(\mu, \sigma) = (3, 0.6)$ unless otherwise stated.

In Type I, I use a log-normal distribution and change the degree of dispersion as measured by the parameter σ. I consider the values $\sigma = 0.2, 0.4, 0.6, 0.8, 1$ and let μ be fixed and equal to 3.

In Type II, I consider skill distributions with a mass point of size b located at the bottom of the distribution. Notice that b corresponds to the b:th percentile of the skill distribution and is defined by the skill level $\underline{n}$ satisfying $P(n <= \underline{n}) = b$. I use the mid-point, $\underline{n}/2$, to represent the the skill range $[0, \underline{n}]$. According to World Health Organization data, the share of people receiving social/disability benefits is around 6% in Nordic countries.[5] Mankiw et al. (2009) attached a 5% mass point at the bottom to reflect the share of disabled workers in the US. In the simulations, I compare five different cases, $b = 0\%, 5\%, 10\%, 15\%, 20\%$.

In Type III and Type IV, I consider different scenarios regarding the right tail of the skill distributions. The probability density function of a Pareto

[5] World Health Organization, 1970-2019, https://gateway.euro.who.int/en/indicators/hfa_415-2721-number-of-people-receiving-socialdisability-benefits/visualizations/#id=19406&tab=table

distribution is given by

$$f_P(x) = \frac{\alpha x_m^\alpha}{x^{\alpha+1}} \quad (x_m, \alpha \in \mathbb{R}_{>0},\ x \in [x_m, \infty)), \tag{10}$$

where α is the shape parameter and x_m is the cutoff point, i.e., the entry point of the Pareto tail. The shape parameter reflects the thickness of the tail, and the smaller the shape parameter, the thicker the Pareto tail. I consider the values $\alpha = 1, 2, 3, 4$. This is in line with the earlier literature that has used values ranging from 0.5 to 5. For example, Mankiw et al. (2009) and Bastani (2015) attach a Pareto parameter $\alpha = 2$ in their parametric wage distributions, and Diamond (1998) uses $0.5, 1.5, 5$.[6] I choose the percentiles of the log-normal distribution $p = 55\%, 65\%, 75\%, 85\%$ as the cutoff points, and the corresponding values of the cutoff points $x_m = 21.65, 25.3, 30.1, 37.4$.

In Type III, I use the value of Pareto parameter $\alpha = 2$ as the benchmark, and I compare the impact of different Pareto parameters (at the same cutoff point) on the pattern of the tax schemes. In Type IV, the benchmark case is a log-normal-Pareto distribution with a cutoff of the Pareto tail at $x_m = 30.1$, and I compare the effect of different cutoff points of the attached Pareto tail (with the same Pareto parameter) to the shape of the marginal tax rates. The baseline parameters are shown in Table 3. In the scenarios Type II, III and IV, the log-normal parameters are fixed at the baseline values $(\mu, \sigma) = (3, 0.6)$.

3.2 Results

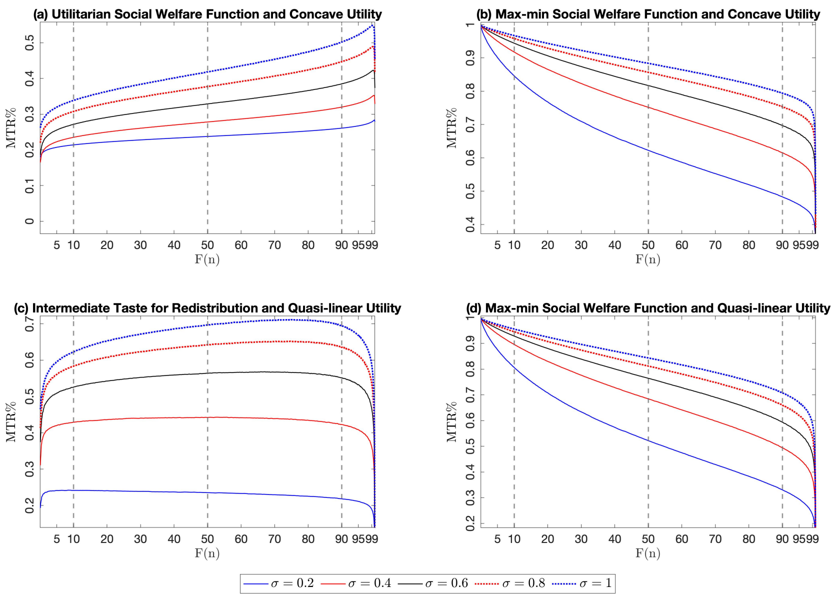
In Figure 1, I summarize the simulation results obtained using Type I skill distributions. Consistent with the findings of previous work, the extent of inherent inequality can dramatically change the pattern of optimal marginal tax rates, see Kanbur and Tuomala (1994). Figure 1(a) shows that optimal tax rates are everywhere increasing when using a concave utility function and a utilitarian social welfare function. The shape of optimal marginal tax rates along the percentiles of the skill distribution tends to be concave if there is no income effect, leading to a decreasing pattern of tax rates for approximately the top 10 percentile of the population, see Figure 1(c). The concave utility function implies declining marginal utility of consumption,

[6] Bastani and Lundberg (2017) report the value of Pareto parameter is around 3 in 1970-2015 in Sweden.

which leads to more redistributive tax schemes (i.e., the second derivative of the tax function is higher everywhere).

The results for the max-min type of social welfare function are in line with the result from Kanbur and Tuomala (1994), in which the optimal tax rates are everywhere decreasing. Figure 1(b) and 1(d) show that the general results for the concave and quasi-linear types of utility function are very similar. The optimal tax scheme decreases faster when there is no income effect. When eliminating one incentive for redistribution, i.e., the declining marginal utility of consumption, optimal tax rates appear to be lower for higher-skilled agents.

Figure 1: *Optimal marginal tax rates using log-normal distributions with different dispersion (Type I)*



In Figure 2, I report optimal tax rates for Type II skill distributions. We observe a universal result across the different panels, which is, the larger the mass point, the lower the optimal tax rates for unskilled agents. Also, optimal marginal tax rates tend to converge for higher skill levels, that is, the size of the mass point does not influence optimal marginal tax rates at high skill levels. The results are invariant to the thickness of the right tail, see Figure A6.[7]

Figure 2: *Optimal marginal tax rates using log-normal distributions with different mass points at the bottom (Type II)*

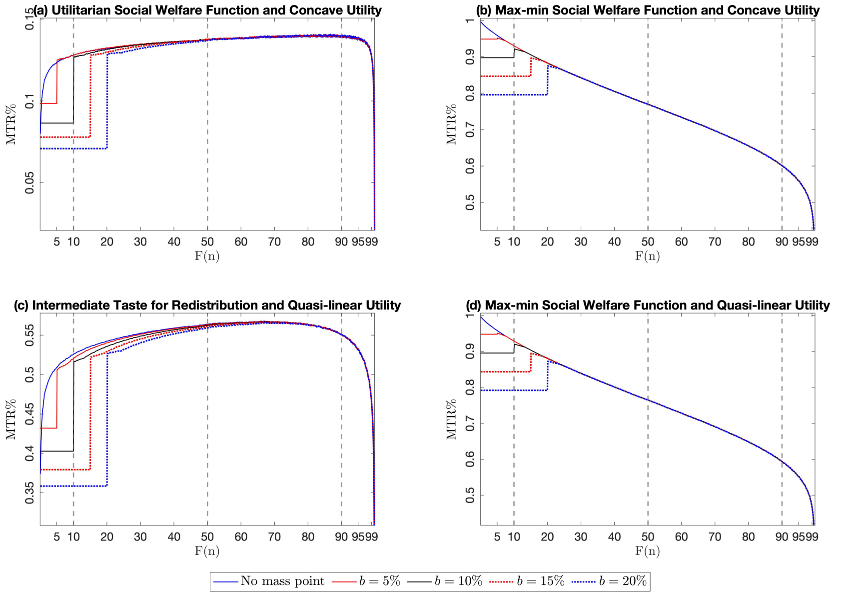
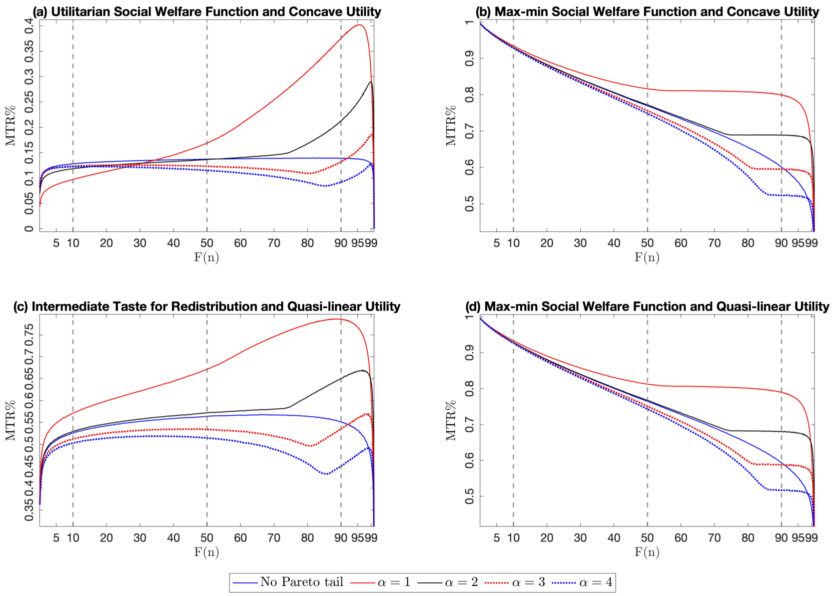


Figure 3 illustrates optimal tax rates for Type III skill distributions. Figure 3(a) and 3(c) show that optimal marginal tax rates are everywhere increasing when attaching a thicker Pareto tail to the log-normal distribution, e.g., $\alpha = 1,2$. The optimal marginal tax rates increase sharply with a thicker Pareto tail, see Figure 3(a). Attaching a Pareto tail affects not only the optimal tax schemes beyond the cutoff points, but also at lower skill levels. Similar to the results from Type II distributions, both the pattern and the level of optimal marginal tax rates are invariant to the choice of utility function when the government is extremely inequality averse.

[7]In appendix Figure A6, I present an alternative version of the Type II skill distribution where I consider a log-normal-Pareto distribution with a mass point at the bottom of the distribution.

Figure 3: *Optimal marginal tax rates using log-normal-Pareto distributions with different shape parameters (Type III)*

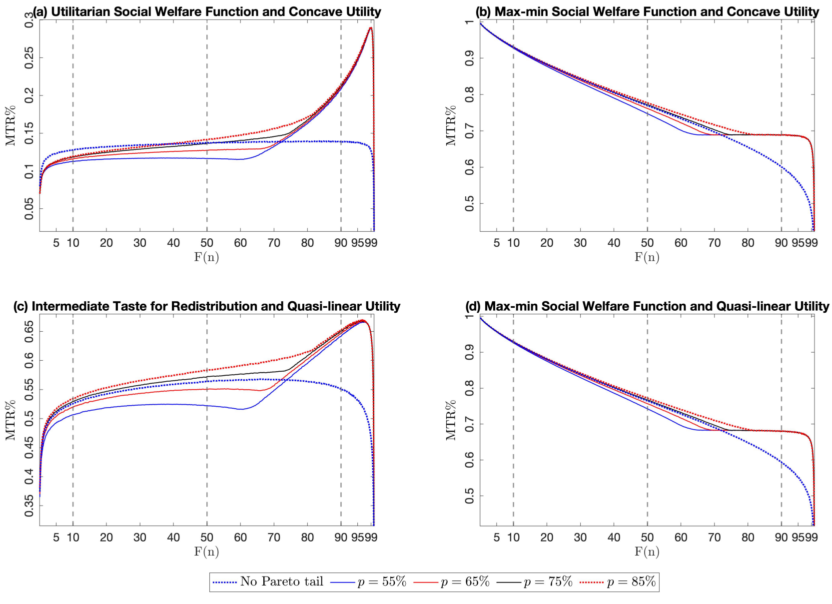


Note: The cutoff point of the Pareto-distribution is fixed at the 75th percentile.

Figure 4 presents simulated optimal tax rates for Type IV skill distributions considering the following percentile cutoff points for the Pareto tail: no cutoff point, p55, p65, p75, p85. In line with the results for Type III skill distributions, optimal marginal tax rates increase for upper-middle skill agents. Attaching a Pareto tail affects not only the optimal tax schemes after the cutoff points, but also the (global) tax schemes before the cutoff points. In particular, the smaller the percentile of the cutoff point, the lower the optimal tax rates before the cutoff point. Moreover, the optimal tax schemes tend to converge to one uniform tax scheme when shape parameters of the Pareto tail are the same across different skill distributions.

Similar to the Type II and Type III skill distributions, the role of the income effect depends on the government's taste for redistribution: the income effect is irrelevant to the pattern of optimal tax schemes if the government has a max-min type of social welfare function.

Figure 4: Optimal marginal tax rates using log-normal-Pareto distributions with different cutoff points for the Pareto-distribution (Type IV)



Note: The shape parameter is fixed at $\alpha = 2$.

4 Part II: Optimal tax rates for 14 OECD countries

In the previous section, I simulated optimal marginal tax rates for different *theoretical* skill distributions. In this section, I conduct simulations using *empirical* skill distributions. This exercise aims to learn about differences in optimal taxes across countries and gain further insights into how the shape of skill distribution affects optimal taxes using data from actual economies. I also explore the sensitivity of optimal marginal tax rates to parametric functional forms and non-parametric distributions.

I employ a simple calibrated model to simulate optimal tax schemes for 14 OECD countries. I focus on the quasi-linear utility function specified in (8). The elasticity of labor supply is as before $e = 0.3$. The government's taste for redistribution is captured by the concave social welfare function $G(u) = \log u$. I use gross hourly wages as proxies for skills, as is common in the literature on optimal income taxation.[8] I apply parametric (log-normal and log-normal-Pareto distributions) and non-parametric methods to fit the empirical wage distributions.

4.1 Data

This paper employs micro-level cross-country data from 2013.[9] For Sweden, I use monthly wage data provided by Statistics Sweden.[10] The survey data covers 50% of private sector employees and all employees in the public sector. For the other 13 countries, I use hourly wages from the Luxembourg Income Study Database. The wages in the Luxembourg Income Study are collected by different government agencies. The data only covers workers who have a formal contract job, i.e., people with zero wage earnings and the non-working population are excluded. The currency is in 2013 Euro, and other currencies are converted to Euro using the average exchange rate in 2013. Summary statistics are reported in Table A1.

4.2 Constructions of wage distributions

Parametric wage distributions for 14 OECD countries

Due to the limitation of the sample size in the Luxembourg Income Study Database, I fit parametric distributions to the wage samples of the 14 OECD countries. The log-normal estimates are shown in Table 4. In Figure 5, I plot the fitted log-normal wage distributions. I categorize countries into three groups based on the dispersion of the empirical wage distributions measured by the variance of the fitted log-normal distribution, see Table 4.

There are two things that need to be considered when attaching a Pareto tail to a log-normal distribution. First, the Pareto tail should fit the observations well. The fit of the right tails of the distributions are depicted in Figure A7. The shape parameter α is estimated by using Maximum Likelihood estimation. The goodness of fit of log-log regressions of the survival functions

[8] It is possible to use different proxies for skill, e.g., the Programme for the International Assessment of Adult Competencies (PIAAC) provides an alternative measure for adult skills, see https://www.oecd.org/skills/piaac/.

[9] Without considerations of wage dynamics, the 2013 data is used due to limitations of the available datasets.

[10] For Sweden, I divide the monthly wage by the average monthly hours of work to obtain a measure of the hourly wage rate.

Table 4: *Fitted Log-normal and Log-normal-Pareto Distributions, 2013*

Country	p	x_m	α	$\triangle f_{ln}(x)$	adjusted R^2	Log-normal (μ, σ)	Variance $(e^{\sigma^2}-1)e^{2\mu+\sigma^2}$	Dispersion
Czech Republic	0.8	6.4	3.29	0.060	0.9876	(1.54, 0.44)	5.64	
Estonia	0.95	11.05	3.25	0.008	0.986	(1.46, 0.54)	8.40	Small
Greece	0.9	12.3	3.57	0.008	0.9756	(1.90, 0.46)	13.02	
Austria	0.85	21.7	3.59	0.001	0.9973	(2.63, 0.48)	62.79	
Finland	0.75	24.6	3.02	0.030	0.997	(2.96, 0.44)	96.54	
UK	0.8	20.26	2.62	0.064	0.9946	(2.55, 0.61)	107.27	
Germany	0.9	26.9	3.71	0.159	0.969	(2.55, 0.65)	131.58	Medium
Japan	0.7	13.7	1.62	0.077	0.963	(2.39, 0.72)	135.88	
Ireland	0.85	30.16	3.03	0.066	0.9909	(2.84, 0.55)	140.03	
Netherlands	0.8	30.1	2.96	0.051	0.9903	(3.06, 0.51)	175.26	
Sweden	0.7	24.95	1.96	0.214	0.9838	(3.15, 0.53)	233.90	
United States	0.85	27.52	2.36	0.064	0.9875	(2.59, 0.76)	247.49	Large
Luxembourg	0.95	56.4	3.69	0.005	0.9924	(3.10, 0.57)	261.78	
Switzerland	0.85	49.39	3.05	0.135	0.9957	(3.33, 0.73)	936.11	

x_m is the cutoff point that a Pareto tail enters the Log-normal distributions, and p is the corresponding percentile of the Log-normal distributions. α is the shape parameter of the Pareto tail. $\triangle f_{ln}(x)$ is defined as the relative difference between the log-normal and Pareto distributions at the cutoff point x_m. Adjusted R^2 is obtained from the log-log regressions of the survival functions of the Pareto distributions.

of the Pareto distributions are shown in Table 4. Second, the log-normal-Pareto distribution is expected to be smooth and continuous. The difference between the log-normal distribution and the Pareto tail at the cutoff point x_m needs to be small. I measure the difference in relative terms, that is,

$$\triangle f_{ln}(x) = \left| \frac{f_P(x = x_m) - f_{ln}(x = x_m)}{f_{ln}(x = x_m)} \right| \quad \text{where } f_P(x = x_m) = \frac{\alpha}{x_m}(1 - p), \quad (11)$$

where p is the corresponding percentile of the threshold x_m. I loop over a set of discrete percentiles (70th, 75th, 80th, 85th, 90th, 95th, 99th) and determine the one that gives the smallest difference between log-normal distribution and the Pareto tail (in relative terms) at the cutoff point x_m. The matched Pareto tails are displayed in Table 4.

Non-parametric wage distribution for Sweden

Given that the Swedish wage data contains over two million observations, I am able to compute an optimal tax scheme using a finer non-parametric wage distribution. In particular, I divide the empirical wage distribution into $10,000$ equally sized bins (each containing the same share of the total population). I use the median value of each bin to represent the wage level of that bin. I then compare optimal tax schemes that result from different parametric and non-parametric distributions.

4.3 Results

I now present simulated optimal income tax schedules for the empirical wage distributions. To be able to easier compare the results for the different countries, as already mentioned, I group and rank them according to the variance of the fitted log-normal distributions, which is a measure of pre-tax inequality. The first group consists of Austria, Czech Republic, Estonia, and Greece; the second group consists of Finland, Germany, Ireland, Japan, and the UK; the third group consists of Luxembourg, Netherlands, Sweden, Switzerland, and the US.[11]

[11]It is worth mentioning that some countries are not ranked as we expected, e.g., Sweden is in the high inequality group. There might be two possible reasons for this ranking: first, the variance measures the pre-tax inequality with no taxes and transfers. Second, the estimates are data-driven, and I access a different dataset for Sweden.

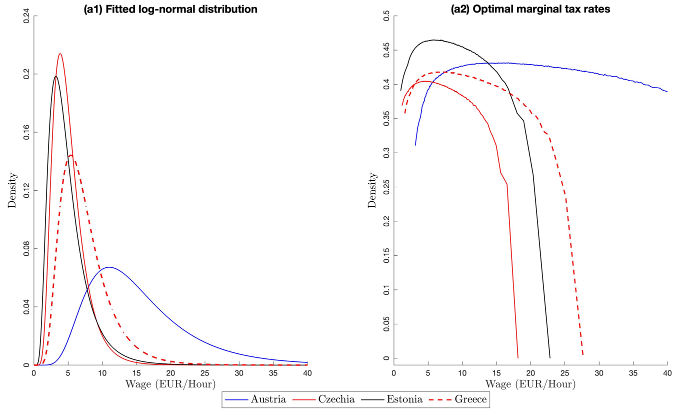
In Figure 5, I visualize fitted log-normal distributions and corresponding optimal marginal tax rates. The first thing to notice is that the pattern of optimal marginal tax rates is sensitive to the shape of the fitted log-normal distributions. I find that the patterns of optimal tax rates are flatter for the more dispersed log-normal distributions (i.e., distributions with large variance), for example, Austria has a flatter tax scheme than Estonia, see Figure 5(a). The marginal tax rates decrease faster for countries with more compressed wage distributions, for example, Czech Republic, Estonia, and Greece. Countries similar in dispersion share a similar pattern of optimal marginal tax rates, such as Sweden and Luxembourg, see Figure 5(c). The general pattern of optimal tax rates decreases for high-income earners (at least for income earners with wage levels above the cutoff point x_m) when applying log-normal distributions to fit wage distributions.

In Figure 6, I show optimal marginal tax rates using fitted log-normal-Pareto distributions. Attaching a Pareto tail alters the pattern of optimal tax rates beyond the cutoff point, and the optimal tax rate at the cutoff point becomes a kink point. We observe a higher tax ladder beyond the kink point for some countries, and it is more salient for Sweden, Japan, Finland, and the Netherlands. It appears that the smaller the percentile of the cutoff point, p, the larger the increase in optimal tax rates, see Table 4. The optimal marginal tax rates for individual countries are reported in Figure A8.

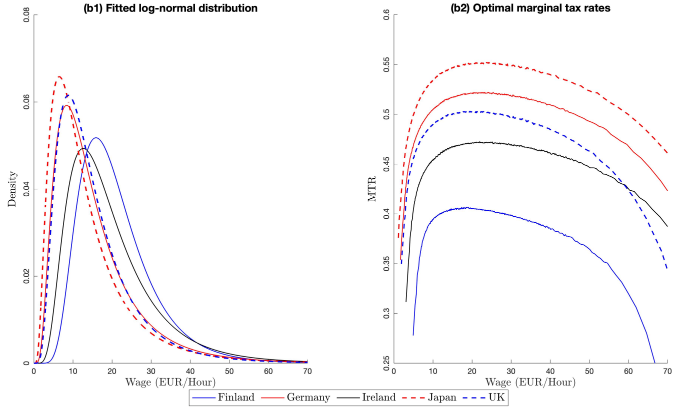
Figure 7 presents the simulated optimal tax rates resulting from the fitted parametric and fitted non-parametric wage distributions. The pattern and level of optimal tax rates are very sensitive to the estimated wage distributions. The log-normal distribution leads to a flatter shape of optimal tax scheme than optimal tax rates simulated using the empirical wage distribution. Based on the evidence from Sweden, it appears that the Pareto tail fits the wage distribution well for high-income earners, and optimal tax schemes of fitted log-normal-Pareto and non-parametric wage distributions move closer for high-income earners.

Figure 5: *Fitted log-normal distributions and optimal marginal tax rates*

(a) *Small dispersion*



(b) *Medium dispersion*



(c) *Large dispersion*

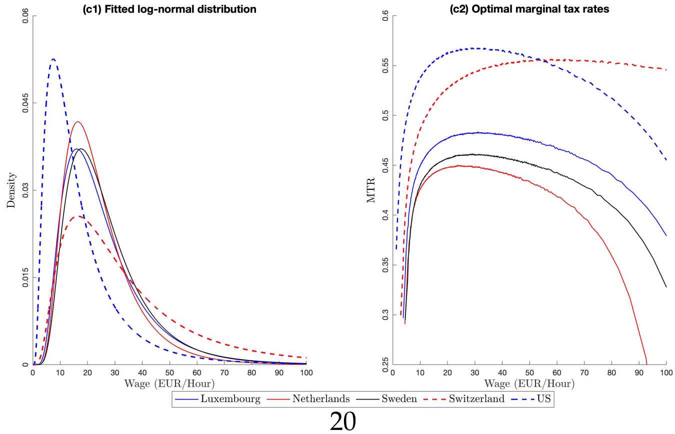
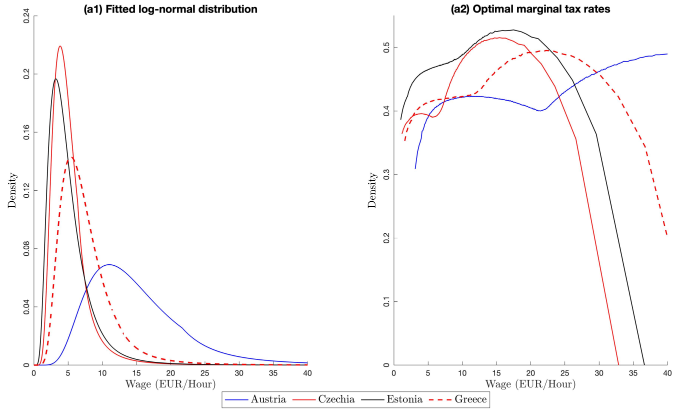
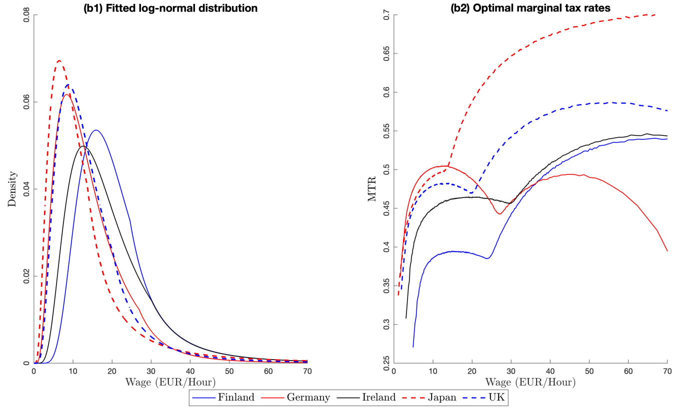


Figure 6: *Fitted log-normal-Pareto distributions and optimal marginal tax rates*

(a) *Small dispersion*



(b) *Medium dispersion*



(c) *Large dispersion*

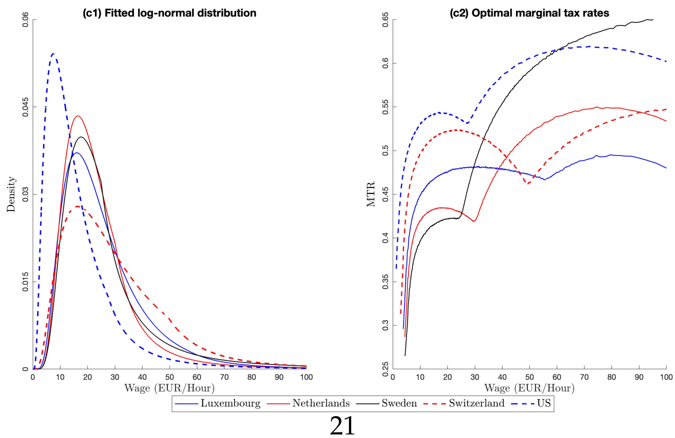
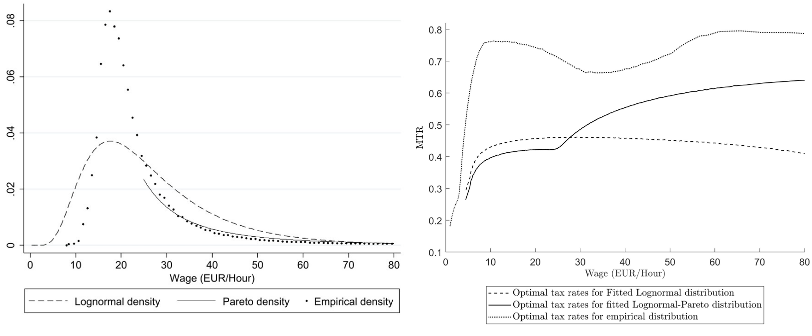


Figure 7: *Optimal marginal tax rates using fitted parametric and fitted non-parametric wage distributions*



5 Conclusions

The paper analyzes the role of the skill distribution in shaping optimal nonlinear income tax schemes. I use theoretical skill distributions as well as empirical skill distributions for 14 OECD countries. I find that a more dispersed log-normal skill distribution implies a more progressive optimal tax schedule. Optimal tax rates should be lower if a greater number of unskilled agents cluster at the bottom, and the scheme should be more progressive if a greater number of agents locate at the top. I also highlight how the impact of the skill distribution is affected by the form of the social welfare function and the utility function.

There are several findings from the analysis using empirical skill distributions. First, the pattern of optimal tax rates are flatter for Luxembourg, Switzerland, and the US. This finding can be attributed to a higher variance of the fitted log-normal distribution for these countries. Second, attaching a Pareto tail alters the pattern of optimal tax rates after the cutoff point, and a discontinuity in the optimal marginal tax rate arises at the cutoff point. The earlier the Pareto tail enters the fitted wage distribution, the more progressive are the optimal tax schemes. The results are sensitive to the type of statistical estimator used to estimate the skill distribution.

Governments can adjust tax schemes by comparing the dispersion of skill distributions in different years or locally across different regions. Since the optimal marginal tax rates can be sensitive to the fitted skill distributions (e.g., parametric and non-parametric distributions), this determines how much the policymaker can trust the result of a given optimal taxation model or simulation.

References

Bastani, S. (2015). Using the discrete model to derive optimal income tax rates. *FinanzArchiv: Public Finance Analysis 71*(1), 106–117.

Bastani, S. and J. Lundberg (2017). Political preferences for redistribution in sweden. *The Journal of Economic Inequality 15*(4), 345–367.

Boadway, R., K. Cuff, and M. Marchand (2000). Optimal income taxation with quasi-linear preferences revisited. *Journal of Public Economic Theory 2*(4), 435–460.

Chetty, R., A. Guren, D. Manoli, and A. Weber (2011). Are micro and macro labor supply elasticities consistent? a review of evidence on the intensive and extensive margins. *American Economic Review 101*(3), 471–75.

Dahan, M. and M. Strawczynski (2000). Optimal income taxation: An example with a u-shaped pattern of optimal marginal tax rates: Comment. *American Economic Review 90*(3), 681–686.

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

Hansen, L. P. and K. J. Singleton (1983). Stochastic consumption, risk aversion, and the temporal behavior of asset returns. *Journal of political economy 91*(2), 249–265.

Hellwig, M. F. (2007). A contribution to the theory of optimal utilitarian income taxation. *Journal of Public Economics 91*(7-8), 1449–1477.

Kanbur, R. and M. Tuomala (1994). Inherent inequality and the optimal graduation of marginal tax rates. *The Scandinavian Journal of Economics 96*(2), 275–282.

Kocherlakota, N. R. (1996). The equity premium: It's still a puzzle. *Journal of Economic literature 34*(1), 42–71.

LIS. Luxembourg Income Study (LIS) Database, http://www.lisdatacenter.org (multiple countries; May 2020 - June 2021), Luxembourg: LIS.

Mankiw, N. G., M. Weinzierl, and D. Yagan (2009). Optimal taxation in theory and practice. *Journal of Economic Perspectives 23*(4), 147–74.

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *The Review of Economic Studies 38*(2), 175–208.

Pålsson, A.-M. (1996). Does the degree of relative risk aversion vary with household characteristics? *Journal of economic psychology 17*(6), 771–787.

Saez, E. (2001). Using elasticities to derive optimal income tax rates. *The Review of Economic Studies 68*(1), 205–229.

Stiglitz, J. E. (1982). Self-selection and Pareto efficient taxation. *Journal of Public Economics 17*(2), 213–240.

Tuomala, M. (1984). On the optimal income taxation: Some further numerical results. *Journal of Public Economics 23*(3), 351–366.

Tuomala, M. (2010). On optimal non-linear income taxation: numerical results revisited. *International Tax and Public Finance 17*(3), 259–270.

Appendix

A Deriving the implicit optimal tax formula

The government problem can be stated as follows:

$$
\begin{aligned}
\max_{\{c_i,y_i\}_{i=1}^N} \quad & \sum_i G(u^i(c_i,y_i))\pi_i \\
\text{subject to} \quad & \sum_i (y_i - c_i)\pi_i \geq 0, \\
& u^{i+1}(c_{i+1},y_{i+1}) \geq u^{i+1}(c_i,y_i),\ \forall i \in \{1,2,\ldots,N-1\}, \\
& c_{i+1} \geq c_i,\ \forall i \in \{1,2,\ldots,N-1\}.
\end{aligned}
$$

The Lagrangian for this maximization problem can be expressed as

$$
\begin{aligned}
\mathcal{L} = & \sum_{i=1}^{N} G(u^i(c_i,y_i))\pi_i + \phi\sum_{i=1}^{N}(y_i - c_i)\pi_i \\
& + \sum_{i=1}^{N-1}\lambda_i(u^{i+1}(c_{i+1},y_{i+1}) - u^{i+1}(c_i,y_i)) + \sum_{i=1}^{N-1}\zeta_i(c_{i+1} - c_i).
\end{aligned} \tag{A1}
$$

Differentiating the Lagrangian with respect to c_i and y_i gives the first-order conditions

$$
\mathcal{L}_{c_i} = \frac{\partial G(\cdot)}{\partial u^i(c_i,y_i)}\frac{\partial u^i(c_i,y_i)}{\partial c_i}\pi_i - \phi\pi_i + \lambda_{i-1}\frac{\partial u^i(c_i,y_i)}{\partial c_i} - \lambda_i\frac{\partial u^{i+1}(c_i,y_i)}{\partial c_i} + \zeta_{i-1} - \zeta_i = 0, \tag{A2}
$$

$$
\mathcal{L}_{y_i} = \frac{\partial G(\cdot)}{\partial u^i(c_i,y_i)}\frac{\partial u^i(c_i,y_i)}{\partial y_i}\pi_i + \phi\pi_i + \lambda_{i-1}\frac{\partial u^i(c_i,y_i)}{\partial y_i} - \lambda_i\frac{\partial u^{i+1}(c_i,y_i)}{\partial y_i} = 0. \tag{A3}
$$

Suppose the consumption monotonicity constraints are not binding, that is, $\zeta_i = 0$ for $i \in \{1,2,\ldots,N\}$. Multiplying equation (A2) by $\frac{\partial u^i(c_i,y_i)/\partial y_i}{\partial u^i(c_i,y_i)/\partial c_i}$, we get

$$
\begin{aligned}
& \frac{\partial G(\cdot)}{\partial u^i(c_i,y_i)}\frac{\partial u^i(c_i,y_i)}{\partial y_i}\pi_i - \phi\pi_i\frac{\partial u^i(c_i,y_i)/\partial y_i}{\partial u^i(c_i,y_i)/\partial c_i} \\
& + \lambda_{i-1}\frac{\partial u^i(c_i,y_i)}{\partial y_i} - \lambda_i\frac{\partial u^{i+1}(c_i,y_i)}{\partial c_i}\frac{\partial u^i(c_i,y_i)/\partial y_i}{\partial u^i(c_i,y_i)/\partial c_i} = 0.
\end{aligned} \tag{A4}
$$

Subtracting equation (A3) from equation (A4), we obtain

$$-\phi\pi_i \frac{\partial u^i(c_i, y_i)/\partial y_i}{\partial u^i(c_i, y_i)/\partial c_i} - \phi\pi_i - \lambda_i \frac{\partial u^{i+1}(c_i, y_i)}{\partial c_i} \frac{\partial u^i(c_i, y_i)/\partial y_i}{\partial u^i(c_i, y_i)/\partial c_i} + \lambda_i \frac{\partial u^{i+1}(c_i, y_i)}{\partial y_i} = 0. \tag{A5}$$

Simplifying the equation above, we obtain

$$\phi\pi_i \Big(\frac{\partial u^i(c_i, y_i)/\partial y_i}{\partial u^i(c_i, y_i)/\partial c_i} + 1 \Big) = \lambda_i \frac{\partial u^{i+1}(c_i, y_i)}{\partial c_i} \Big(\frac{\partial u^{i+1}(c_i, y_i)/\partial y_i}{\partial u^{i+1}(c_i, y_i)/\partial c_i} - \frac{\partial u^i(c_i, y_i)/\partial y_i}{\partial u^i(c_i, y_i)/\partial c_i} \Big). \tag{A6}$$

Individuals face a tax regime and maximize their utility

$$\begin{aligned} \max_{\{c_i, y_i\}_{i \in S}} \quad & u^i(c_i, y_i) \\ \text{subject to} \quad & c_i \leq y_i - T(y_i), \end{aligned}$$

and the first-order condition gives us

$$\frac{\partial u^i(c_i, y_i)/\partial y_i}{\partial u^i(c_i, y_i)/\partial c_i} = -(1 - T'(y_i)). \tag{A7}$$

Inserting equation (A7) into (A6), we obtain

$$T'(y_i) = \frac{\lambda_i}{\phi\pi_i} \frac{\partial u^{i+1}(c_i, y_i)}{\partial c_i} \Big(\frac{\partial u^{i+1}(c_i, y_i)/\partial y_i}{\partial u^{i+1}(c_i, y_i)/\partial c_i} - \frac{\partial u^i(c_i, y_i)/\partial y_i}{\partial u^i(c_i, y_i)/\partial c_i} \Big). \tag{A8}$$

Defining the marginal rate of substitution for an agent of skill level n_i at the point (c_i, y_i) of an indifference curve as

$$\theta(c_i, y_i, n_i) := -\frac{\partial u^i(c_i, y_i)/\partial y_i}{\partial u^i(c_i, y_i)/\partial c_i}. \tag{A9}$$

The optimal tax formula is given by

$$T'(y_i) = \frac{\lambda_i \frac{\partial u^{i+1}(c_i, y_i)}{\partial c_i}}{\phi\pi_i} (\theta(c_i, y_i, n_i) - \theta(c_i, y_i, n_{i+1})). \tag{A10}$$

B Sensitivity analysis

B.1 The curvature of the individual utility function

As discussed in section 3.1, the combination of a concave social welfare function and a concave utility function leads to an arbitrary level of curvature of the social welfare function. How sensitive are the patterns of optimal tax schemes to the curvature of the individual utility function?

To examine this effect, I employ a utilitarian social welfare function and consider the following utility function:

$$u(c,l) = \frac{c^{1-\gamma}}{1-\gamma} - \frac{\eta}{1+1/e} l^{1+1/e}, \tag{B11}$$

where γ is the coefficient of the relative risk aversion in consumption. My sensitivity analysis considers three different coefficients of relative risk aversion in consumption, i.e., $\gamma = 0.9, 1, 1.1$. The benchmark value of γ is 1, in which case the utility function is expressed as equation (7) in the main text.[12] In Figure A1-A5, I report the results from this sensitivity analysis. It appears that the general patterns of optimal marginal tax rates are not very sensitivity to the value of γ.

[12]For example, Hansen and Singleton (1983) concludes the coeffficient of relative risk aversion is between 0 and 2 by using historical monthly data. Pålsson (1996) shows that the relative risk aversion on Swedish data is in the range of 2 and 4. Kocherlakota (1996) claims that a coefficient of the relative risk aversion above 10 reflects individual highly improbable behavior. Saez (2001) adopts a logarithmic form of consumption that corresponds to a coefficient of relative risk aversion of 1 in his simulations. Mankiw et al. (2009) uses 1.5 as the value of coefficient to simulate optimal marginal tax rates in US.

Figure A1: *Optimal marginal tax rates using different log-normal distributions (Type I, sensitivity analysis)*

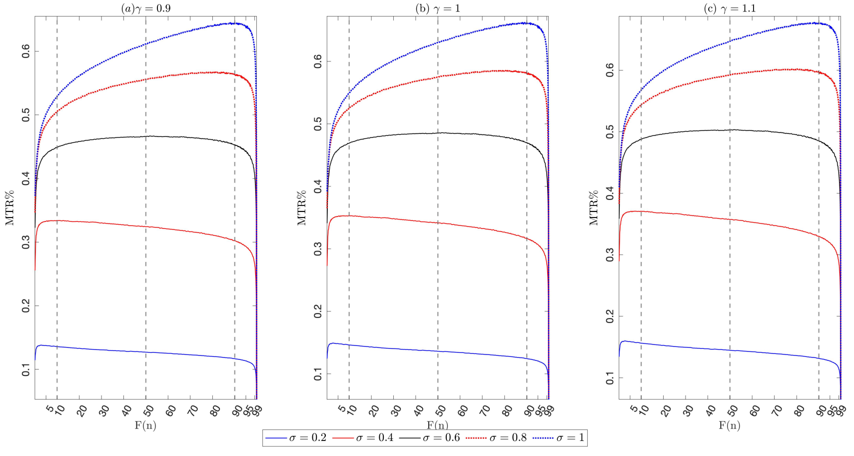


Figure A2: *Optimal marginal tax rates using a log-normal distribution considering different mass points at the bottom (Type II, sensitivity analysis)*

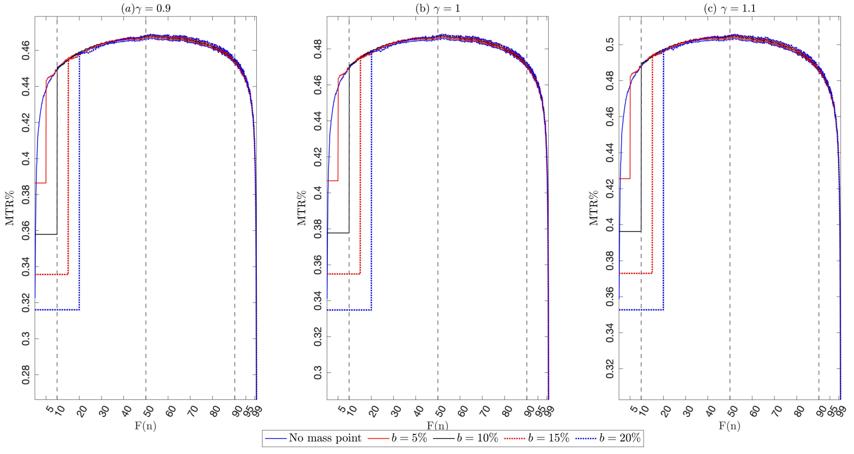
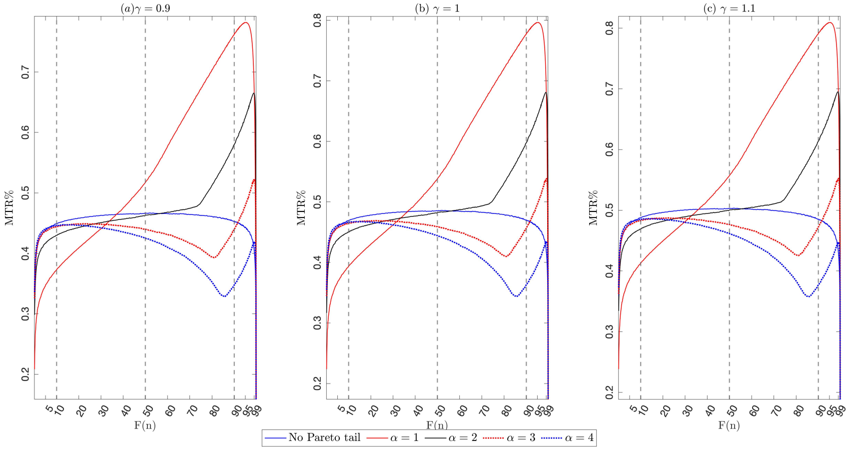
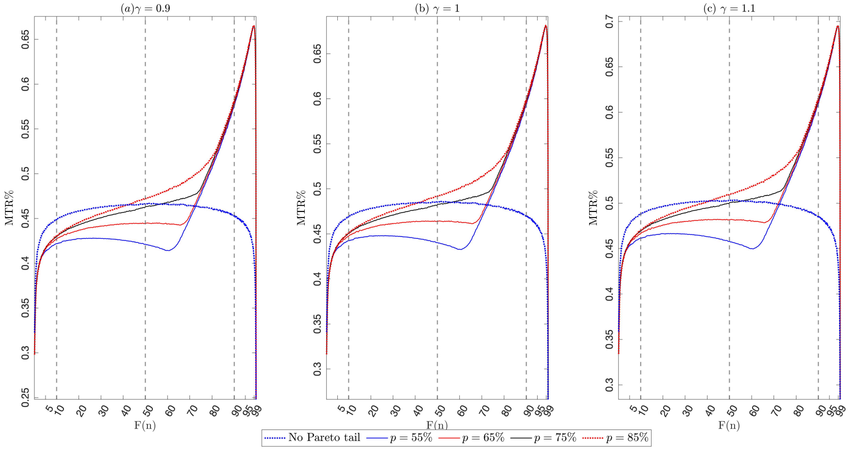


Figure A3: *Optimal marginal tax rates using a log-normal-Pareto distribution considering different shape parameters (Type III, sensitivity analysis)*



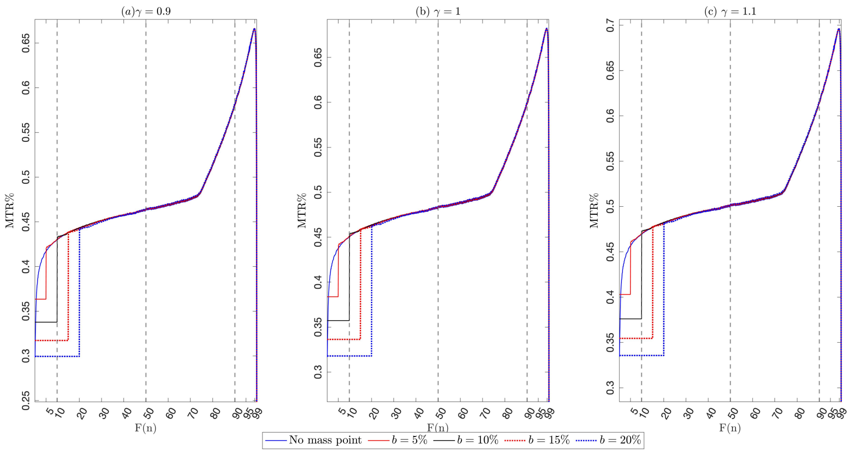
Note: The cutoff point of the Pareto-distribution is fixed the 75th percentile.

Figure A4: *Optimal marginal tax rates using a log-normal-Pareto distribution considering different cutoff points for the Pareto-distribution (Type IV, sensitivity analysis)*



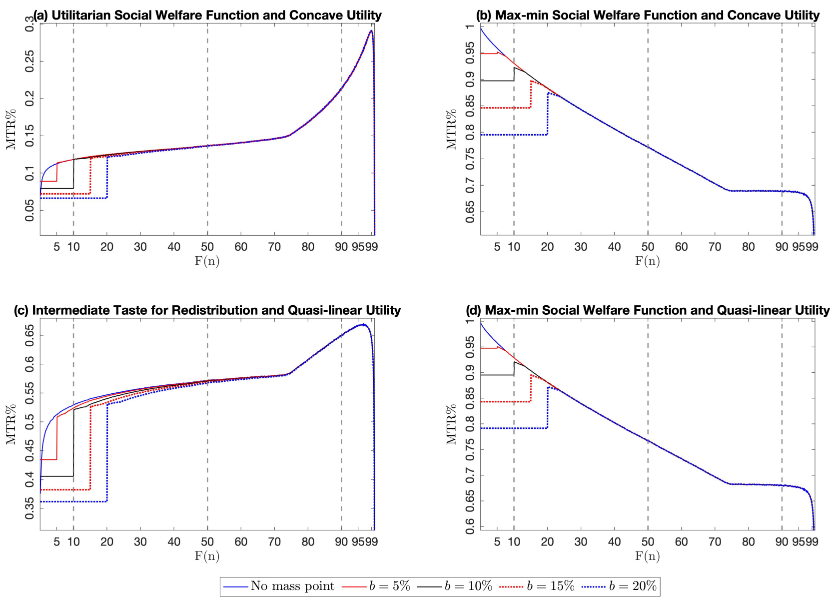
Note: The shape parameter is fixed at $alpha = 2$.

Figure A5: *Optimal marginal tax rates using a log-normal-Pareto distribution considering different mass points at the bottom (sensitivity analysis)*



B.2 Existence of a Pareto tail in a log-normal distribution considering different mass points at the bottom

Figure A6: *Optimal marginal tax rates using a log-normal-Pareto distribution considering different mass points at the bottom*



C Summary statistics

Summary statistics are reported in Table A1.

D Right tails of the wage distributions

Figure A7 depicts the fit of the right tails of the wage distributions.

E Simulated optimal marginal tax rates by country

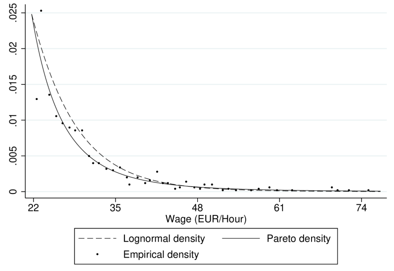
In Figure A8, I show optimal marginal tax rates for individual countries.

Table A1: *Descriptives, 2013*

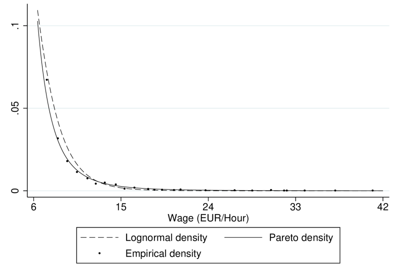
Country	Age	Female	Immigrant	Secondary education	University	Married	Wage	Observations
Austria	40.21	0.480	0.163	0.695	0.191	0.498	15.57	5,019
	(11.67)	(0.500)	(0.369)	(0.461)	(0.393)	(0.500)	(8.26)	
Czech Republic	42.90	0.467	0.011	0.756	0.209	0.577	5.18	6,230
	(11.47)	(0.499)	(0.104)	(0.430)	(0.406)	(0.494)	(3.41)	
Estonia	43.45	0.533	0.148	0.595	0.272	0.475	5.03	5,140
	(13.24)	(0.499)	(0.355)	(0.491)	(0.445)	(0.499)	(3.42)	
Finland	43.84	0.526		0.428	0.485	0.617	21.41	7,087
	(11.66)	(0.499)		(0.495)	(0.500)	(0.486)	(12.06)	
Germany	41.83	0.515	0.233	0.580	0.307	0.596	15.52	14,976
	(11.95)	(0.500)	(0.423)	(0.494)	(0.461)	(0.491)	(14.07)	
Greece	41.49	0.447	0.074	0.419	0.457	0.648	7.47	2,125
	(10.01)	(0.497)	(0.262)	(0.494)	(0.498)	(0.478)	(4.14)	
Ireland	41.67	0.529	0.202	0.313	0.541	0.576	20.63	3,994
	(11.65)	(0.499)	(0.402)	(0.464)	(0.498)	(0.494)	(29.76)	
Japan	49.04	0.450		0.443	0.483	0.845	15.20	2,467
	(12.39)	(0.498)		(0.497)	(0.500)	(0.362)	(19.99)	
Luxembourg	40.45	0.465	0.572	0.432	0.280	0.668	26.22	4,010
	(10.73)	(0.499)	(0.495)	(0.495)	(0.449)	(0.471)	(17.53)	
Netherlands	44.54	0.482	0.050	0.415	0.432	0.632	24.56	9,078
	(11.36)	(0.500)	(0.218)	(0.493)	(0.495)	(0.482)	(18.66)	
Sweden	43.47	0.574	0.142	0.442	0.481	0.449	30.37	2,351,092
	(12.55)	(0.495)	(0.349)	(0.497)	(0.500)	(0.497)	(64.63)	
Switzerland	42.85	0.476	0.243	0.476	0.413	0.564	34.78	6,084
	(12.62)	(0.499)	(0.429)	(0.499)	(0.492)	(0.496)	(29.00)	
UK	41.99	0.511		0.478	0.428	0.706	15.65	16,885
	(12.53)	(0.500)		(0.500)	(0.495)	(0.456)	(14.92)	
United States	41.79	0.481	0.190	0.446	0.459	0.589	18.10	58,039
	(13.54)	(0.500)	(0.392)	(0.497)	(0.498)	(0.492)	(25.72)	

For Sweden, I use large-scale survey data collected by Statistics Sweden. For the 13 other countries, I use data from the Luxembourg Income Study. I utilize the large-scale wage survey data from Statistics Sweden, as the data provides more observations than the Luxembourg Income Study Database (see Section 4.1).

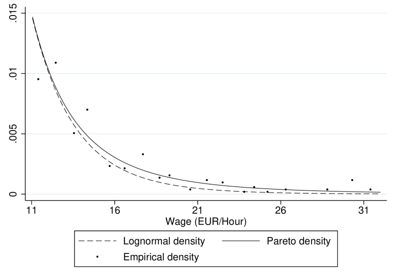
Figure A7: *Right Tails of the Wage Distributions, 2013*



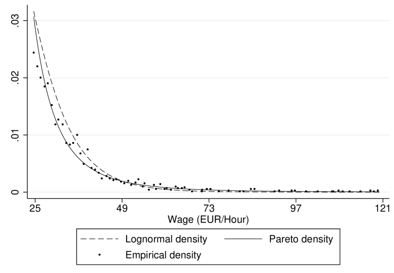
(a) *Austria*



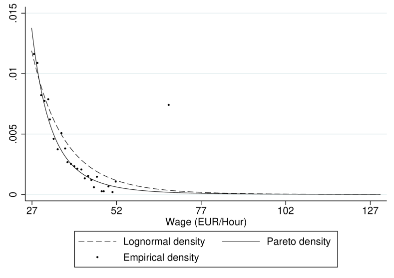
(b) *Czechia*



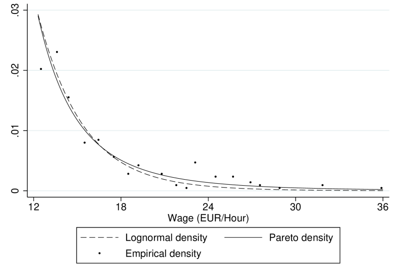
(c) *Estonia*



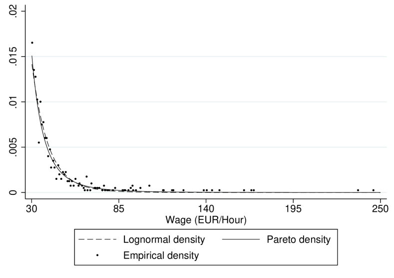
(d) *Finland*



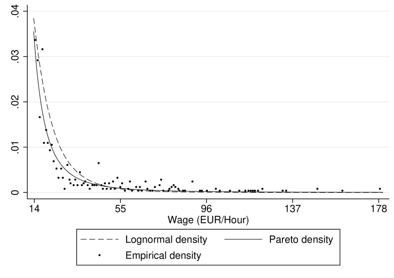
(e) *Germany*



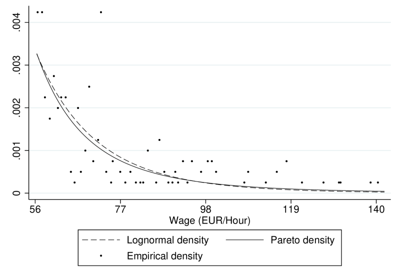
(f) *Greece*



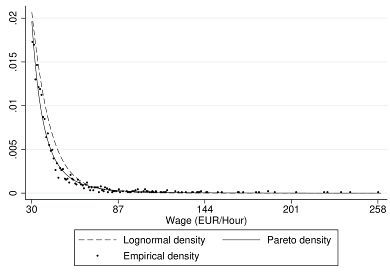
(g) *Ireland*



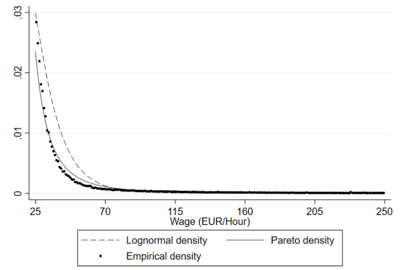
(h) *Japan*



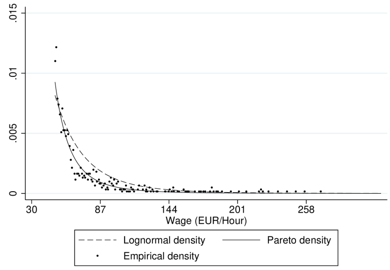
(i) *Luxembourg*



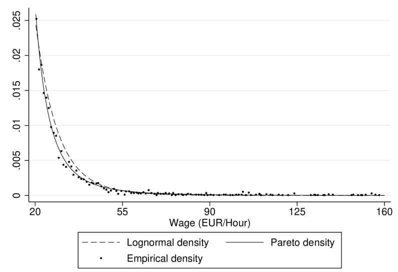
(j) *Netherlands*



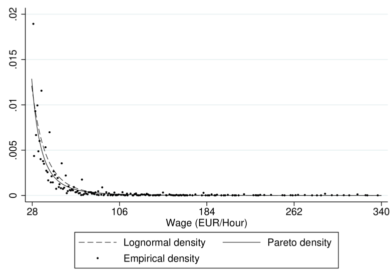
(k) *Sweden*



(l) *Switzerland*

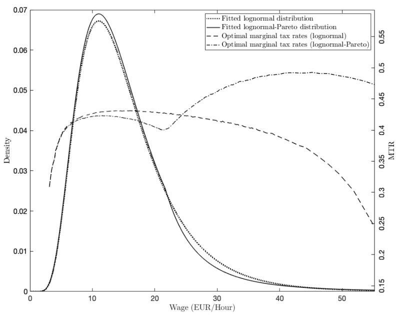


(m) *UK*

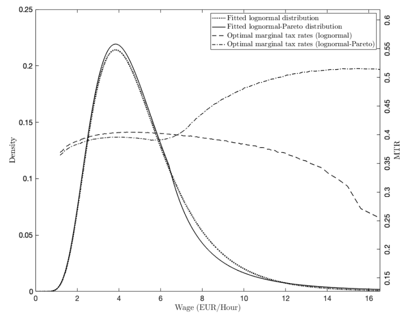


(n) *US*

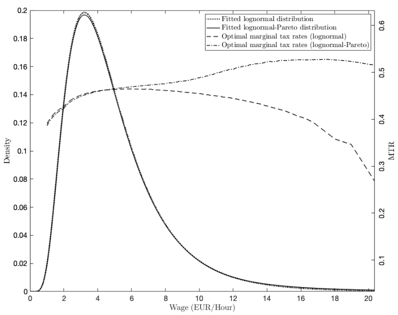
Figure A8: *Optimal marginal tax rates, by country*



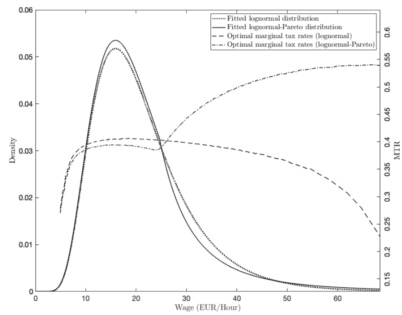
(a) *Austria*



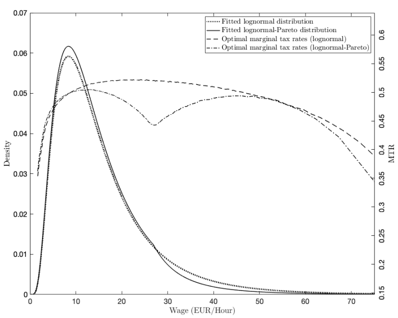
(b) *Czechia*



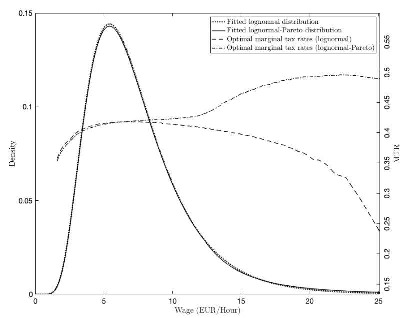
(c) *Estonia*



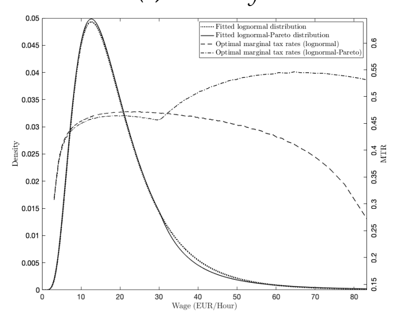
(d) *Finland*



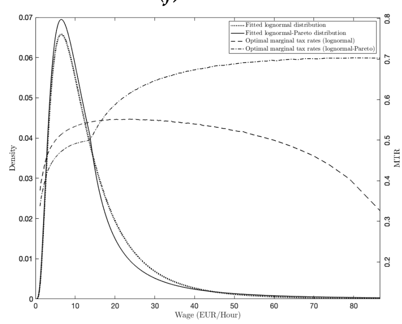
(e) *Germany*



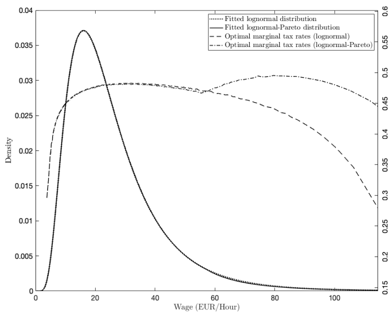
(f) *Greece*



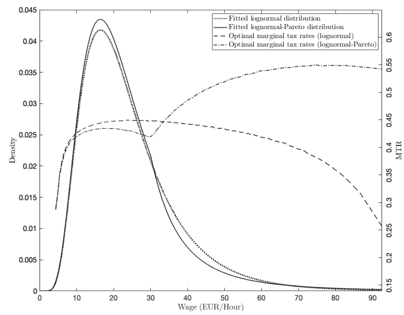
(g) *Ireland*



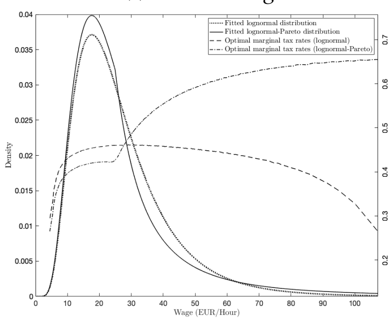
(h) *Japan*



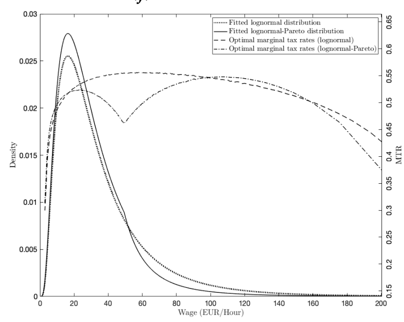
(i) *Luxembourg*



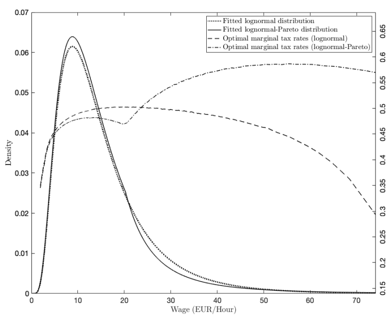
(j) *Netherlands*



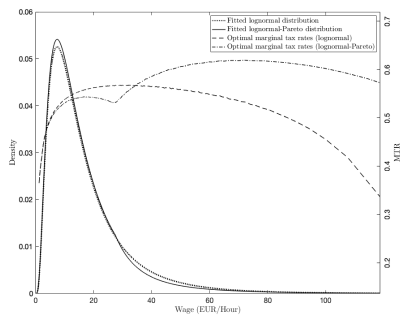
(k) *Sweden*



(l) *Switzerland*



(m) *UK*



(n) *US*