Tagging and Optimal Income Taxation: An Application to Income-Contingent Student Loans*

Sebastian Findeisen
University of Konstanz

Lea Fricke
University of St. Gallen

Dominik Sachs
University of St. Gallen

February 8, 2022

Abstract

We analyze income-contingent student loans within an optimal income-taxation framework in the spirit of Mirrlees (1971). The government uses the observed education degree as a tag and sets an education-dependent tax schedule. We calibrate the model to five European countries (Germany, UK, France, Spain and Italy) and simulate optimal education-dependent marginal tax rates. We find that college graduates should face a larger tax rate than high school graduates until the income level where Pareto tail starts. The implied optimal repayment is non-negative, increasing in income and finite. This is in line with real world policies. We carefully evaluate our results to several other specifications. We find that bandwidth selection plays a crucial role in determining the shape of optimal tax rates.

JEL Codes: H21, H23, I21

Keywords: Optimal Taxation, Education, Education Finance, Student Loans

---

*sebastian.findeisen@uni-konstanz.de, lea.fricke@student.unisg.ch, dominik.sachs@unisg.ch.

The results presented here are based on EUROMOD version I3.0+. Originally maintained, developed and managed by the Institute for Social and Economic Research (ISER), since 2021 EUROMOD is maintained, developed and managed by the Joint Research Centre (JRC) of the European Commission, in collaboration with EUROSTAT and national teams from the EU countries. We are indebted to the many people who have contributed to the development of EUROMOD. The results and their interpretation are the authors' responsibility.

1 Introduction

The idea of income-contingent student loans goes back to Milton Friedman who proposed the following:

> "A governmental body could offer to finance or help finance the training of any individual [...] by making available not more than a limited sum per year for not more than a specified number of years [...]. The individual would agree in return to pay to the government in each future year x per cent of his earnings in excess of y dollars for each $1,000 that he gets in this way. This payment could easily be combined with payment of income tax and so involve a minimum of additional administrative expense." (Friedman 1955)

The main feature of income-contingent student loans is that the repayment is proportional to the income earned after graduation and it is retracted through the tax system. Thus, the repayment is similar to an additional tax on income. Income-contingent loans have the advantage that they provide an insurance against poor labor market outcomes since repayment only accrues when earnings exceed a certain threshold and the repayment is proportional to the income. This eliminates the potential risk of large repayment burdens after graduation (Chapman (2006), Lochner and Monge-Naranjo (2016)).

Several countries like the United Kingdom and the United States have already introduced income-contingent student loans as a policy instrument to finance higher education. In the United Kingdom, students can take out a loan to cover the full amount of costs during college. After graduation, individuals have to repay 9% of earned gross income above the minimum income threshold of around £ 20,000 - £ 27,000. In addition, an interest rate is charged. The repayment schedule involves debt forgiveness after 25 to 30 years depending on the repayment plan. More recently, the United States implemented a voluntary income-contingent loan schedule. Borrowers repay 10% of their income above a minimum income threshold. As in the United Kingdom, an interest rate is charged on debt. After a maximum period of 20 years, all income-contingent repayments are terminated. However, all outstanding debt is treated as income and thus, taxed accordingly. This dissolves the income-contingency of repayment. For example, it may happen that the tax charged on outstanding debt transcends the earned income. This bounded income-contingency clearly limits the insurance feature of income-contingent student loans in the United States (Britton, van der Erve, and Higgins (2019), Lochner and Monge-Naranjo (2016)).

It is a pervasive perception that income-contingent student loans distort labor supply as they act as an additional marginal tax rate on labor income. In 2015, The Economist concluded in an article on governmental student finance that "with repayments linked to income, graduates are discouraged from working."[1] Besides the public debate, this perception is also widespread among economists. Krueger and Bowen (1993, p. 200) raised the concern that "(a)ny income-contingent loan plan is likely to distort labor supply incentives by raising marginal tax rates, perhaps substantially." Barr (1999, p. 19) circulated in a policy consulting paper that "(t)he disadvantage of income-contingent loans are [...] (that) they may be perceived as a tax, with potential disincentive effects" and Barr, Chapman, Dearden, and Dynarski (2019, p. 37) wrote on this issue that "(a) higher repayment rate [...] creates a larger potential distortion to labour supply."[2]

In this paper, we analyze income-contingent student loans within an optimal income taxation framework in the spirit of Mirrlees (1971). This enables us to study the widespread conception from a new perspective. Our results show that a Rawlsian social planner will find it optimal to set higher marginal tax rates for college graduates. This implies that income-contingent repayment schedules are optimal even when distortionary effects of taxes on labor supply are taken into account. Thus, our results challenge the widespread perception that income-contingent student loans are a poor instrument to finance higher education because they are assumed to distort labor supply.

First, we consider a model in which the education decision of an individual is exogenous to the tax system. In the model, individuals differ in innate ability and in their education degree. The government observes total income as well as the education degree of an individual. The observed education degree is used as a tag implying that the government sets an education-dependent tax schedule. To solve the problem, we apply the tax perturbation method established by Piketty (1997) and Saez (2001). We find that, within each education group, the formula for optimal marginal tax rates is identical to the standard optimal income tax formula obtained by Diamond (1998). Further, we find that under a Rawlsian planner and identical elasticities of taxable income, the education group with a larger inverse hazard rate at a certain income level faces the larger marginal tax rate. A relatively larger inverse hazard rate implies that the social planner can generate more tax revenues from distorting labor supply in that education group as the mechanical revenue gain is relatively larger than the behavioral revenue loss. The same intuition applies to the top tax rate. The education group with the thicker top tail of the income distribution should face the higher top tax rate.

Second, we present an empirically driven application of the model with a Rawlsian planner to five European countries. We calibrate country-specific income distributions conditional on education based on the German Socio-Economic Panel and the European Union Statistics on Income and Living Conditions. As top income data is underrepresented in survey data, we ap-

[1] `https://www.economist.com/finance-and-economics/2015/08/20/graduate-stock`

[2] See also Dynarski and Kreisman (2013), who stated that a higher repayment rate raises distortions of labor supply.

pend country-specific Pareto tails. To obtain an approximation of the current tax-and-transfer system, we simulate effective marginal tax rates including benefit phase-out rates with the microsimulation model EUROMOD. Then, we infer the skill distributions from the calibrated income distributions and the simulated effective marginal tax rates. Further, we allow for cross-country differences in the elasticities of taxable income by using country-specific estimates from the literature. The main innovation of the quantitative part is that we account for finite sample limitations by computing confidence intervals for the optimal education-dependent tax schedule. For doing so, we apply a bootstrapping approach. We find that college graduates should face a larger marginal tax rate up to the income level where the Pareto tail starts. For top incomes, the tax rate is identical for college and high school graduates. The education-dependent tax schedule can be interpreted as the repayment of an income-contingent student loan. The high school tax schedule is the common labor income tax schedule and the repayment is the difference in the tax payment under the college and high school tax schedule. The implied repayment schedule from the applied framework is non-negative, strongly increasing in income and finite. This is in line with the properties of repayment schedules of income-contingent student loans in the real-world. Importantly, this finding is identical across all countries.

A general concern is that the shape of simulated marginal tax rates is sensitive to calibration choices of the model. Therefore, we consider various alternative calibration specifications. The corresponding simulation results are summarized in Table 1. First, we extend our model to account for a ternary tag. We find that college graduates with a master degree should face a larger marginal tax rate than college graduates with a bachelor degree. Second, we use a parametric approach with and without an augmented Pareto tail to calibrate the income distribution. As result, the main findings are stable with the exception of small changes in France. Here, the simulated education-dependent marginal tax rates intersect in a small income region before the Pareto tail starts. In that area, high school graduates face a slightly larger tax rate than college graduates. Then, we investigate the influence of the specified bandwidth in the non-parametric calibration of the income distribution on the education-dependent tax rates. We find that the choice of the bandwidth is crucial for the finding of a larger college tax rate in France and Italy. In both countries, college graduates should face a smaller marginal tax rate in a small income region before the Pareto tail starts when using a smaller bandwidth.

Related Literature. This paper is related to the literature on optimal income taxation that goes back to Mirrlees (1971) and Diamond (1998). Piketty (1997) and Saez (2001) proposed the idea to use a small perturbation of the initial tax schedule to solve for the optimal non-linear tax schedule which significantly improved the intuition of underlying mechanisms. Saez (2002) and Jacquet, Lehmann, and Van der Linden (2013) extended the workhorse model of optimal income taxation to account for an extensive margin representing labor force participation. Rothschild

Table 1: Summary of Simulation Results under Different Calibration Specifications

Specification	Marginal Tax Rates (MTR)	Top Tax Rate (TTR)	Section
Baseline	College MTR > High School MTR	Identical TTR	Section 3
Ternary Tag	Master MTR > Bachelor MTR > High School MTR	Identical TTR	Section 4.1
Parametric Approach w/ Pareto Tail	College MTR > High School MTR FRA: College MTR $\gtreqless$ High School MTR	Identical TTR	Section 4.2
Parametric Approach w/o Pareto Tail	College MTR > High School MTR FRA: College MTR $\gtreqless$ High School MTR	Identical TTR FRA: College TTR < High School TTR	Section 4.2
Smaller Bandwidth in KDE	College MTR > High School MTR FRA & ITA: College MTR $\gtreqless$ High School MTR	Identical TTR	Section 4.3
Male Sub-Sample	College MTR > High MTR	Identical TTR	Section 4.2
Country-Specific Pareto Threshold	College MTR > High School MTR	Identical TTR	Section 4.4
More Recent Pareto Parameter	College MTR > High School MTR	Identical TTR	Section 4.4
HSV Tax Schedule	College MTR > High School MTR	Identical TTR	Section 4.4
Identical Elasticities across Countries	College MTR > High School MTR	Identical TTR	Section 4.4
Education Dependent Elasticities	College MTR $\gtreqless$ High School MTR	College TTR < High School TTR	Section 4.4

Notes: College MTR > High School MTR denotes the result that college graduates should face a larger marginal tax rate than high school graduates until the income level where the Pareto tail starts. College MTR $\gtreqless$ High School MTR denotes the result that college graduates should face a larger marginal tax rate than high school graduates except for a small income region before the Pareto tail starts. In this income region, college graduates should face a smaller marginal tax rate than high school graduates. Identical TTR denotes the results that college and high school graduates should face the same top tax rate. College TTR > (<) High School TTR denotes the result that college graduates should face a larger (smaller) top tax rate than high school graduates. See text for details.

and Scheuer (2013) and Sachs, Tsyvinski, and Werquin (2020) include general equilibrium effects. The general quantitative result is that optimal marginal tax rates follow a U-shape pattern with a constant top tax rate. For a recent review of the literature on optimal income taxation, see Piketty and Saez (2013).

Within the field of optimal income taxation, our paper is closely related to two subfields. First, it connects to the tagging literature where taxes are conditional on income and other observed personal characteristics. Akerlof (1978) analyzed the optimal income taxation problem with a categorical tag. His model considers two types of workers, namely low and high skilled workers, that are partitioned into two groups based on an observable characteristic. One group consists of only low skilled workers and one group consists of both types of workers. His results show that tagging allows for a greater redistribution to individuals of the low-skilled group compared to a situation where taxes are only conditional on income.[3]
Cremer, Gahvari, and Lozachmeur (2010) analyze the optimal income taxation problem with an exogenous binary tag in a setting where preferences are quasilinear and the social planner is Rawlsian. They analytically identify winners and losers from tagging under the assumption of a log-normal skill distribution. If the hazard rates of the skill distributions do not cross, all individuals of the group with the lower average skill level benefit from tagging. For individuals of the other group, tagging can be inferior or superior to a situation without tagging. If,

[3]See also Nichols and Zeckhauser (1982), Immonen, Kanbur, Keen, and Tuomala (1998), Boadway and Pestieau (2006) for a formal analysis of the optimal income taxation problem when the population can be tagged into groups.

however, the hazard rates cross, losers and winners from tagging change after the crossing. Further, they show that the untagged marginal tax rate is enclosed by the two tagged marginal tax rates. Finally, they calibrate their model to the US using gender as a tag. Their results show that tagging entails large welfare improvements. The only losers from tagging are male high-wage earners. Mankiw and Weinzierl (2010) and Weinzierl (2011) use the example of body height and age as an exogenous tag, respectively. Both come to the conclusion that other factors besides income should be considered for determining taxes. Further, Weinzierl (2011) shows that age dependent taxes lead to substantial welfare gains. We extend the framework of Cremer, Gahvari, and Lozachmeur (2010) to the more general case of a weighted Utilitarian planner. A key distinction to Cremer, Gahvari, and Lozachmeur (2010), Mankiw and Weinzierl (2010) and Weinzierl (2011) is that we apply the intuitive tax perturbation approach by Piketty (1997) and Saez (2001) to derive optimal tagged tax rates. In the quantitative part, we calibrate the skill distribution from income distributions that are augmented by a Pareto tail instead of using a log-normal wage distribution as in Cremer, Gahvari, and Lozachmeur (2010).

Second, our paper is related to the literature on integrated education and tax systems. Bovenberg and Jacobs (2005) include endogenous education to a Mirrleesian optimal income taxation model and analyze optimal tax and education policies. They show that education subsidies and progressive taxation are siamese twins as education subsidies can eliminate the distortionary impact of labor income taxes on education.[4] Gary-Bobo and Trannoy (2015) consider a combined two-type model of student loans and optimal income taxation. They find that optimal student loan repayments are always income-contingent, either by implementing a progressive graduate tax or by offering income-contingent student loan schedules. In contrast to their work, we focus on the interpretation of an income-contingent student loan schedule rather than a graduate tax as income-contingent loans have the advantage of not harming horizontal equity concerns and thus, are more realistic to be implemented in the real-world. Further, we use a continuous type distribution instead of only two types, which is standard in the large literature on optimal income taxation.

Our paper is closely related to Findeisen and Sachs (2016) who study the effects of education on optimal income taxes in a dynamical framework.[5] They consider a model where individuals make a binary college education decision and draw their labor market ability before entering

[4]Boháček and Kapička (2008) analyze optimal income taxes and optimal education subsidies in a dynamic model and come to a similar conclusion as Bovenberg and Jacobs (2005). Anderberg (2009) analyzes optimal tax and education policies in an environment with risky human capital where education investments influence the degree of wage risk. The main finding is that there should be a positive education premium if education increases the wage risk.

[5]See also Stantcheva (2017) and Kapička and Neira (2019). Stantcheva (2017) focuses on optimal income and education policies when human capital is accumulated over the life-cycle. Kapička and Neira (2019) focus on optimal income taxation in a dynamic model with risky human capital that is acquired during the working age. Our paper focuses on an earlier stage of human capital acquisition, namely before the labor market entry.

the labor market. Without making any ex-ante assumption on the instruments of the government, they find that income-contingent student loans can always be designed in a way that they are second-best Pareto efficient. Further, they apply their framework to US data and simulate the optimal tax and repayment schedule. Similar to our results, they find that college graduates should face a larger marginal tax rate up to the point where the Pareto tail kicks in. The top tax rate is almost similar for college and high school graduates. The approximated repayment increases linearly in income and stays constant afterwards. Further, they show that income-contingent repayment schedules lead to large welfare gains.[6] In contrast to their work, we consider a static model where all decisions are made within one period. This has the advantage that the elementary trade-offs can be described comprehensibly. Our results can easily be extended to a dynamic setting. Further, we assume an exogenous education decision. Findeisen and Sachs (2016) argue that the quantitative effects of endogenous education are rather small.

Finally, our paper is related to the small empirical literature on the effects of income-contingent student loans on labor supply. Britton and Gruber (2020) investigate the influence of an additional marginal tax rate emerging from an income-contingent student loan on labor supply. Using an administrative dataset they estimate labor supply reactions through a bunching at kinks approach in the United Kingdom. Their results suggest that there is no negative impact of income-contingent repayment schedules on labor supply in the United Kingdom. Chapman and Leigh (2009) use a bunching at notches estimator to identify the impact of the Australian income-contingent student loan system on labor supply. They only find that individuals bunch below the threshold, however, the effect is economically small. Herbst (2020) studies the effects of income-contingent repayments on various outcomes of borrowers, including labor supply. He deduces from his results that income-contingent student loans have only small distortionary effects on labor supply.

Roadmap. In Section 2, we describe the model. In Section 3, we calibrate the model to five European countries and evaluate optimal education-dependent marginal tax rates quantitatively. Section 4 evaluates optimal education-dependent marginal tax rates under a ternary tag and various alternative calibration specifications. Conclusions are drawn in Section 5.

[6] Matsuda and Mazur (2020) analyze welfare effects of income-contingent student loans with uninsurable college dropout risk analytically and quantitatively. Similar to Findeisen and Sachs (2016), they find that income-contingent student loans together with an income tax can implement the second-best allocation. This implies that student income-contingent student loans can increase the overall welfare of an economy. In addition, they find that student income-contingent student loans raise college enrollment. However, student income-contingent student loans also lower incentives for educational effort during college.

2 The Model

We start by introducing the theoretical model that is used in the simulation exercises later. For simplicity, we consider a static model, where the education decision of individuals is independent of the tax-and-transfer system.[7] The government sets an education-dependent tax schedule, which can be interpreted as an income-contingent student loan in the real-world. In particular, the government would offer a compulsory income-contingent loan schedule to finance tuition fees and consumption during college. After graduation, the repayment is collected along with the labor income tax. This implies that the government uses the observed education degree as a pure exogenous tag.

2.1 Structure

Set Up Consider an economy with a continuum of individuals normalized to 1. Individuals differ in innate ability $\theta \in \Theta = [\underline{\theta}, \overline{\theta}]$ and in their education degree $e_i \in \{e_{hs}, e_{col}\}$, where the index hs refers to a high school degree and col to a college degree. Innate ability is private information and cannot be observed by the government, whereas education is publicly observable. The government uses the observed education degree as a tag and partitions individuals into two mutually exclusive groups with exogenous population proportions π_{e_i}, where $\sum_{i=col,hs} \pi_{e_i} = 1$. One group consists of individuals with a high school degree and the other group consists of individuals with a college degree.

Heterogeneity In both groups, innate ability θ is distributed according to the cumulative distribution function $F_{e_i}(\theta)$, where the support is education-independent and given by $[\underline{\theta}, \overline{\theta}]$ with $\underline{\theta} \geq 0$. The corresponding probability density function is noted by $f_{e_i}(\theta)$ and assumed to be strictly positive and differentiable for all $\theta \in [\underline{\theta}, \overline{\theta}]$. The average level of innate ability is higher in the college group than in the high school group as ability and education are positively correlated.

Preferences All individuals have identical preferences over consumption c and labor effort l that are characterized by the quasi-linear utility function $U(c - v(l))$, where $v(l)$ describes the disutility of labor effort.[8] $U(.)$ is assumed to be concave in consumption and $v(l)$ is assumed to be strictly convex with $v(0) = 0$. Individuals earn a gross income $y = \theta \cdot l$ and consumption

[7] This implies that our model cannot capture foregone earnings during college nor borrowing constraints. However, it still captures the main trade-offs of the government when designing education-dependent tax schedules. Findeisen and Sachs (2016)) consider a dynamic model of an integrated education finance and tax systems. Their quantitative results are similar to our findings.

[8] The assumption of quasi-linear preferences is widely used in the literature on optimal income taxation, e.g. Diamond (1998), Bovenberg and Jacobs (2005), Rothschild and Scheuer (2013), Sachs, Tsyvinski, and Werquin (2020) among others and also in the literature on tagging, e.g. Cremer, Gahvari, and Lozachmeur (2010) and

equals $c = y - T_{e_i}(y)$, where $T_{e_i}(y)$ is a twice-continuously differentiable education-dependent tax schedule. Individuals choose their labor effort by maximizing utility:

$$\max_{l_{e_i}} U(c_{e_i} - v(l_{e_i})) \text{ s.t. } c_{e_i} = y_{e_i} - T_{e_i}(y_{e_i}). \tag{1}$$

The optimal level of labor effort is determined by the first order condition of the utility maximization problem:

$$v'(l_{e_i}) = (1 - T'_{e_i}(y_{e_i}))\theta. \tag{2}$$

Tax Schedule The government uses the observed education degree as a tag and sets an education-dependent tax schedule $T_{e_i}(y)$. In the real world, education-dependent taxes can be interpreted as the repayment of an income-contingent student loan. The high school tax schedule $T_{e_{hs}}(y)$ is the common labor income tax schedule that applies to all individuals, i.e. $T(y) = T_{e_{hs}}(y)$. The repayment $R(y)$ of the income-contingent student loan is then set as the difference between both tax schedules, i.e. $R(y) = T_{e_{col}}(y) - T_{e_{hs}}(y)$.[9] Findeisen and Sachs (2016) show that such an allocation is incentive-compatible and can always be implemented as a second-best Pareto optimum.

2.2 The Planner's Problem

The social planner sets an education-dependent non-linear income tax schedule $\{T_{e_{hs}}(y(\theta)), T_{e_{col}}(y(\theta))\}$ by maximizing a weighted Utilitarian welfare function. The social planner is constrained by a resource constraint and by taking individuals behavioral responses into account. For the ease of notation, we sometimes suppress the dependence of variables. Formally, the optimization problem reads as follows:

$$\max_{\{T_{e_{hs}(y(\theta))}, T_{e_{col}}(y(\theta))\}} \sum_{i=hs,col} \int_{\underline{\theta}}^{\bar{\theta}} U\left(y_{e_i} - T_{e_i}(.) - v\left(\frac{y_{e_i}}{\theta}\right)\right) s(\theta) f_{e_i}(\theta) d\theta, \tag{3}$$

subject to a resource constraint

$$\sum_{i=hs,col} \pi_{e_i} \int_{\underline{\theta}}^{\bar{\theta}} T_{e_i}(y_{e_i}(\theta, T_{e_i}(.))) f_{e_i}(\theta) d\theta \geq E, \tag{4}$$

Weinzierl (2011). Furthermore, the empirical literature finds only very small income effects, e.g. Gruber and Saez (2002).

[9] As mentioned by Findeisen and Sachs (2016), the repayment schedule may be decreasing in income as the repayment depends on whether marginal tax rates for college graduates are larger or smaller than marginal tax rates of high school graduates. In contrast to this theoretical result, the repayments of income-contingent student loans in the real-world are restricted to be increasing in income, non-negative and finite. The quantification of our model yields strong support for all three properties in five European countries. In particular, repayments are strongly increasing in income as college graduates face larger marginal tax rates than high school graduates.

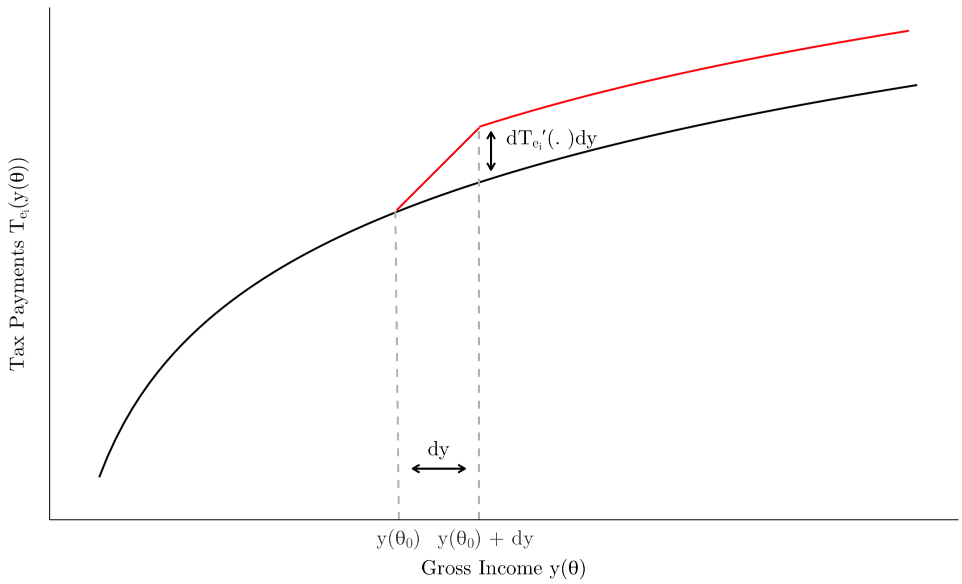


Figure 1: Tax Perturbation

where E denotes the aggregate exogenous revenue requirement and taking individual behavior into account:

$$y(\theta, T_{e_i}(.)) = \arg\max_{y_{e_i}} \; U\Big(y_{e_i} - T_{e_i}(y_{e_i}) - v\Big(\frac{y_{e_i}}{\theta}\Big)\Big). \tag{5}$$

$s(\theta)$ represents a set of social marginal welfare weights that are normalized to 1, i.e. $\sum_{i=hs,col} \pi_{e_i} \int_{\underline{\theta}}^{\bar{\theta}} s(\theta) f_{e_i}(\theta) d\theta = 1$.

2.3 Optimal Education-Dependent Taxes

To find the optimal education-dependent tax schedule, we use the perturbation method as in Piketty (1997), Saez (2001), Golosov, Tsyvinski, and Werquin (2014), Raj (2019), Jacquet and Lehmann (2020), Lockwood (2020) and Bierbrauer, Boyer, and Peichl (2021). The idea is to perturb the initial tax schedule $T_{e_i}(.)$ by a marginal increase of $dT'_{e_i}(.)$ on the small income interval $[y(\theta_0); y(\theta_0) + dy]$. Figure 1 outlines the tax perturbation. In the optimum the first-order welfare effect of this small tax perturbation must be zero, which pins down the necessary condition for the optimal tax schedule.

First, let $H_{e_i}(y(\theta))$ be the cumulative income distribution functions and $h_{e_i}(y(\theta))$ the corresponding income probability density functions. The skill and income distributions are related through $F_{e_i}(\theta) = H_{e_i}(y(\theta))$. As in Saez (2001), denote the virtual income density that would occur under a linear tax schedule as $h^*_{e_i}(y(\theta))$. Furthermore, define λ_{e_i} as the marginal value of public funds and $g_{e_i}(\theta) = \frac{U'_c(.)s(\theta)}{\lambda_{e_i}}$ as the social marginal welfare weight of an individual of type θ for each education group e_i, following Piketty and Saez (2013).

Tax Perturbation The perturbation of the initial tax schedule leads to three effects: a mechanical revenue effect, a welfare effect and a behavioral effect. The mechanical revenue effect captures the increase in tax revenues induced by the small increase of dT'_{e_i}. Individuals with an income of $y > y(\theta_0)$ pay $dT'_{e_i}dy$ more taxes leading to an mechanical increase of tax revenues by

$$dM_{e_i}(y(\theta_0)) = dT'_{e_i}dy \int_{y_0}^{\overline{y}} h_{e_i}(y(\theta))dy. \tag{6}$$

This also induces a welfare loss for individuals with an income of $y > (\theta_0)$, captured by the welfare effect

$$dW_{e_i}(y(\theta_0)) = -dT'_{e_i}dy \int_{y_0}^{\overline{y}} g_{e_i}(y(\theta))h_{e_i}(y(\theta))dy, \tag{7}$$

where the social planner values the welfare loss of individuals with an income of $y > y(\theta_0)$ by their welfare weight $g_{e_i}(y(\theta))$.

Additionally, individuals with an income of $y \in [y(\theta_0); y(\theta_0) + dy]$ respond to the increase in the marginal tax rate by reducing labor effort as a larger marginal tax rate decreases the marginal benefit of gross income. This translates into a loss in tax revenues captured by the behavioral effect

$$dB_{e_i}(y(\theta_0)) = -dT'_{e_i}dy \frac{T'_{e_i}(y(\theta_0))}{1 - T'_{e_i}(y(\theta_0))} y(\theta_0)\varepsilon(y(\theta_0))h^*_{e_i}(y(\theta_0)), \tag{8}$$

where $\varepsilon_{e_i}(y(\theta))$ is the income elasticity with respect to the net-of-tax rate at income $y(\theta)$ of education group e_i. The reduction of labor supply has no first-order welfare effect on individual utility as the envelope theorem applies. Note, that the behavioral effect only captures the substitution effect as the income effect is zero due to the assumption of quasi-linear preferences.

Optimal Education-Dependent Taxes In the optimum, the sum of all three welfare effects must be zero:

$$dM_{e_i}(y(\theta_0) + dW_{e_i}(y(\theta_0)) + dB_{e_i}(y(\theta_0)) = 0. \tag{9}$$

Rearranging this condition yields the formula for education-dependent optimal marginal tax rates summarized in Proposition 1. Similar, the welfare effect of a small increase in the lump-sum element must be zero:

$$dM_{e_i}(y(\underline{\theta}) + dW_{e_i}(y(\underline{\theta})) = 0, \tag{10}$$

which characterizes the optimal education-dependent lump-sum element. Note that a small increase in the lump-sum element does not involve a behavioral effect as preferences are assumed to be quasi-linear.

Proposition 1 (Optimal Education-Dependent Marginal Tax Rates). *The optimal marginal tax rates for college and high school graduates at income $y(\theta_0)$ are given by*

$$\frac{T'_{e_i}(y(\theta_0))}{1-T'_{e_i}(y(\theta_0))} = \left(1+\frac{1}{\varepsilon_{e_i}(y(\theta_0))}\right)\cdot\left(1-\bar{g}_{e_i}(\theta_0)\right)\cdot\frac{1-F_{e_i}(\theta_0)}{f_{e_i}(\theta_0)\theta_0},$$

where $\bar{g}_{e_i}(\theta_0) \equiv \frac{\int_{\theta_0}^{\bar{\theta}} g_{e_i}(\theta)f_{e_i}(\theta)d\theta}{1-F_{e_i}(\theta)}$ *is the average social marginal welfare weight for individuals with* $y > y(\theta_0)$*.*

Proof. Rearranging (9) as well as changing variables from $y(\theta_0)$ to θ_0 and using $y(\theta_0)h^*(y(\theta_0))(1+\varepsilon_{e_i}(y(\theta_0))) = \theta_0 f_{e_i}(\theta_0)$, which follows directly from Lemma 1 in Saez (2001), yields the result. □

Within each education group, the optimal tagged tax rates are equivalent to those obtained in a standard Mirrleesan income taxation problem by Diamond (1998). Under tagging, all terms are group specific and do not refer to the whole population. Furthermore, the derived optimal marginal tax rates are independent of the properties of the other group.[10]

The tagged optimal tax schedule is also closely related to the formulas derived by Cremer, Gahvari, and Lozachmeur (2010) in an optimal income taxation model with an exogenous tag. The only difference is that Cremer, Gahvari, and Lozachmeur (2010) consider a Rawlsian planner that maximizes the utility of the least off individual. Now, suppose that the social planner is Rawlsian by setting the welfare weights $s(\theta)$ to $s(\underline{\theta}) = 1$ and $s(\theta) = 0 \; \forall \; \underline{\theta} \neq \theta$. This implies that $\bar{g}_{e_i} = 0 \; \forall i \; \in \{hs, col\}$ and thus, the optimal tax schedule in (1) collapses to

$$\frac{T'_{e_i}(y(\theta_0))}{1-T'_{e_i}(y(\theta_0))} = \left(1+\frac{1}{\epsilon_{e_i}(y(\theta_0))}\right)\cdot\frac{1-F_{e_i}(\theta_0)}{f_{e_i}(\theta_0)\theta_0}, \tag{11}$$

which is identical to the formula derived by Cremer, Gahvari, and Lozachmeur (2010). Under an identical elasticity across education groups, differences between the college and high school marginal tax rate are traced back to the inverse hazard rate $\frac{1-F_{e_i}(\theta_0)}{f_{e_i}(\theta_0)\theta_0}$, which captures the relative size of the mechanical revenue gain and the revenue loss due to behavioral responses. At a certain income level, the social planner sets a higher tax rate for college graduates if and only if the inverse hazard rate of the college group is larger than of the high school group, i.e. $\frac{1-F_{e_{col}}(\theta_0)}{f_{e_{col}}(\theta_0)\theta_0} > \frac{1-F_{e_{hs}}(\theta_0)}{f_{e_{hs}}(\theta_0)\theta_0}$. A relatively larger inverse hazard rate implies that the social planner can generate more tax revenues from distorting labor supply in the college group than in the high school group as the mechanical revenue gain is relatively larger than the behavioral revenue loss. Thus, it is optimal to set a higher tax rate for college graduates.

[10]This result is identical to Cremer, Gahvari, and Lozachmeur (2010) who derive optimal income tax formulas with an exogenous tag under a Rawlsian social planner.

2.4 Optimal Tagged Top Tax Rates

Now, we derive the characteristics of the asymptotic tax rate assuming that the income distribution is unbounded. First, rewrite the inverse hazard ratio $\frac{1-F_{e_i}(\theta_0)}{f_{e_i}(\theta_0)\theta_0}$ in terms of the income distribution. Using the relationship between the income and wage distribution $h^*(y(\theta_0))y'(\theta_0) = f_{e_i}(\theta_0)$, the formula for the optimal tagged marginal tax rates in (1) can be rewritten as follows:

$$\frac{T'_{e_i}(y(\theta_0))}{1-T'_{e_i}(y(\theta_0))} = \left(\frac{1}{\epsilon_{e_i}(y(\theta_0))}\right) \cdot \left(1-\bar{g}_{e_i}(\theta_0)\right) \cdot \frac{1-H_{e_i}(y(\theta_0))}{h^*_{e_i}(y(\theta_0))y(\theta_0)}. \tag{12}$$

It is a prominent fact that the tail of an income distribution can be approximated by a Pareto distribution.[11] Assume that both tails of the education-dependent income distributions $h_{e_i}(y(\theta))$ are Paretian with the Pareto parameter a_{e_i}. This entails, that the inverse hazard rate of income $\frac{1-H_{e_i}(y(\theta_0))}{h^*_{e_i}(y(\theta_0))y(\theta_0)}$ tends to $\frac{1}{a_{e_i}}$ at the top of the income distribution. Furthermore, assume that the government does not value redistribution between individuals at the very top of the distribution implying that $g_{e_i}(\theta)$ converges to a constant $\bar{g}_{e_i}$. Then, the top tax rate τ_{e_i} in the college and high school optimal tax schedule is described by

$$\frac{\tau_{e_i}}{1-\tau_{e_i}} = \frac{1}{a_{e_i} \cdot e_{e_i}} \cdot (1-\bar{g}_{e_i}), \tag{13}$$

where e_{e_i} is the constant income elasticity with respect to the net-of-tax rate in both education groups.

To obtain a better understanding of differences in the top tax rate between both groups, consider a Rawlsian planner and assume that elasticities do not vary across high school and college graduates, which is consistent with empirical evidence.[12] Then, the optimal top tax rate boils down to

$$\frac{\tau_{e_i}}{1-\tau_{e_i}} = \frac{1}{a_{e_i} \cdot e}. \tag{14}$$

By now, differences in the optimal top tax rate only arise due to differences in the Pareto parameter a_{e_i}. If both tails of the education-dependent income distributions have the same thickness, i.e. $a_{col} = a_{hs}$, , then college and high school graduates should face the same top tax rate. Contrarily, if the tail of the college income distribution is thicker (thinner) than the tail of the high school distribution, then college graduates should face a higher (smaller) top tax rate.

[11] For example, Saez (2001) shows that the top tail of the empirical US earnings distribution can be well approximated by a Pareto distribution.

[12] For example, Bargain, Orsini, and Peichl (2014) find no statistically significant differences in wage elasticities across education groups.

Example *Assume that the elasticity is the same across education groups and takes the value of* $e = 0.25$*, which is an empirically plausible estimate according to Saez, Slemrod, and Giertz (2012). First, consider a college and high school income distribution where the Pareto parameter is the same and takes the value of* $a = 2$*. This implies that the top tax rate equals* $\tau^*_{e_i} = 0.67 \ \forall \ i \in \{col, hs\}$ *in both education groups. Now, consider an economy where the college income distribution has a thicker tail than the high school income distribution. Let the Pareto parameter of the college income distribution be* $a_{e_{col}} = 1.5$ *and of the high school income distribution be* $a_{e_{hs}} = 2$*. This implies that the top tax rate of college graduates equals* $\tau^*_{col} = 0.72$ *and of high school graduates* $\tau^*_{hs} = 0.67$*.*

3 Quantitative Exploration of the Simplified Model

We calibrate the model with a Rawlsian planner to five European countries assuming quasi-linear preferences of the form of $U(c, l) = c - \frac{l^{1+\epsilon_{e_i}}}{1+\epsilon_{e_i}}$, where ϵ_{e_i} is the income elasticity with respect to the net-of-tax rate. The set of countries include Germany, the United Kingdom (UK), France, Spain and Italy. For doing so, we need country-specific skill distributions and values for the elasticity of taxable income. We infer the skill distributions by inverting the individual labor supply first order condition as in Saez (2001). This requires the calibration of country-specific income distributions and an approximation of effective marginal tax rates implied by the current tax-and-transfer system, as described in Section 3.1.[13] Then, we simulate education-dependent optimal marginal tax rates and the implied repayment schedule for all countries. In addition, we compute 95% confidence intervals for the education-dependent optimal marginal tax rates by applying a bootstrapping approach. So far, the literature on optimal income taxation has not accounted for finite sample limitations of the underlying income distributions.[14] For policy implications as in the case of an education-dependent tax schedule it is essential, to account for the statistical sensitivity of the tax schedule to avoid misleading policy implications. Appendix A.2 describes the simulation and bootstrapping procedure in detail.

3.1 Calibration

Data To obtain country-specific income distributions conditional on education, we use data on annual gross labor incomes and education from the 2017 German Socio-Economic Panel (GSOEP) for Germany and the 2018 cross-sectional European Union Statistics on Income and Living Conditions (EU-SILC) for the UK, Spain, France and Italy. All income distributions

[13]The calibration of the model is close to the calibration of the workhorse model of optimal income taxation in Ayaz, Fricke, Fuest, and Sachs (2021)

[14]Beresteanu and Dahan (2002) are an exemption who compute confidence intervals for the hazard rate. However, they do not investigate the statistical sensitivity of the tax schedule.

refer to the year 2017. We divide the sample into a college and a high school group based on the highest reported education level. The college group contains all individuals with a tertiary education degree and the high school group contains all individuals without a tertiary education degree. See Appendix A.1.1 for a detailed description of both data sources. Table 2 outlines the characteristics of the samples used for the calibration of the income distributions. First, the share of females in each of the education groups is similar to the share in the whole sample. This limits the potential threat that education is used as an indirect tag for gender. Second, the shares of college graduates are close to the ones reported by the OECD (2021), though our samples yield a little bit higher college shares. Lastly, college graduates obtain a wage premium reflected by a higher annual income compared to the income of high school graduates across all countries.

Table 2: Data Sources and Parameters for the Calibration of Income and Skill Distributions

	Germany	UK	Spain	France	Italy
Data					
Source	GSOEP	EU-SILC	EU-SILC	EU-SILC	EU-SILC
Year of Income Distribution	2017	2017	2017	2017	2017
Number of Observations	15 406	7 647	13 ,775	9 914	18 ,763
Female Share	50.9 %	52.2 %	47.2 %	50.3 %	46.0 %
Female Share - College	49.4 %	54.8 %	53.4 %	54.5 %	53.7 %
Female Share - High School	51.5 %	49.3 %	42.8 %	47.6 %	44.0 %
College Share	30.8 %	52.1 %	41.0 %	39.6 %	21.3 %
OECD College Share	28.6 %	45.7 %	36.4 %	35.2 %	18.7 %
Mean Income	34 839 €	34 203 €	20 892 €	29 976 €	27 277 €
Mean Income - College	48 750 €	41 818 €	27 971 €	39 810 €	37 039 €
Mean Income - High School	28 636 €	25 933 €	15 972 €	23 534 €	24 641 €
Calibration					
Pareto Threshold	150 000 €	150 000 €	150 000 €	150 000 €	150 000 €
Pareto Parameter	1.67	1.78	2.11	2.20	2.22
Mass of People with Zero Earnings	4.4 %	7.0 %	3.8 %	5.6 %	3.2 %
Elasticity	0.54	0.25	0.45	0.20	0.45

Notes: EU-SILC reports all data in Euros. The values of the Pareto parameter are from Atkinson, Piketty, and Saez (2011). The mass of people with zero earnings matches the shares of recipients of disability benefits reported by OECD (2009). For France, we use the average across OECD countries. For the elasticity of taxable income, we use the estimate of Doerrenberg, Peichl, and Siegloch (2017) for Germany, Brewer, Saez, and Shepard (2010) for the UK, Almunia and Lopez-Rodriguez (2019) for Spain, Lehmann, Marical, and Rioux (2013) for France and the approximate midpoint of estimated elasticities as reported by Saez, Slemrod, and Giertz (2012) for Italy.

Income Distributions To calibrate the country-specific income distributions conditional on education, a kernel density estimation is applied to obtain a smooth distribution.[15] For incomes above 150 000 €, we add a Pareto distribution with a constant Pareto parameter matching the

[15]See Appendix A.1.2 for a detailed description of the applied kernel density estimation.

country-specific values from Atkinson, Piketty, and Saez (2011). To the best of our knowledge, there exists no evidence on top income distributions conditional education. Therefore, we assume that the Pareto tail is identical for both education groups. Moreover, a constant share of the population with zero income is assumed. The shares correspond to the country-specific shares of recipients of disability benefits as in Mankiw, Weinzierl, and Yagan (2009).[16] Table 2 summarizes the choices of the country-specific values used in the calibration of the income distributions.

Current Tax Systems We approximate the current income tax systems with the tax-benefit microsimulation model EUROMOD and EU-SILC as input data.[17] First, we simulate effective marginal tax rates including benefit phase-out rates based on the 2017 tax-and-transfer system. The simulated effective marginal tax rates include income taxes, mean-tested benefits, pension and social insurance contributions.[18] Then, we calculate average effective marginal tax rates for income bins with a size of 5 000 €. To smooth the average marginal tax rates, we use a local polynomial estimator with a linearly increasing bandwidth. Subsequently, we perform a second smoothing of the predicted marginal tax rates to overcome kinks at the points where the bandwidth changes. For doing so, we use a local polynomial estimator with a constant bandwidth.[19] The simulated average and smoothed marginal tax rates are illustrated in Figure A1 in Appendix A.1.3.

Skill Distributions We infer the skill distributions from the calibrated income distributions and the simulated effective marginal tax rates by inverting the individual labor supply first-order condition as in Saez (2001). For the income elasticity, we use empirical evidence to match the country-specific values.[20] Table 2 contains details on the elasticities used in the quantitative exercise. According to empirical evidence from Bargain, Orsini, and Peichl (2014) and Keane

[16]The shares of disability benefits recipients are taken from the Employment Outlook of the OECD (2009), containing data up to the year 2007. Unfortunately, country-specific data for France is missing in the dataset. Thus, the average across OECD countries is used for France.

[17]For detailed information about the tax-benefit calculator EUROMOD, see Sutherland and Figari (2013).

[18]EUROMOD simulates effective marginal tax rates based on a marginal increase of gross earnings of 3%. Within one household, gross earnings are increased for each member with positive earnings. Finally, effective marginal tax rates are calculated based on the difference in household disposable income and stored on individual level. For a detailed description, see Jara and Tumino (2013).

[19]Restricting the first local polynomial estimator to a constant bandwidth requires a very large bandwidth as the number of observations with high incomes is limited. A very large bandwidth leads to a loss in accuracy for low incomes implying that the strong increase in marginal tax rates is not captured anymore. To avoid the problem, a linearly increasing bandwidth between 2 500 € and 70 000 € over a grid with 1001 nodes is used. For the second local polynomial estimator, the bandwidth is set to a constant value of 10 000 €.

[20]We use the estimate of Doerrenberg, Peichl, and Siegloch (2017) for Germany, Brewer, Saez, and Shepard (2010) for the UK, Almunia and Lopez-Rodriguez (2019) for Spain and Lehmann, Marical, and Rioux (2013) for France. To the best of our knowledge, there exists no evidence on the elasticity of taxable income for Italy. Therefore, we use the approximate midpoint of estimated elasticities as reported by Saez, Slemrod, and Giertz (2012)

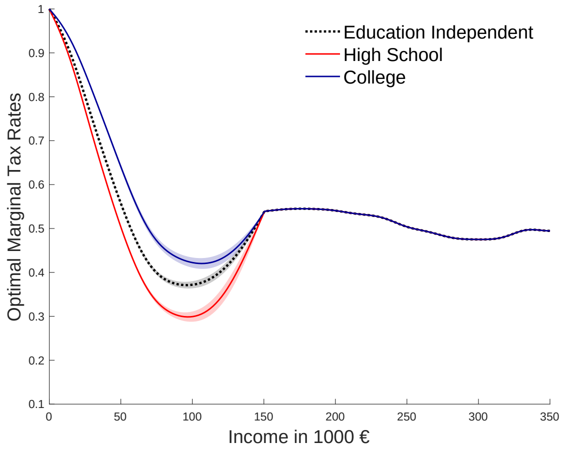
and Wasi (2016), we assume that college and high school graduates have identical elasticities. In Section 4.4, we show how our results change assuming that college graduates have a higher elasticity.

3.2 Results

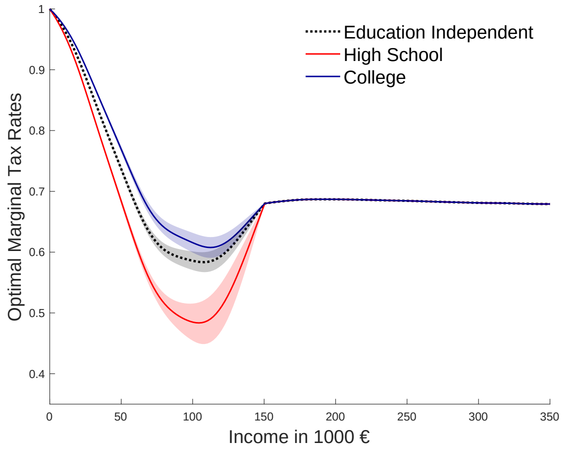
Optimal Marginal Tax Rates Figure 2 displays the optimal tagged marginal tax rates for college and high school graduates as a function of annual gross income up to 350 000 €. For comparison, the optimal education-independent marginal tax rates are illustrated as well. The corresponding 95% confidence intervals are displayed as bands in a lighter shade.

In all countries, the optimal tagged and education-independent marginal tax rates follow a U-shape pattern, which is consistent with the literature on optimal income taxation. In general, the inverse hazard rate $\frac{1-F_{e_i}(\theta)}{f_{e_i}(\theta)\theta}$ plays an important role in determining the optimal tax schedule. In particular, $1 - F_{e_i}(\theta)$ captures the mass of individuals from whom the absolute tax payment increases and $f_{e_i}(\theta)\theta$ describes the mass of individuals which labor supply is distorted by a small increase in the marginal tax rate. At low income levels, the share of individuals is large whose absolute tax payments increase. At the same time, the mass of individuals is small whose labor supply is distorted. This implies that the inverse hazard rate is rather high and that the mechanical revenue gain dominates the behavioral revenue loss. Thus, it is optimal to set high marginal tax rates at low income levels. Up to an income of around 100 000 €, optimal marginal tax rates are decreasing. With increasing income, the mass of individuals with distorted labor supply increases making the behavioral revenue loss more and more severe. At the same time, the revenue gains from the mechanical effect decrease. Afterwards, marginal tax rates start to converge to a constant tax rate. At an income of around 150 000 € the marginal tax rates stay constant on a level of 50-70%.

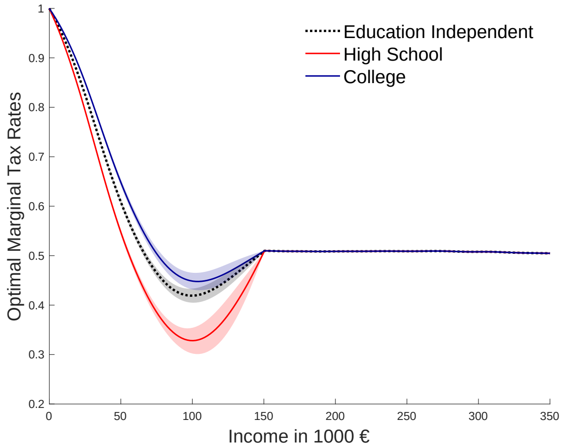
The optimal education-dependent marginal tax rates enclose the education-independent optimal tax rate until the Pareto tail starts at an income of around 150 000 €. Up to an income of 150 000 €, college graduates always face a higher marginal tax rate than high school graduates. Importantly, the finding of a higher college tax rate is robust across all countries. The intuition is as follows: The relative revenue gain from a small increase of the marginal tax rate is larger in the college group since the college hazard rate is relatively larger. Thus, it is optimal to set a higher marginal tax rate for college graduates. For top incomes, college and high school graduates face the same marginal tax rate as the Pareto tail is assumed to be education-independent.



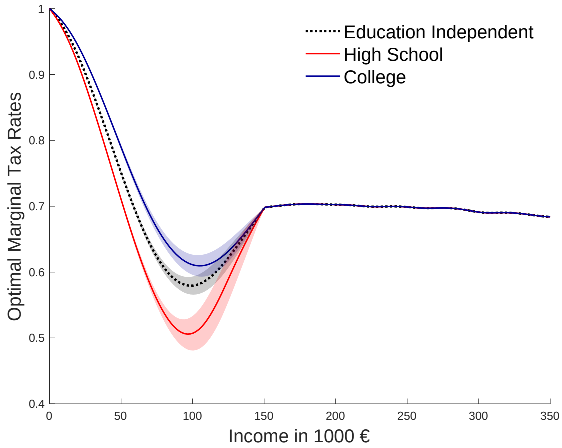
(a) Germany



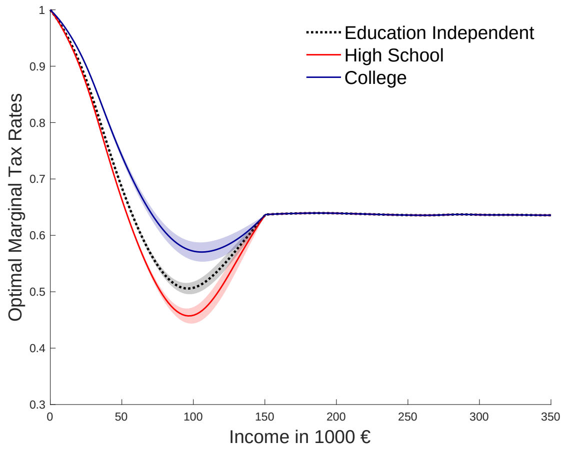
(b) United Kingdom



(c) Spain



(d) France



(e) Italy

Figure 2: Optimal Marginal Tax Rates Conditional on Education
Notes: Confidence intervals are illustrated as bands in a lighter shade and are calculated on a 5% significance level by bootstrapping with 500 replications.

In France and Germany, the simulated optimal top tax rates have a jagged shape. This results from the approximation of the current income tax system by simulating effective marginal tax rates with the microsimulation model EUROMOD. Restricting the current income tax schedule to a HSV functional form yields smooth optimal marginal tax rates across all countries.

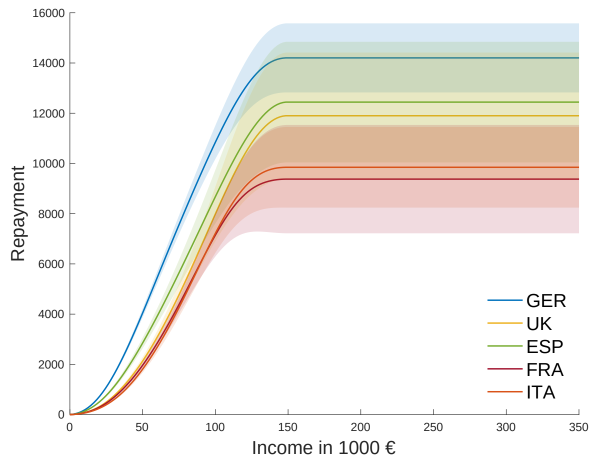
Our quantitative results are very similar to the results of Findeisen and Sachs (2016). They calibrate a dynamic optimal income taxation framework with an endogenous college decision to US data and find that the optimal college tax rate is larger than the one of high school graduates until the Pareto tail kicks in. Then, the pattern reverses and high school graduates face a higher tax rate than college graduates. Though, the difference between both tax rates is rather small. In the top income tail, both tax rates converge to a constant top tax rate of similar magnitude, where the college tax rate is a little bit smaller than the one of high school graduates. This small difference arises due to the endogeneity of college enrollment with respect to the tax schedule. For the social planner, it is optimal to set a slightly smaller tax rate for college graduates to increase the attractiveness of obtaining a college degree. In contrast, we find an identical top tax rate across education groups because education is assumed to be exogenous implying that the effect of education incentive compatibility does not arise. However, Findeisen and Sachs (2016) argue that the effect of endogenous education on marginal tax rates is rather small.

Our finding of a higher marginal tax rate in the college group is also consistent with the quantitative findings of Cremer, Gahvari, and Lozachmeur (2010) who use gender as a tag. They calibrate their model such that the mean income in the male population is higher than in the female population. Cremer, Gahvari, and Lozachmeur (2010) find that men face a higher marginal tax rate and that the simulated tagged marginal tax rates enclose the untagged tax rate at every income level. In all countries, the average income of college graduates is higher than the income of high school graduates. Similar to Cremer, Gahvari, and Lozachmeur (2010), we also find that the group with a higher mean income should face a higher marginal tax rate and that the tagged marginal tax rates are bracketed by the untagged tax rate. However, this finding only holds up to the income level where the Pareto tail starts. Cremer, Gahvari, and Lozachmeur (2010) assume that the income distribution is log-normally distributed over the whole support. In contrast, we append a Pareto tail for top incomes which is in line with the empirical literature on income distributions (see for example Saez (2001) for the US and Jenkins (2017) for the UK).

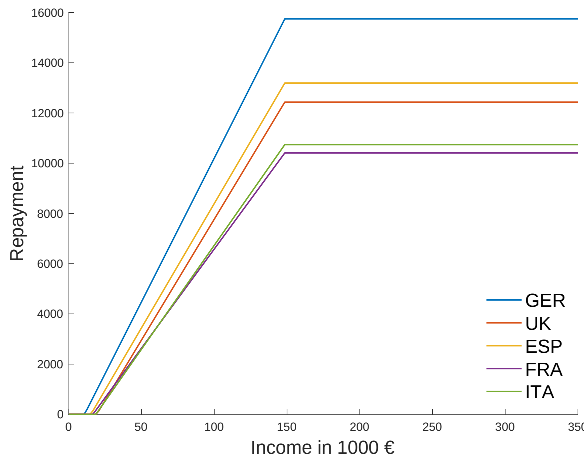
Importantly, our simplified model does not consider the endogeneity of college enrollment as the education decision is assumed to be independent of the tax system. Taking this into account, would lead to a smaller college tax rate and a higher high school tax rate as argued in the theoretical part. This would imply that high school and college graduates would face a slightly different top tax rate as in Findeisen and Sachs (2016). However, Findeisen and Sachs

(2016) argue that the additional term arising due to the endogeneity of college enrollment is rather small.

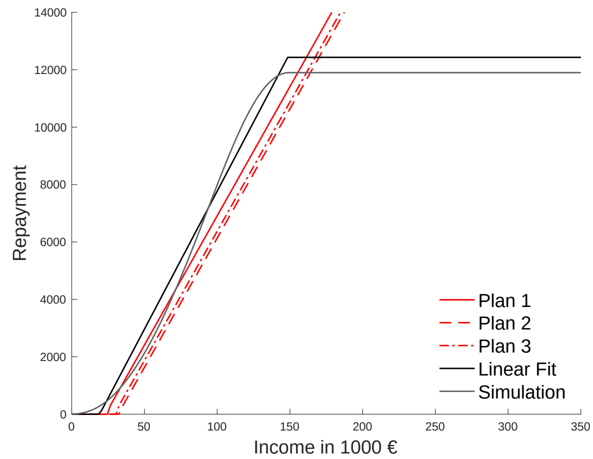
Simulated Repayment Schedule Next, we analyze the optimal income-contingent repayment schedules implied by the applied framework. As outlined in Section 2, the education-dependent tax schedules can be interpreted as the repayment of an income-contingent student loan. This would imply that the high school tax schedule is the common labor income tax schedule and the repayment is the difference in the tax payment under the college and high school tax schedule, i.e. $R(y) = T_{e_{col}}(y) - T_{e_{hs}}(y)$. Figure 3(a) illustrates the simulated repayment schedules as a function of annual income up to 350 000 € for all five countries.



(a) Simulated Repayment Schedule



(b) Approximated Repayment



(c) Current Repayment Schedules in the UK

Figure 3: Simulated and Approximated Repayment Schedules

Notes: Confidence intervals are illustrated as bands in a lighter shade and are calculated on a 5% significance level by bootstrapping with 500 replications. In Panel c, the corresponding 95% confidence interval of the simulated repayment schedule is not displayed for the sake of greater graphical clarity. The red lines illustrate the different repayment plans currently offered by the government of the United Kingdom. The current repayment schedules in the UK are plotted based on the repayment rate and minimum income thresholds published by `https://www.gov.uk/repaying-your-student-loan`. The values of the minimum income thresholds are converted into €

The slope of the repayment schedule is given by the difference in the education-dependent marginal tax rates. Across all countries, the repayment increases in income up to the point where the Pareto tail starts. Afterwards, the repayment stays constant. Except for very low incomes, the slope of the repayment schedule is almost linear as the difference between the college and high school optimal tax rate is nearly constant. The constant repayment for top incomes is a consequence of identical top tax rates which results from the assumption of an identical Pareto tail across all education groups. The intercept of the repayment schedule is given by the difference in the lump-sum transfers across education groups. Here, the lump-sum transfer is assumed to be identical in both groups implying that the optimal repayment for individuals with a zero income is zero as well, i.e. $R(0) = 0$. Figure 3(a) shows growing confidence intervals with increasing income. This is caused by an accumulation of sample uncertainties in the process of calculating the absolute tax payments.

In all countries, the simulated repayment schedule can be well approximated by a schedule with a minimum income threshold and a repayment linearly increasing in income up to 150 000 € and stays constant afterwards. The approximated repayment schedules are illustrated in Figure 3(b). The minimum income threshold ranges between 10 780 € in Germany and 19 173 € in the UK. Individuals with an income below a minimum income threshold do not need to repay their loan. Then, the repayment is linearly increasing in income with a constant repayment rate of 8-11%. For incomes above roughly 150 000 €, the repayment stays constant. Table 3 summarizes the details of the approximated repayment schedules in all countries.

Table 3: Properties of the Approximated Repayment Schedules

	Germany	UK	Spain	France	Italy
Repayment Rate	11.4 %	9.6 %	9.9 %	7.9 %	8.3 %
Minimum Income Threshold	10 780 €	19 173 €	15 140 €	16 539 €	18 587 €
Repayment Cap	148 463 €	148 463 €	148 463 €	148 463 €	148 463 €

Notes: The repayment rate is estimated by a linear approximation of the simulated repayment schedule (ordinary least squares). The minimum income threshold corresponds to the income level where the repayment, based on the approximated repayment schedule, equals zero. The repayment cap corresponds to the income level of the maximum of the simulated repayment.

Within the set of countries, the UK is the only country providing income-contingent student loans. At the moment, the government offers three different plans for undergraduate students. Students cannot choose one repayment plan voluntarily as the eligibility depends on the start year of their undergraduate studies. All three plans have a minimum income threshold where graduates do not need to pay anything back. The exact amount varies between the different plans. The threshold of the first, second and third plan is 23 200 €, 31 840 € and 29 165 €, respectively. The repayment rate is identical across all three plans and set to 9% of gross income above the minimum income threshold.[21] It is a natural exercise to compare the implied repayment schedule of our framework with the actual schedule. Figure 3(c) shows the simulated and approximated repayment schedule as well as all three effective schedules. Our approximated repayment rate is almost identical to the effective repayment rate of 9%. Only the minimum income thresholds differ slightly where the thresholds in the UK are slightly higher. However, all repayment plans are uncapped implying that the repayment is always increasing in income.

In summary, the quantification of our simplified model shows that college graduates should face a higher marginal tax rate up to the income level where the Pareto tail starts. For top incomes, college and high school graduates should face identical marginal tax rates. The implied repayment schedule can be well approximated by a simple schedule with a minimum income threshold, a repayment that increases linearly in income and a cap on repayments. Further, the approximated repayment schedule comes close to the effective repayment schedule in the UK. The only main difference is that annual repayments are not capped.

4 Extensions and Robustness

Now, we evaluate the sensitivity of our results to alternative specifications of the model. Additional details and the simulated optimal education-dependent marginal tax rates as well as the implied repayment schedule are included in Appendix B. All results are summarized in Table 1.

4.1 Ternary Tag

So far, we have only simulated optimal marginal tax rates using a binary tag. Next, we consider that college graduates differ in their obtained education degrees. In particular, we simulate optimal marginal tax rates for master, bachelor and high school graduates in Germany.[22] Figure B1 in Appendix B.1 illustrates optimal education-dependent marginal tax rates with a ternary tag and the implied repayment schedule for bachelor and master graduates. College graduates with a master degree face the highest marginal tax rate up to the point where the Pareto tail starts, though the difference to the bachelor tax rate is rather small. College graduates with a bachelor degree still face a larger tax rate than high school graduates. The top tax rate is identical across all three education groups. Now, the applied framework implies a repayment schedule conditional on the obtained education degree. The slope of the repayment schedule

[21] See `https://www.gov.uk/repaying-your-student-loan` for details on the different repayment plans. The values of the minimum income thresholds are converted into €.

[22] A precise simulation in the other countries is not possible due to data limitations in EU-SILC.

for college graduates with a master degree is slightly steeper. This entails that the repayment of college graduates with a master degree is larger compared to those with a bachelor degree.

4.2 Parametric Calibration of the Income Distribution

Instead of applying a kernel density estimation for smoothing the income distribution, we apply a parametric approach. See Appendix B.2 for details on the parametric calibration of the income distributions. Across all countries, college graduates face a larger marginal tax rate than high school graduates up to the point where the Pareto tail starts. In France, the pattern is slightly different. Here, the optimal education-dependent marginal tax rates intersect before the Pareto tail starts. This implies that college graduates face a smaller marginal tax rate for a small income range before the Pareto tail begins. Across all countries, the top tax rate is identical for high school and college graduates. Moreover, the implied repayment schedules are similar to the schedules of the baseline simulation. Figure B2 in Appendix B.2.1 illustrates the optimal education-dependent marginal tax rates and the repayment schedule with a parametric calibration of the income distribution.

Further, we calibrate the income distributions by applying the parametric approach from above, but without appending a Pareto tail. Now, college graduates face a larger marginal tax rate over the whole income range. As before, the finding is slightly different in France. Here, the optimal education-dependent marginal tax rates intersect at an income level of around 80 000 €. For incomes below 80 000 € college graduates face a larger marginal tax rate and for incomes above 80 000 € high school graduates face a larger marginal tax rate. By now, the implied repayment schedule is increasing in income over the whole income range. In France, the smaller marginal tax rates for college graduates for incomes above 80 000 € lead to a repayment schedule that is decreasing in income and negative. The simulated optimal education-dependent marginal tax rates and the repayment schedule are summarized in Figure B3 in Appendix B.2.2.

4.3 Choice of Bandwidth in the Kernel Density Estimation

The key element of the non-parametric calibration of the income distributions is a kernel density estimation which requires the specification of a bandwidth. This specification has a significant influence on the shape of the simulated optimal marginal tax rates.[23] To analyze the influence of different bandwidths on the shape of optimal marginal tax rates, we compare simulated marginal tax rates from a true income distribution with simulated marginal tax rates from a

[23]The kernel density estimation requires the choice of a kernel as well. We do not investigate the choice of the kernel on the shape of optimal marginal tax rates as the same degree of smoothness of two kernel density estimators with different bandwidths can be achieved by adjusting the bandwidths.

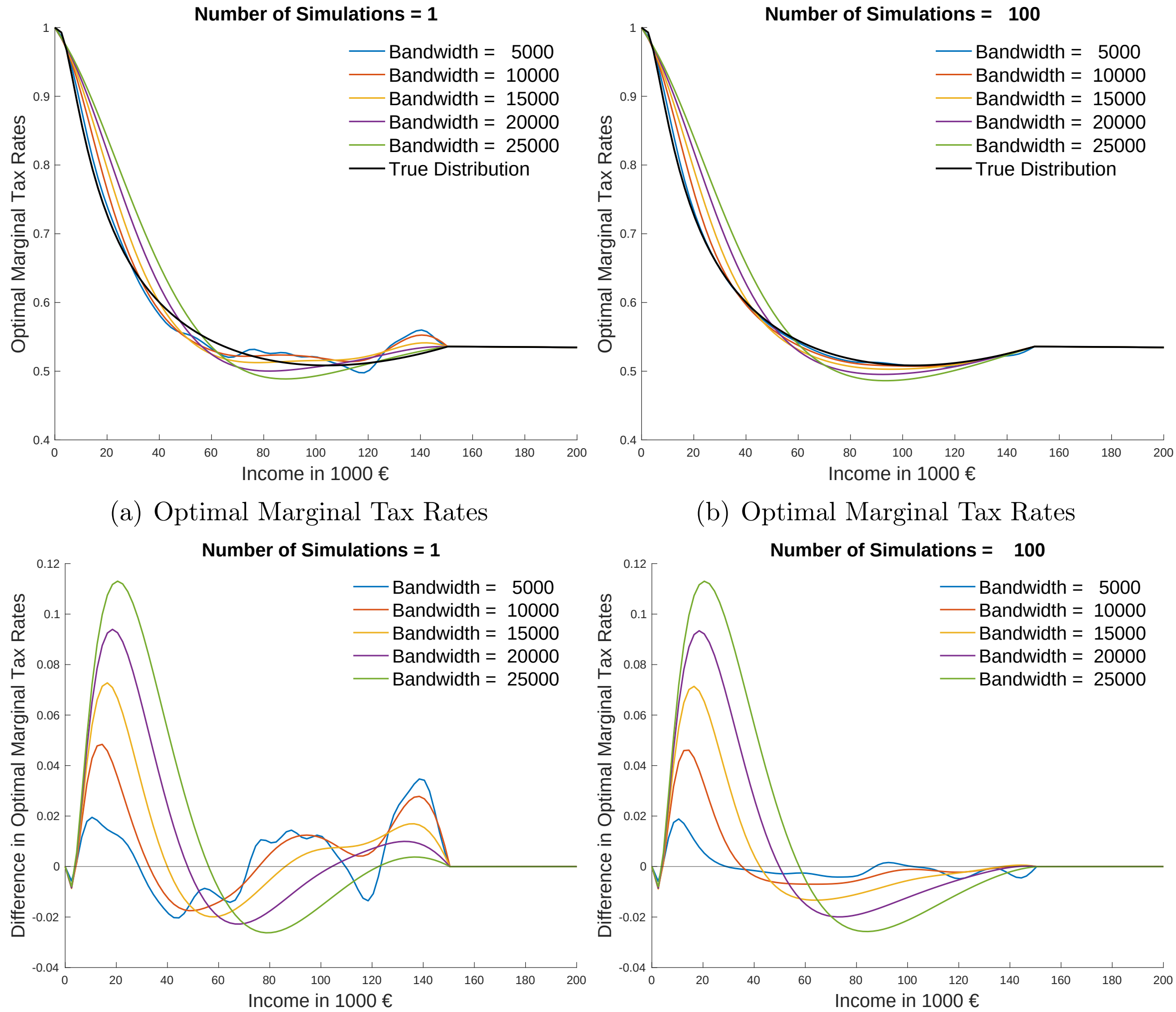
sample of incomes drawn from the true income distribution. Details on the calibration of the true income distribution and the random sample are described in Appendix B.3.

Figure 4 illustrates simulated marginal tax rates for five different sizes of the bandwidth ranging from 5 000 € to 25 000 € and compares them with the marginal tax rates from the true income distribution. The left panel presents the results from one simulation and the right hand side presents the average over 100 simulations. The simulated tax schedules with a bandwidth of 5 000 € and 10 000 € have a jagged shape at intermediate incomes whereas the simulated tax schedules with larger bandwidths have a smooth shape over the whole income range. When taking the average of over 100 simulations, the shape gets smoother even for the two lowest bandwidths. Panel c and d show the difference between the simulated and true tax schedule. For low incomes, marginal tax rates simulated from a random sample of income are larger than the true marginal tax rates. This pattern reverses for intermediate incomes up to the point where the Pareto tail starts. Across all 100 simulations, the smallest bandwidth of 5 000 € performs best. Note, that the bias increases with the size of the bandwidth but at the same time the smoothness of the tax schedule increases. This implies that there is a trade-off between simulating an accurate tax schedule and simulating a smooth tax schedule.

Optimal bandwidth choice has not received much attention in the literature on optimal income taxation despite the non negligible influence on the shape of simulated optimal marginal tax rates.[24] Lockwood (2020) and Choné and Laroque (2010) note that a larger bandwidth is useful to get a smooth tax schedule. However, a larger bandwidth increases the bias between the simulated and true tax schedule. In general, the bandwidth does not significantly impact the general finding of a U-shape pattern. Though, the specified bandwidth can matter for policy implications as in the case of education-dependent marginal tax rates. Here, the choice of the bandwidth can make a difference whether the college tax rate lies above the high school tax rate or not. Figure 5 displays optimal marginal tax rates simulated with a smaller bandwidth in France and Italy. For comparison, the optimal tax schedule of the baseline simulation is illustrated in a lighter shade. Using a smaller bandwidth reveals the same pattern as in the simulation exercise before. A smaller bandwidth leads to smaller marginal tax rates for lower incomes and higher marginal tax rates for intermediate incomes. In these two countries, high school graduates face a slightly larger tax rate for a small income range before the Pareto tail starts. Now, the repayment slightly decreases before staying constant. In addition, a trade-off between accuracy and precision of simulated marginal tax rates can be observed. A smaller bandwidth increases the accuracy of the tax schedule but also increases the variance and thus, confidence intervals get larger. In all other countries, a smaller bandwidth does not change the

[24]Beresteanu and Dahan (2002) are an exemption who employ an adaptive bandwidth with a cross validation procedure. However, they do not discuss the role of different bandwidths on the shape of the inverse hazard rate or simulated marginal tax rates.

main finding of the baseline simulation. Figure B4 in Appendix B.3 show optimal marginal tax rates conditional on education and the implied repayment schedule with a lower specified bandwidth than in the baseline simulation for all countries.



(a) Optimal Marginal Tax Rates

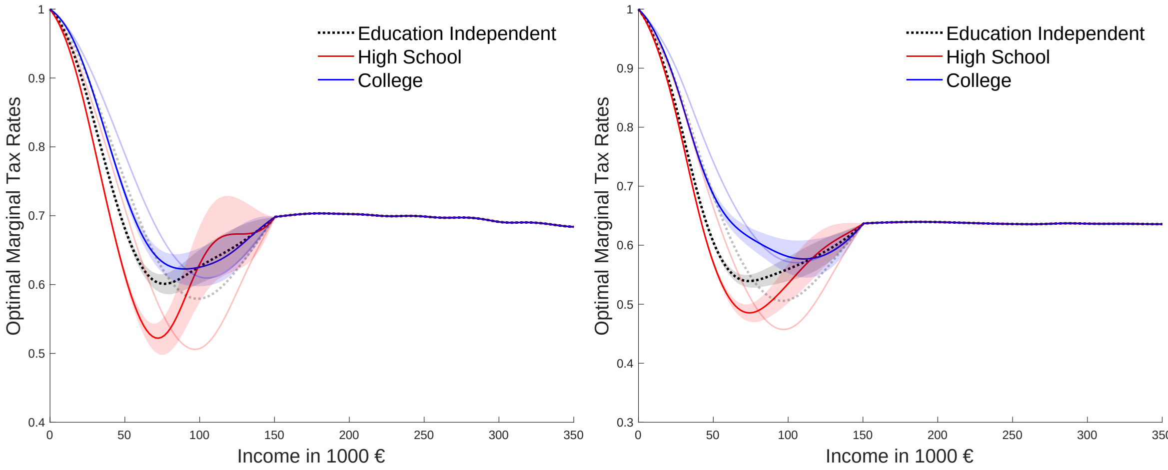
(b) Optimal Marginal Tax Rates

(c) Difference between True and Simulated Optimal Marginal Tax Rates

(d) Difference between True and Simulated Optimal Marginal Tax Rates

Figure 4: Optimal Marginal Tax Rates under Different Values of the Bandwidth

Notes: The left panels show the simulated optimal marginal tax rates and the differences between the true and simulated optimal marginal tax rates for one simulation. The right panels show the average across 100 simulations.



(a) France (b) Italy

Figure 5: Optimal Marginal Tax Rates Conditional on Education and Repayment Schedule and with a Smaller Bandwidth

Notes: Confidence intervals are illustrated as bands in a lighter shade and are calculated on a 5% significance level by bootstrapping with 500 replications. The simulated optimal marginal tax rates of the baseline simulation are displayed in a lighter shade. The corresponding 95% confidence intervals of the baseline simulation and the repayment schedule are not displayed for the sake of greater graphical clarity. The kernel density estimation uses a bandwidth of 20,000 €.

4.4 Further Robustness

Male Sub-Sample To limit potential interactions between education and gender in determining incomes after graduation, we restrict our samples of incomes to male individuals. This eliminates the potential threat that education is used as an indirect instrument of implementing a gender based taxation. Our results of Section 3.2 are not sensitive to this restriction.

Country-Specific Pareto Threshold Instead of assuming an identical Pareto threshold of 150 000 € across countries, we use the country-specific estimates of the Pareto threshold by Bartels and Metzing (2019), which are displayed in Table B2 in Appendix B.4. The main finding of a higher marginal tax rate for college graduates and an identical top tax rate is still present. The repayment is still non-negative, increasing in income and finite.

More Recent Values of the Pareto Parameter Now, we use more recent country-specific values for the Pareto parameter estimated by Bartels and Metzing (2019). Table B2 in Appendix B.4 contains the estimates from Bartels and Metzing (2019) and the baseline values from Atkinson, Piketty, and Saez (2011). In general, both estimated Pareto parameters are very similar. Only in Spain and France, the more recent estimates by Bartels and Metzing (2019) are a little bit smaller implying that the top income tail got thicker over time. Intuitively, these two countries are the only countries where the simulated optimal marginal tax rates slightly differ from the baseline simulation. The main finding from the baseline simulation of Section 3.2 is reconfirmed.

HSV Tax Function Instead of simulating current effective marginal tax rates with EUROMOD, we assume that the current income tax schedules can be approximated by a functional form that was proposed by Feldstein (1969) and most recently applied by Bénabou (2002) and Heathcote, Storesletten, and Violante (2017). Hence, we denote tax revenues as a function of gross income:

$$T(y) = y - \lambda y^{1-\tau}, \tag{15}$$

where τ refers to the degree of tax progressivity and λ determines the average level of taxation in the economy. Table B3 in Appendix B.5 contains the country-specific values for τ and λ. Our main findings are not sensitive to the specification of the current income tax schedule.

Identical Elasticities across Countries Instead of using country-specific values for the elasticity of taxable income, we use the midpoint estimate of $\epsilon = 0.25$ by Saez, Slemrod, and Giertz (2012) for all countries. The simulated optimal marginal tax rates do not change in the UK and Italy as the country-specific elasticity is identical to the midpoint elasticity. In Germany and Spain, the tax schedule shifts upwards and in France downwards. An upward (downward) shift is observed in Germany and Spain (France) because the country-specific elasticity is larger (smaller) than the midpoint estimate. The tagged optimal marginal college and high school tax rates shift by almost the same amount implying that college graduates still face a larger marginal tax rate until the Pareto tail starts. The shape of the repayment schedules is almost identical to the schedule of the baseline simulation.

Now, optimal marginal tax rates can be easily compared across countries. Figure B6 in Appendix B.6 illustrates the optimal education-dependent marginal tax rates for each education group separately, the education-independent marginal tax rates and the repayment schedules. Across all countries, the shape of the simulated marginal tax rates is almost identical. The top tax rates slightly differ across countries, where Germany has the highest top tax rate and Italy the lowest one. This can be traced back to the assumption of a country-specific Pareto tail as differences in the top tax rate arise now only due to different values of the Pareto parameter. Across all countries, Germany has the thickest top income tail and Italy the thinnest tail. As in the baseline simulation, the repayment schedules are very similar across all countries.

Education-Dependent Elasticities So far, we have assumed that college and high school graduates have identical elasticities of taxable income. This is in line with the empirical literature on labor supply elasticities which either finds similar labor supply elasticities or smaller elasticities for more educated individuals. For example, Bargain, Orsini, and Peichl (2014) find no statistically significant differences in wage elasticities across education groups. Keane and Wasi (2016) find that the compensated and uncompensated labor supply elasticities are almost similar for college and high school graduates over the life-cycle. Further, they find a lower Frisch elasticity of labor supply for college graduates over the life-cycle. Blundell, Costa Dias, Meghir, and Shaw (2016) find a lower Frisch and Marshallian labor supply elasticity for women with tertiary education over the whole life-cycle.

Assuming a lower elasticity for college graduates would shift the college marginal tax rate further up leading to even larger marginal and top tax rates for college graduates. If college graduates had a higher elasticity than high school graduates, the college tax schedule would shift downwards. Then, it may occur that college graduates face a smaller marginal tax rate than high school graduates. The intuition is that the behavioral revenue loss from a small increase in the marginal tax rate increases as college graduates reduce their labor supply to a larger extent. Hence, it is optimal to set a lower college tax rate.

Therefore, we increase the elasticity for college graduates such that the college marginal tax rate is less than or equal to the high school marginal tax rate. This gives us an approximation by how much the elasticity of college graduates can be larger such that the social planner will not find it optimal to set a higher tax rate for college graduates over the whole income range. On average, the elasticity of college graduates needs to be larger by a factor of 1.7. The country-specific factors as well as the corresponding elasticities from this scenario are presented in Table B4 in Appendix B.6.1. Figure B7 in Appendix B.6.1 illustrates the optimal education-dependent tax schedules with the elasticities from Table B4.

5 Conclusion

In this paper, we explore the optimal design of education-dependent income taxes and their application to income-contingent student loans. For doing so, we develop an optimal income taxation framework in the spirit of Mirrlees (1971) where the government uses the observed education degree as a tag. We calibrate the model to five European countries. Optimal education-dependent marginal tax rates as well as the implied repayment schedule of an income-contingent student loan are simulated. The main novelty of the quantitative part is that we account for finite sample limitations by computing confidence intervals based on a bootstrapping approach. Further, we carefully evaluate the robustness of the simulated education-dependent tax schedules to various alternative calibration specifications of the model.

The main finding of the theoretical part is that the classical formula for income taxes from Diamond (1998) applies within each education group. Under a Rawlsian planner and identical elasticities, the education group with the larger inverse hazard rate faces the larger marginal tax rate. The same applies to the top tax rate. The education group with the thicker top income tail faces the larger top tax rate.

The main findings of the quantitative part are that college graduates should face a higher marginal tax rate until the income level where the Pareto tail starts and that college and high school graduates should face an identical top tax rate. The repayment schedule comes close to a schedule where (i) repayment strongly increases in income, (ii) a minimum income threshold exits and (iii) repayment is capped. This is in line with the properties of repayment schedules in the real-world. Importantly, these findings are robust across countries and not sensitive to several alternative calibration specifications of the model. Only in France and Italy, the finding slightly changes for some alternative specifications. In these specifications, high school graduates should face a larger marginal tax rate in a small income region before the Pareto tail starts. The implications for optimal tax and education policy are that college graduates should face a larger marginal tax rate from a revenue maximizing perspective. The top tax rate should be the same for college and high school graduates. This can be easily implemented by an income-contingent student loan as has been realized in the United Kingdom. Our results show that income-contingent repayment schedules are optimal even when distortionary effects of taxes are taken into account.

Our paper has abstracted from several aspects that can be addressed in future research. First, we only consider a static model implying that forgone earnings during college and borrowing constraints cannot be captured. Second, we limit the quantitative exploration to the case of exogenous college enrollment and a Rawlsian social planner. Future work could extend the theoretical framework to a dynamic setting and extend the quantitative exploration to account for endogenous college enrollment. Further, the optimal education-dependent tax schedules could be simulated under different assumptions of redistributive preferences. For example, redistributive preferences could be encoded through declining marginal social welfare weights where high school graduates receive a larger weight. Alternatively, an inverse-optimum approach could be employed which has the advantage that it does not rely on assumed redistributive preferences Bourguignon and Amedeo (2012). Another fruitful area for future research would be the analysis of the effects of skill-biased technological change on education-dependent tax schedules and the repayment schedule of income-contingent student loans.

References

Akerlof, G. A. (1978): “The Economics of "Tagging" as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning,” *The American Economic Review*, 68(1), 8–19.

Almunia, M., and D. Lopez-Rodriguez (2019): “The Elasticity of Taxable Income in Spain: 1999-2014,” *SERIEs*, 10, 281–320.

Anderberg, D. (2009): “Optimal Policy and the Risk Properties of Human Capital Reconsidered,” *Journal of Public Economics*, 93(9), 1017–1026.

Atkinson, A., A.-C. Guio, and E. Marlier (2017): “Monitoring Social Inclusion in Europe,” .

Atkinson, A. B., T. Piketty, and E. Saez (2011): “Top Incomes in the Long Run of History,” *Journal of Economic Literature*, 49(1), 3–71.

Ayaz, M., L. Fricke, C. Fuest, and D. Sachs (2021): “Who Should Bear the Burden of Covid-19 Related Fiscal Pressure? An Optimal Income Taxation Perspective,” ifo Working Paper No. 358.

Bargain, O., K. Orsini, and A. Peichl (2014): “Comparing Labor Supply Elasticities in Europe and the United States: New Results,” *The Journal of Human Resources*, 49(3), 723–838.

Barr, N. (1999): “Higher Education Finance: Lessons from International Experience,” .

Barr, N., B. Chapman, L. Dearden, and S. Dynarski (2019): “The US College Loans System: Lessons from Australia and England,” *Economics of Education Review*, 71, 32–48, Higher Education Financing: Student Loans.

Bartels, C., and M. Metzing (2019): “An Integrated Approach for a Top-Corrected Income Distribution,” *The Journal of Economic Inequality*, 17, 125–143.

Beresteanu, A., and M. Dahan (2002): “An Optimal Shape of Income Tax: Evidence from Zero Income Tax Countries- Paraguay Uruguay,” .

Bierbrauer, F. J., P. C. Boyer, and A. Peichl (2021): “Politically Feasible Reforms of Nonlinear Tax Systems,” *American Economic Review*, 111(1), 153–91.

Blundell, R., M. Costa Dias, C. Meghir, and J. Shaw (2016): “Female Labor Supply, Human Capital, and Welfare Reform,” *Econometrica*, 84(5), 1705–1753.

Boadway, R., and P. Pestieau (2006): “Tagging and Redistributive Taxation,” *Annales d’Économie et de Statistique*, (83/84), 123–147.

Boháček, R., and M. Kapička (2008): “Optimal Human Capital Policies,” *Journal of Monetary Economics*, 55(1), 1–16, Carnegie-Rochester Conference Series on Public Policy: Advances in Dynamic Public Finance April 20-21, 2007.

Bourguignon, F., and S. Amedeo (2012): “Tax–Benefit Revealed Social Preferences,” *The Journal of Economic Inequality*, 10(1), 75–108.

Bovenberg, L., and B. Jacobs (2005): “Redistribution and Education Subsidies are Siamese Twins,” *Journal of Public Economics*, 89(11), 2005–2035.

Brewer, M., E. Saez, and A. Shepard (2010): *Means Testing and Tax Tates on Earnings*Oxford University Press.

Britton, J., and J. Gruber (2020): “Do Income Contingent Student Loans Reduce Labor Supply?,” *Economics of Education Review*, 79, 102061.

Britton, J., L. van der Erve, and T. Higgins (2019): “Income Contingent Student Loan Design: Lessons From Around the World,” *Economics of Education Review*, 71, 65–82, Higher Education Financing: Student Loans.

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

Chapman, B. (2006): *Government Managing Risk Income Contingent Loans for Social and Economic Progress*vol. 2 of *Handbook of the Economics of Education*, pp. 1435–1503. Routledge.

Chapman, B., and A. Leigh (2009): “Do Very High Tax Rates Induce Bunching? Implications for the Design of Income Contingent Loan Schemes,” *Economic Record*, 85(270), 276–289.

Choné, P., and G. Laroque (2010): “Negative Marginal Tax Rates and Heterogeneity,” *American Economic Review*, 100(5), 2532–47.

Cremer, H., F. Gahvari, and J.-M. Lozachmeur (2010): “Tagging and Income Taxation: Theory and an Application,” *American Economic Journal: Economic Policy*, 2(1), 31–50.

Diamond, P. A. (1998): “Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates,” *The American Economic Review*, 88(1), 83–95.

Doerrenberg, P., A. Peichl, and S. Siegloch (2017): “The Elasticity of Taxable Income in the Presence of Deduction Possibilities,” *Journal of Public Economics*, 151, 41–55.

Dynarski, S., and D. Kreisman (2013): “Loans for Educational Opportunity: Making Borrowing Work for Today’s Students,” .

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

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

Friedman, M. (1955): “The Role of Government in Education,” Rutgers University Press.

García-Miralles, E., N. Guner, and R. Ramos (2019): “The Spanish Personal Income Tax: Facts and Parametric Estimates,” *SERIEs*, 10, 439–477.

Gary-Bobo, R. J., and A. Trannoy (2015): “Optimal Student Loans and Graduate Tax under Moral Hazard and Adverse Selection,” *The RAND Journal of Economics*, 46(3), 546–576.

Golosov, M., A. Tsyvinski, and N. Werquin (2014): "A Variational Approach to the Analysis of Tax Systems," NBER Working Paper No. 20780.

Gruber, J., and E. Saez (2002): "The Elasticity of Taxable Income: Evidence and Implications," *Journal of Public Economics*, 84(1), 1–32.

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

Herbst, D. (2020): "Liquidity and Insurance in Student-Loan Contracts: The Effects of Income-Driven Repayment on Borrower Outcomes," .

Holter, H. A., D. Krueger, and S. Stepanchuk (2019): "How Do Tax Progressivity and Household Heterogeneity Affect Laffer Curves?," *Quantitative Economics*, 10(4), 1317–1356.

Immonen, R., R. Kanbur, M. Keen, and M. Tuomala (1998): "Tagging and Taxing: The Optimal Use of Categorical and Income Information in Designing Tax/Transfer Schemes," *Economica*, 65(258), 179–192.

Jacquet, L., and E. Lehmann (2020): "Optimal Income Taxation with Composition Effects," *Journal of the European Economic Association*, 19(2), 1299–1341.

Jacquet, L., E. Lehmann, and B. Van der Linden (2013): "Optimal Redistributive Taxation with Both Extensive and Intensive Responses," *Journal of Economic Theory*, 148(5), 1770–1805.

Jara, H. X., and A. Tumino (2013): "Tax-Benefit Systems, Income Distribution and Work Incentives in the European Union," *International Journal of Microsimulation*, pp. 27–62.

Jenkins, S. P. (2017): "Pareto Models, Top Incomes and Recent Trends in UK Income Inequality," *Economica*, 84(334), 261–289.

Kapička, M., and J. Neira (2019): "Optimal Taxation with Risky Human Capital," *American Economic Journal: Macroeconomics*, 11(4), 271–309.

Keane, M. P., and N. Wasi (2016): "Labour Supply: The Roles of Human Capital and The Extensive Margin," *The Economic Journal*, 126(592), 578–617.

Kindermann, F., L. Mayr, and D. Sachs (2020): "Inheritance Taxation and Wealth Effects on the Labor Supply of Heirs," *Journal of Public Economics*, 191, 104127.

Krueger, A. B., and W. G. Bowen (1993): "Policy Watch: Income-Contingent College Loans," *Journal of Economic Perspectives*, 7(3), 193–201.

Lehmann, E., F. Marical, and L. Rioux (2013): "Labor Income Responds Differently to Income-Tax and Payroll-Tax Reforms," *Journal of Public Economics*, 99, 66–84.

Lochner, L., and A. Monge-Naranjo (2016): *Student Loans and Repayment: Theory, Evidence, and Policy*vol. 5 of *Handbook of the Economics of Education*, pp. 397–478. Elsevier.

Lockwood, B. B. (2020): "Optimal Income Taxation with Present Bias," *American Economic Journal: Economic Policy*, 12(4), 298–327.

MANKIW, N. G., AND M. WEINZIERL (2010): "The Optimal Taxation of Height: A Case Study of Utilitarian Income Redistribution," *American Economic Journal: Economic Policy*, 2(1), 155–76.

MANKIW, N. G., M. WEINZIERL, AND D. YAGAN (2009): "Optimal Taxation in Theory and Practice," *Journal of Economic Perspectives*, 23(4), 147–74.

MATSUDA, K., AND K. MAZUR (2020): "College Education and Income Contingent Loans in Equilibrium: Theory and Quantitative Evaluation," .

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

NICHOLS, A. L., AND R. J. ZECKHAUSER (1982): "Targeting Transfers through Restrictions on Recipients," *The American Economic Review*, 72(2), 372–377.

OECD (2009): "OECD Employment Outlook: Tackling the Job Crisis," .

——— (2021): "Long-Term Unemployment Rate (Indicator)," Accessed on 16 April 2021.

PIKETTY, T. (1997): "La Redistribution Fiscale Face au Chomage," *Revue francaise d'economie*, 12, 157–201.

PIKETTY, T., AND E. SAEZ (2013): "Optimal Labor Income Taxation," *Handbook of Public Economics*, Volume 5, 391–474.

RAJ, A. (2019): "Optimal Income Taxation in the Presence of Networks of Altruism," .

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

SACHS, D., A. TSYVINSKI, AND N. WERQUIN (2020): "Nonlinear Tax Incidence and Optimal Taxation in General Equilibrium," *Econometrica*, 88(2), 469–493.

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

——— (2002): "Optimal Income Transfer Programs: Intensive versus Extensive Labor Supply Responses," *The Quarterly Journal of Economics*, 117(3), 1039–1073.

SAEZ, E., J. SLEMROD, AND S. H. GIERTZ (2012): "The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review," *Journal of Economic Literature*, 50(1), 3–50.

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

SUTHERLAND, H., AND F. FIGARI (2013): "EUROMOD: The European Union Tax-Benefit Microsimulation Model," *International Journal of Microsimulation*, 6(1), 4–26.

WAGNER, G. G., J. R. FRICK, AND J. SCHUPP (2007): "The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements," *Journal of Applied Social Science Studies*.

WEINZIERL, M. (2011): "The Surprising Power of Age-Dependent Taxes," *The Review of Economic Studies*, 78(4), 1490–1518.

A Numerical Appendix

A.1 Details on Calibration

A.1.1 Data

For the calibration of the country-specific income distributions conditional on education, we utilize data on labor income and education restricted to individuals with positive income. Furthermore, we limit the sample to the working age population (25-65) and to individuals not being in education currently. Details on the data sources and calibration of each country are outlined below.

Germany For the calibration of the German income distributions conditional on education, we use the 2017 German Socio-Economic Panel v35. The GSOEP is a representative longitudinal household survey on German households that has been conducted since 1984 and contains data on labor income and the education degree of individuals (see Wagner, Frick, and Schupp (2007) for details). We calculate gross labor income as the sum of employment and self-employment income from all jobs in the previous month of the interview. This implies that the reference year of the German income distribution is 2017. To calculate annual incomes, we multiply monthly income by twelve.

To construct the binary education variable, we use the highest reported education level of an individual based on the ISCED Classification. Individuals with tertiary education are considered as college graduates. This includes all individuals with a bachelor, master or doctoral degree. All individuals without tertiary education are considered as high school graduates. In particular, the group of high school graduates contains individuals with a higher vocational education, a vocational education in combination with an A-level, a middle vocational education and general elementary education.

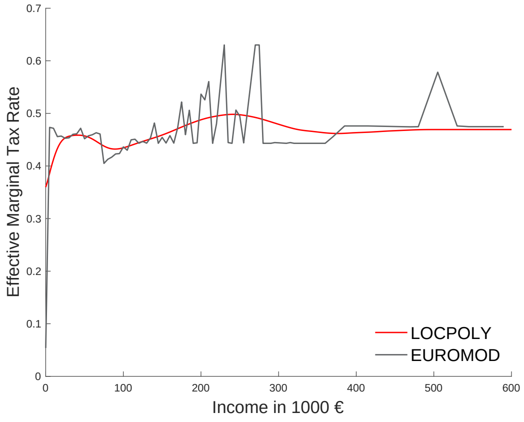
UK, France, Spain and Italy For the calibration of the UK, French, Spanish and Italian income distributions conditional on education, we use the 2018 European Union Statistics on Income and Living Conditions (EU-SILC). The EU-SILC contains annual income data in a harmonized framework which allows for cross-country comparisons (see Atkinson, Guio, and Marlier (2017) for details). We calculate gross labor income as the sum of employment and self-employment. Annual incomes are reported for the previous year of the survey leading to 2017 as the reference year of the income distribution.

To construct the binary education variable, we use the highest education level of an individual based on the ISCED Classification. All individuals with tertiary education are considered as college graduates. This includes all individuals with a bachelor, master or doctoral degree. All individuals without tertiary education are considered as high school graduates. In particular, this includes all individuals with primary education, lower secondary education, upper secondary education post-secondary non-tertiary education and all kinds of vocational education.

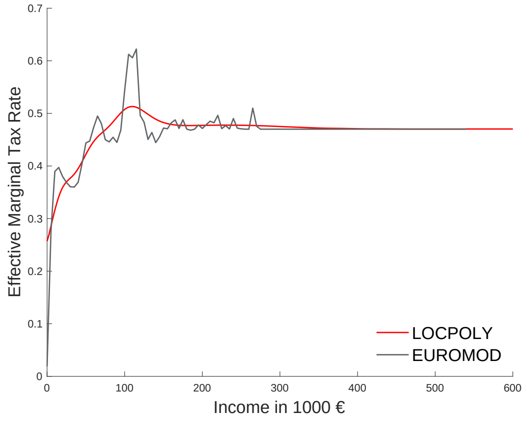
A.1.2 Kernel Density Estimation

To smooth the income distributions obtained from the GSOEP and the EU-SILC, we apply a standard kernel density estimation based on a normal kernel function. For all countries, we use an education-independent income grid with 1000 nodes that are evenly spaced between 2 500 € and 2 000 000 €. As in Lockwood (2020) and Choné and Laroque (2010), we use a large bandwidth for the kernel density estimation to obtain smooth optimal marginal tax rates in the simulations. For all countries, we set the bandwidth to an education-independent value of 30 000 €. As the number of observations is limited in the GSOEP and EU-SILC data, such a large bandwidth is necessary. See Section 4.3 for a detailed description of the impact of the bandwidth on simulated optimal marginal tax rates.

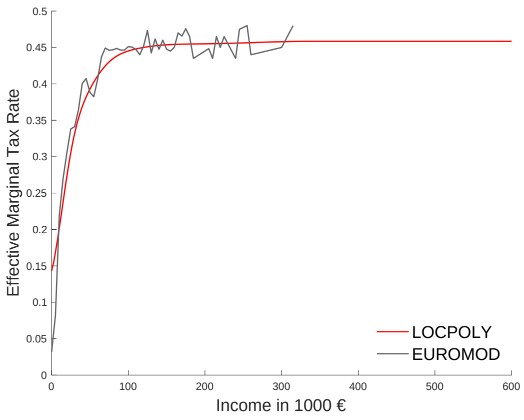
A.1.3 Current Tax Systems



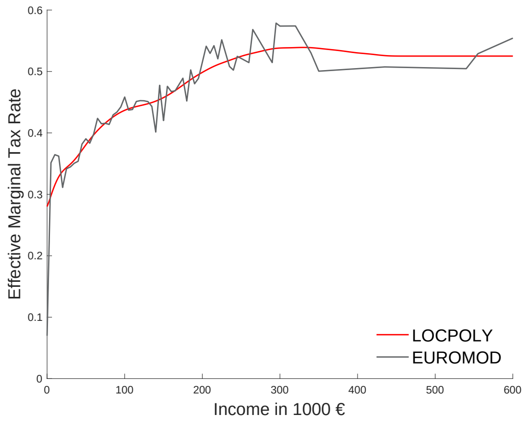
(a) Germany



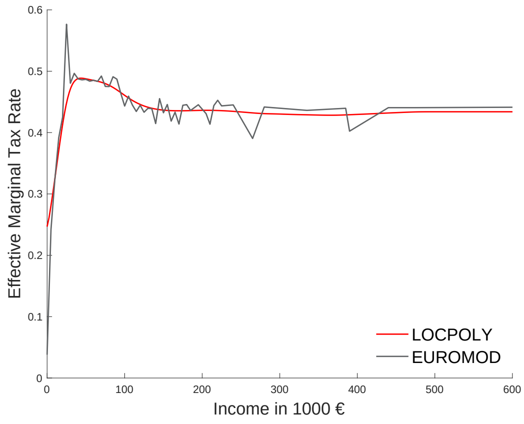
(b) United Kingdom



(c) Spain



(d) France



(e) Italy

Figure A1: Simulated Average and Smoothed Marginal Tax Rates with EUROMOD
Notes: The grey line illustrates the simulated average marginal tax rates with EUROMOD. The red line illustrates the smoothed marginal tax rates by a locally polynomial estimator (LOCPOLY).

A.2 Numerical Simulation Details

Optimal Education-Dependent Marginal Taxes We use a fixed-point algorithm to compute optimal education-dependent tax schedules. For doing so, we discretize the type space and use a discrete version of the formula for education-dependent optimal marginal tax rates from Proposition 1:

$$\frac{T'_{e_i}(y(\theta_i))}{1-T'_{e_i}(y(\theta_i))} = \left(1+\frac{1}{\epsilon}\right)\cdot(\theta_{i+1}-\theta_i)\cdot\left(1-\frac{\frac{1}{\lambda_{e_i}}\sum_{j=i}^{n}U'_j(.)s(\theta_j)}{1-F_{e_i}(\theta_i)}\right)\cdot\frac{1-F_{e_i}(\theta_i)}{f_{e_i}(\theta_i)*\theta_i}.$$

As a starting point, we use the approximation of the current income tax system by EUROMOD as initial tax schedule. Then, the algorithm is as follows:

1. Given the (initial) tax schedule, compute the allocation of the economy. For doing so, we assume an exogenous revenue requirement of 10% of GDP.
2. Given the allocation, compute the optimal tax schedule by using the discrete version of the optimal tax rate formula in Equation (A.2).
3. Update the (initial) tax schedule by the new tax schedule.
4. Repeat steps 1-3 until the tax schedule converges.

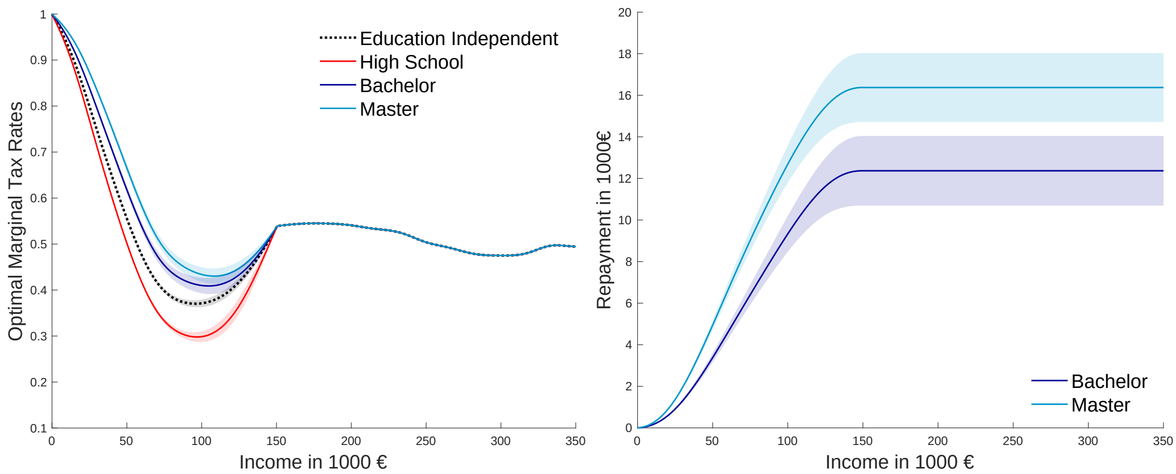
Confidence Intervals I compute 95% confidence intervals by bootstrapping with 500 replications. The proceeding is as follows:

1. Repeatedly draw observations from the sample of incomes conditional on education.[25]
2. Calibrate the conditional income distributions as described in Section 3.1 and simulate the optimal education-dependent marginal tax rates as well as the repayment schedule in every bootstrap sample.
3. Compute the 95% confidence interval from the bootstrap sampling distribution.

[25]The education-independent sample is resampled as follows: First, the samples of incomes conditional on education are resampled independently. Then, the education-independent sample is constructed as the composite of both education-dependent samples. This ensures that the college share is the same across all bootstrap samples.

B Extensions and Robustness Appendix

B.1 Ternary Tag



(a) Optimal Marginal Tax Rates (b) Simulated Repayment Schedule

Figure B1: Optimal Marginal Tax Rates Conditional on Education and Repayment Schedule with a Ternary Tag
Notes: Confidence intervals are illustrated as bands in a lighter shade and are calculated on a 5% significance level by bootstrapping with 500 replications.

B.2 Parametric Calibration of the Income Distribution

We use a lognormal approximation of the income distribution with a Pareto tail for incomes above 150 000 €. To obtain an estimate for the mean and standard deviation, we use the same data on incomes as in the baseline simulation and perform a maximum likelihood estimation. As in the baseline calibration, we assume that a fixed mass of the population earns an income of zero and the Pareto tail is the same in both education groups. Table B1 contains the results of the Maximum Likelihood estimation of the mean and standard deviation.[26]

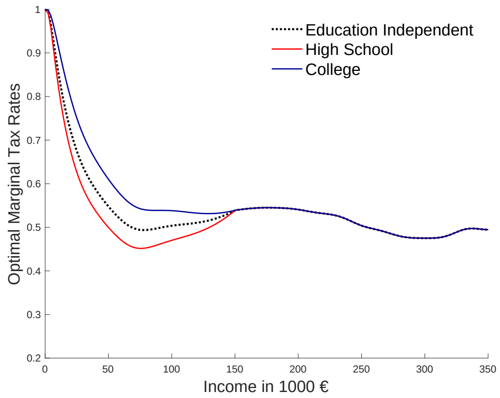
[26] Note that the simulation of the education-dependent optimal marginal tax rates with a parametric calibration does not directly allow for bootstrapped confidence intervals.

Table B1: Maximum Likelihood Estimation Results

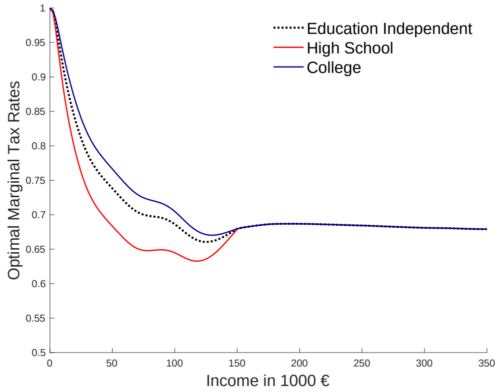
	Germany	UK	Spain	France	Italy
Mean					
Unconditional Distribution	10.15	10.06	9.51	9.96	9.91
College Distribution	10.50	10.27	9.86	10.31	10.19
High School Distribution	10.00	9.84	9.27	9.73	9.83
Standard Deviation					
Unconditional Distribution	0.87	0.91	1.15	1.02	0.87
College Distribution	0.88	0.93	1.07	0.88	0.88
High School Distribution	0.83	0.84	1.13	1.04	0.85

Notes: The mean and standard deviation of the income distributions are estimated under the assumption of a log-normal distribution by Maximum Likelihood. The sample of incomes are obtained from the GSOEP and EU-SILC.

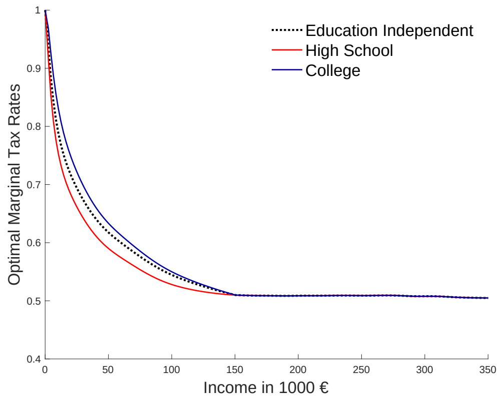
B.2.1 Parametric Calibration with Pareto Tail



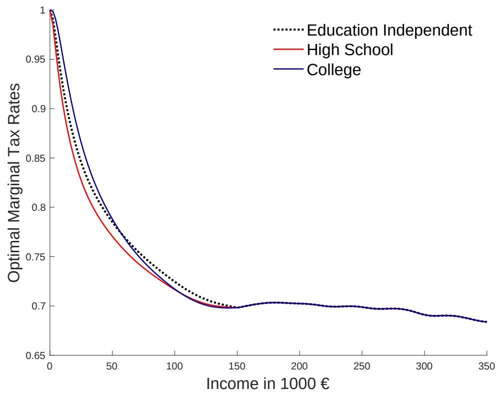
(a) Germany



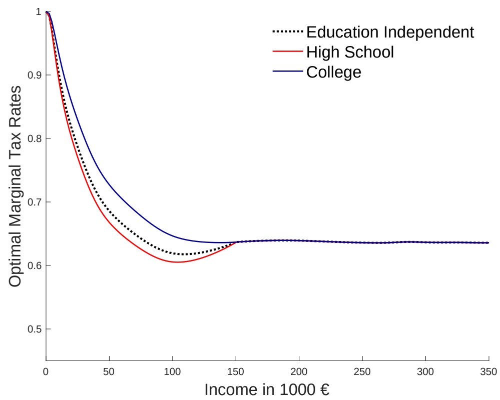
(b) United Kingdom



(c) Spain



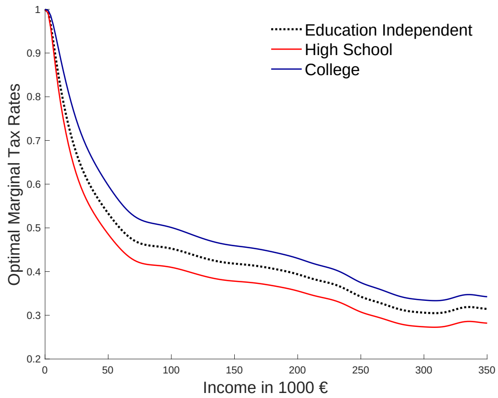
(d) France



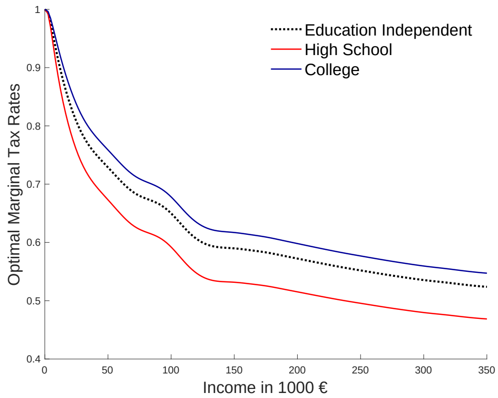
(e) Italy

Figure B2: Optimal Marginal Tax Rates Conditional on Education and Repayment Schedules with a Parametric Calibration of the Income Distribution

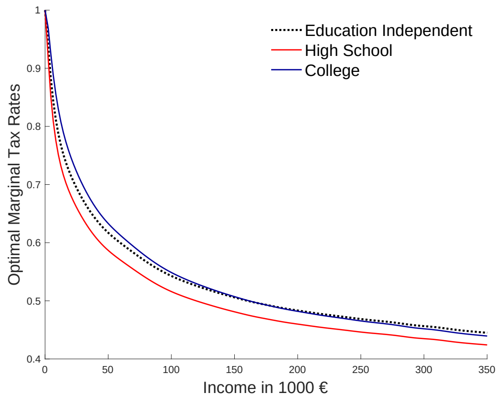
B.2.2 Parametric Calibration without Pareto Tail



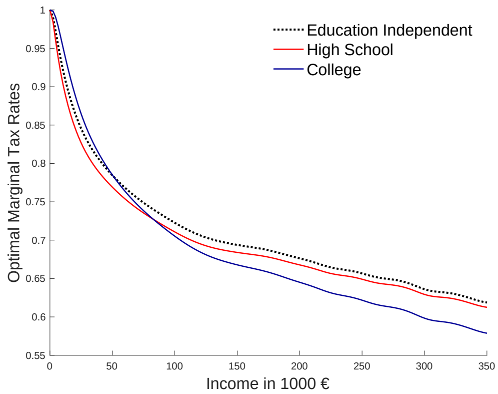
(a) Germany



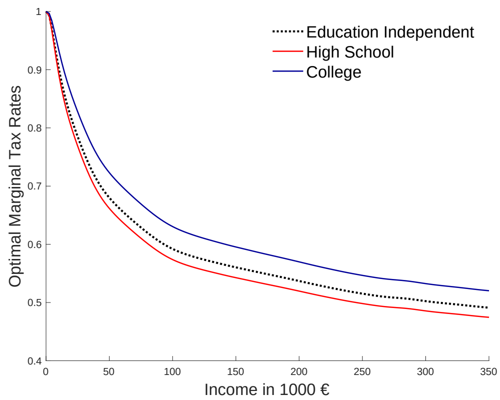
(b) United Kingdom



(c) Spain



(d) France



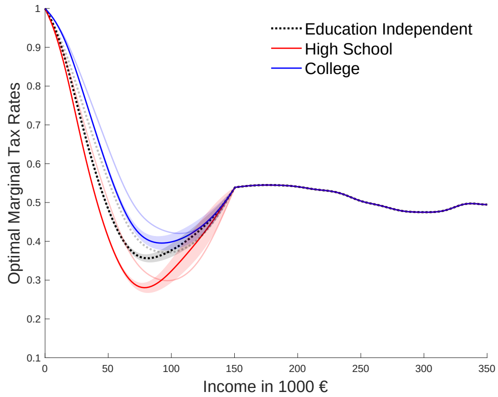
(e) Italy

Figure B3: Optimal Marginal Tax Rates Conditional on Education and Repayment Schedules with a Parametric Calibration of the Income Distribution

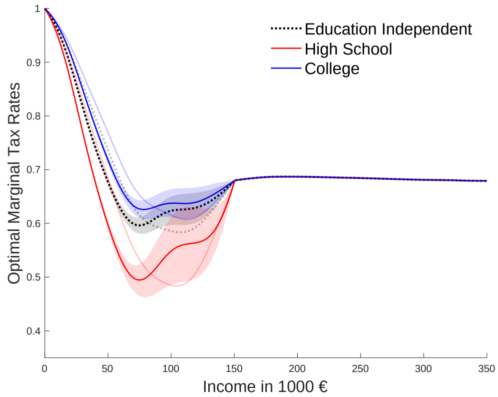
B.3 Choice of Bandwidth in the Kernel Density Estimation

Calibration of True Income Distribution and Random Sample of Incomes To calibrate a true income distribution, we assume that the income distribution is lognormal augmented with a Pareto tail. We estimate the mean and standard deviation from the GSOEP data by performing a maximum likelihood estimation. This yields a mean of 10.15 and a standard deviation of 0.87. Then, we append a Pareto tail for incomes above 150 000 € with a Pareto parameter of 1.67. Further, we assume that a mass of 4.4% always earns an income of zero. Next, we draw a random sample of incomes with a sample size of 7 500 from the calibrated true income distribution and calibrate an income distribution from that random sample. The calibration process is identical to the calibration in Section 3.1 for Germany. For the kernel density estimation at the beginning, we use five different bandwidths ranging from 5 000 € to 25 000 €.

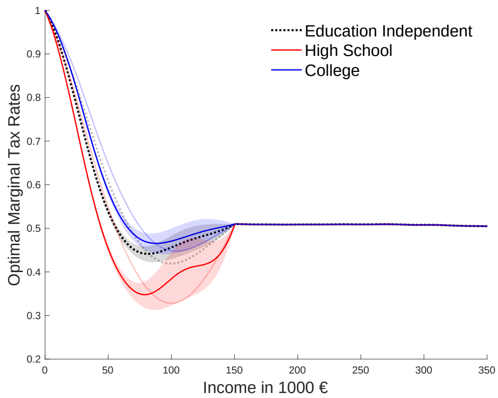
Optimal Marginal Tax Rates Figure B4 show optimal marginal tax rates conditional on education and the implied repayment schedule with a lower specified bandwidth than in the baseline simulation. Figure B4 uses a medium bandwidth of 20 000 €. For comparison, the optimal marginal tax rates of the baseline simulation with a large bandwidth of 30 000 €are illustrated in a lighter shade.



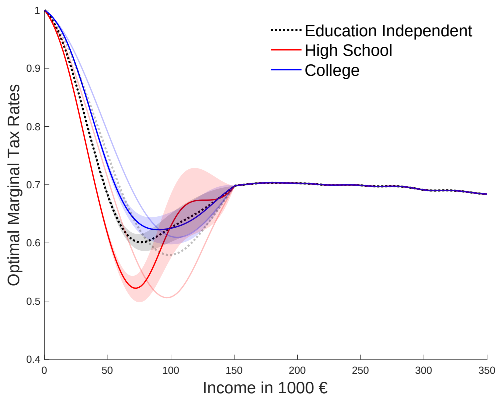
(a) Germany



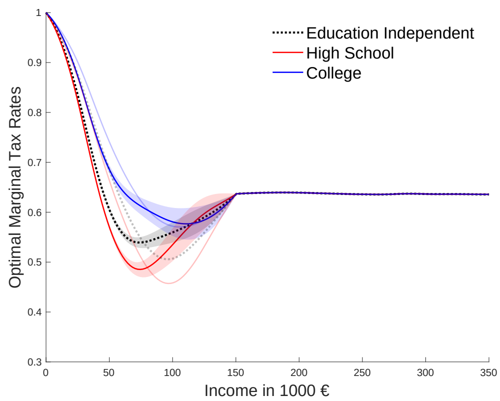
(b) United Kingdom



(c) Spain



(d) France



(e) Italy

Figure B4: Optimal Marginal Tax Rates Conditional on Education and Repayment Schedules with a Medium Bandwidth

Notes: Confidence intervals are illustrated as bands in a lighter shade and are calculated on a 5% significance level by bootstrapping with 500 replications. The simulated optimal marginal tax rates of the baseline simulation are displayed in a lighter shade. The corresponding 95% confidence interval of the simulated repayment schedule is not displayed for the sake of greater graphical clarity. The simulated repayment schedules of the baseline simulation are depicted with a dashed line.

B.4 Alternative Specifications of the Pareto Tail

Table B2: Properties of the Pareto Tail

	Germany	UK	Spain	France	Italy
Pareto Threshold					
Baseline	150 000 €	150 000 €	150 000 €	150 000 €	150 000 €
Bartels and Metzing (2019)	163 361 €	120 599 €	79 038 €	148 475 €	91 748 €
Pareto Parameter					
Baseline	1.67	1.78	2.11	2.2	2.22
Bartels and Metzing (2019)	1.64	1.79	1.96	1.92	2.18

Notes: The values of the Pareto parameter of the baseline simulation are from Atkinson, Piketty, and Saez (2011).

B.5 HSV Tax Function

Country-specific values for the degree of tax progressivity τ and the average level of taxation λ are taken from Kindermann, Mayr, and Sachs (2020) for Germany, García-Miralles, Guner, and Ramos (2019) for Spain and Holter, Krueger, and Stepanchuk (2019) for the UK, France and Italy and are summarized in Table B3. Note that the estimates from Holter, Krueger, and Stepanchuk (2019) refer to single households without children. Figure B5 illustrates the HSV functional form for all countries.

Table B3: Country Specific Parameters of HSV Specification

	Germany	UK	Spain	France	Italy
τ	0.128	0.168	0.148	0.138	0.153
λ	0.679	0.836	0.898	0.850	0.822

Notes: The values for τ and λ are taken from Kindermann, Mayr, and Sachs (2020) for Germany, García-Miralles, Guner, and Ramos (2019) for Spain and Holter, Krueger, and Stepanchuk (2019) for UK, France and Italy.

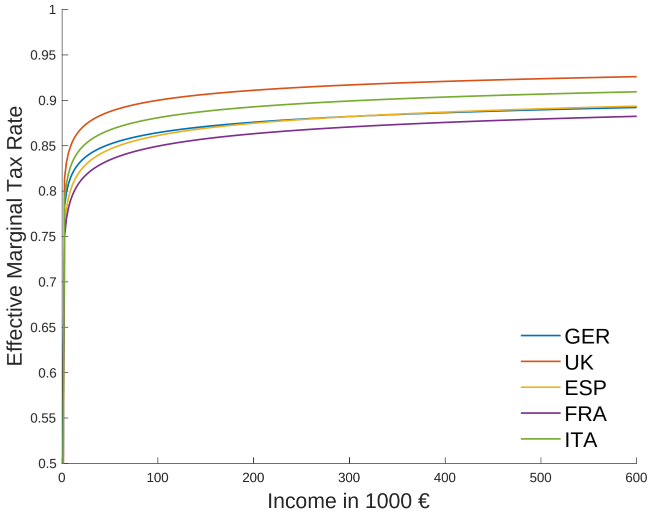
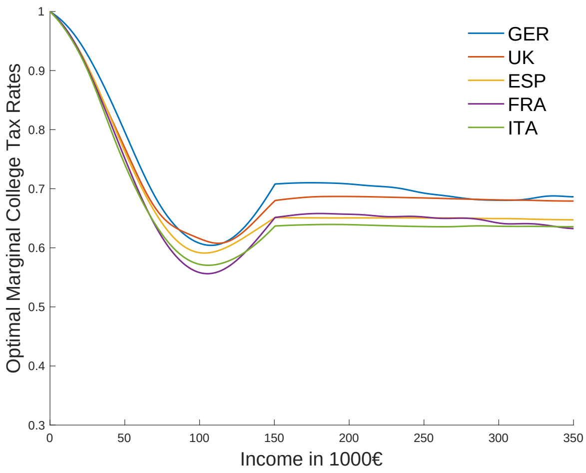
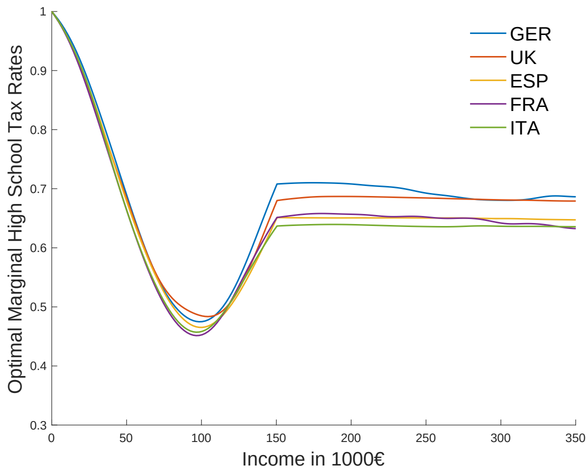


Figure B5: HSV Approximation of Marginal Tax Rates

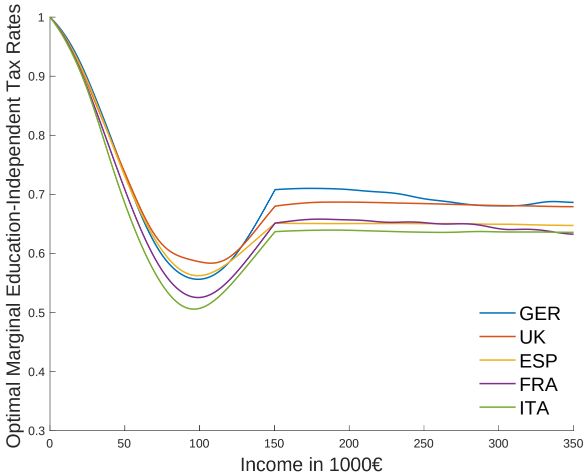
B.6 Identical Elasticities across Countries



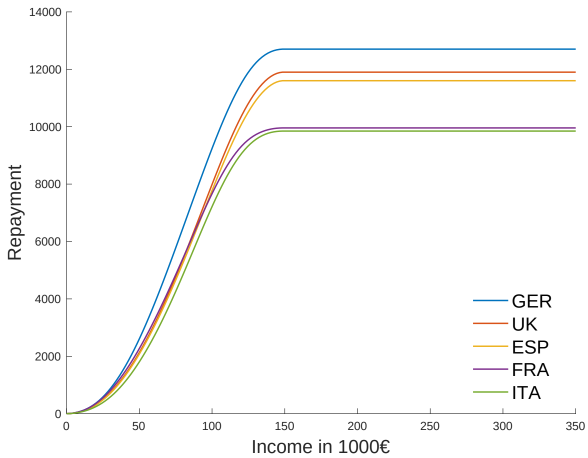
(a) College Tax Rates



(b) High School Tax Rates



(c) Education-Independent Tax Rates



(d) Simulated Repayment Schedule

Figure B6: Optimal Marginal Tax Rates Conditional on Education and Repayment Schedules with Identical Elasticities of Taxable Income across Countries

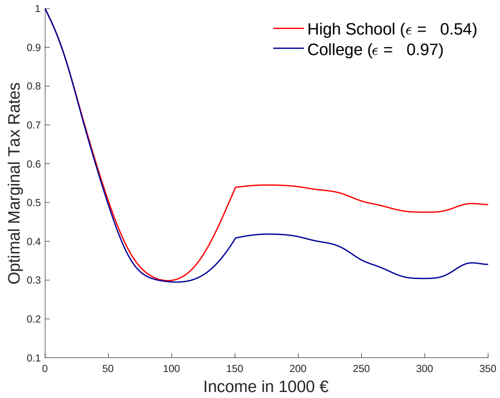
Notes: The corresponding 95% confidence intervals of the simulated optimal marginal tax rates and repayment schedule are not displayed for the sake of greater graphical clarity.

B.6.1 Education-Dependent Elasticities of Taxable Income

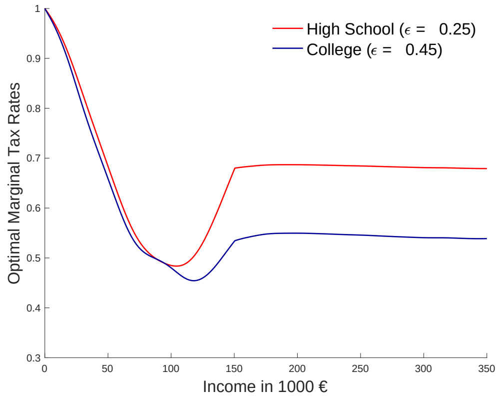
Table B4: Maximum Values of College Elasticity of Taxable Income

	Germany	UK	Spain	France	Italy
High School Elasticity	0.54	0.25	0.45	0.20	0.25
College Elasticity	0.97	0.45	0.76	0.32	0.40
Factor	1.80	1.80	1.69	1.60	1.60

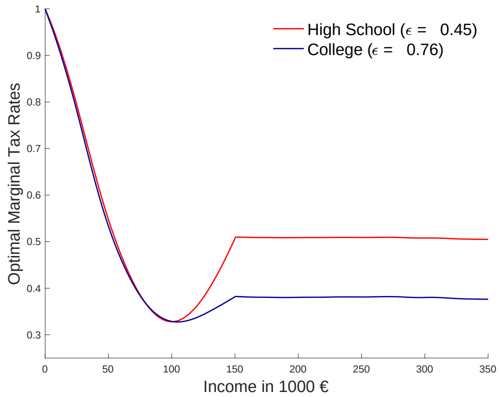
Notes: The values of the elasticity of taxable income of high school graduates are taken from the baseline calibration. In particular, I use the estimate of Doerrenberg, Peichl, and Siegloch (2017) for Germany, Brewer, Saez, and Shepard (2010) for the UK, Almunia and Lopez-Rodriguez (2019) for Spain, Lehmann, Marical, and Rioux (2013) for France and the approximate midpoint of estimated elasticities as reported by Saez, Slemrod, and Giertz (2012) for Italy. See Section 4.4 for details.



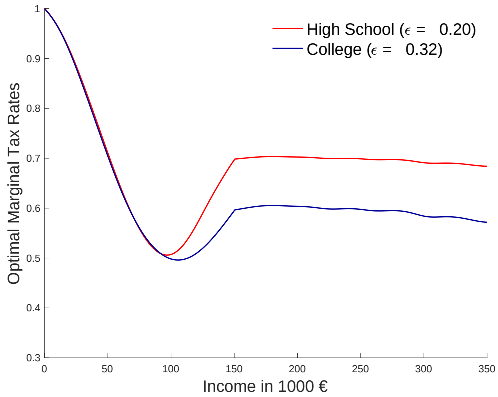
(a) Germany



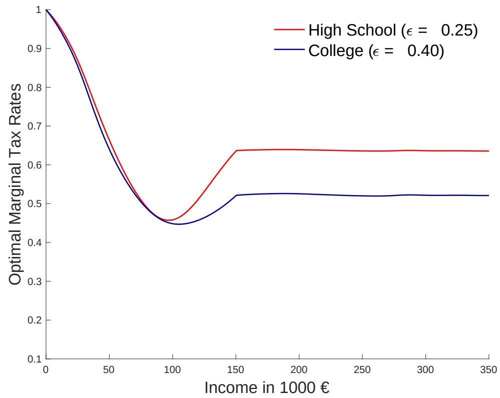
(b) United Kingdom



(c) Spain



(d) France



(e) Italy

Figure B7: Overlapping Optimal Marginal Tax Rates Conditional on Education with Education-Dependent Elasticities

Notes: The corresponding 95% confidence intervals of the simulated optimal education-dependent marginal tax rates are not displayed for the sake of greater graphical clarity.