NBER WORKING PAPER SERIES

OPTIMAL TAXATION OF TOP LABOR INCOMES: A TALE OF THREE ELASTICITIES

Thomas Piketty
Emmanuel Saez
Stefanie Stantcheva

Working Paper 17616
http://www.nber.org/papers/w17616

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2011

We thank co-editor Karl Scholz, Marco Bassetto, Wojciech Kopczuk, Laszlo Sandor, Florian Scheuer, Joel Slemrod, two anonymous referees, and numerous seminar participants for useful discussions and comments. Rolf Aaberge, Markus Jantti, Brian Nolan, Esben Schultz, and Floris Zoutman helped us gather international top marginal tax rate data. We are very thankful to Miguel Ferreira for sharing the international CEO data from Fernandes, Ferreira, Matos, and Murphy (2013) with us. We acknowledge ″nancial support from the Center for Equitable Growth at UC Berkeley and the MacArthur foundation. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities
Thomas Piketty, Emmanuel Saez, and Stefanie Stantcheva
NBER Working Paper No. 17616
November 2011, Revised March 2013
JEL No. H21

ABSTRACT

This paper presents a model of optimal labor income taxation where top incomes respond to marginal tax rates through three channels: (1) standard labor supply, (2) tax avoidance, (3) compensation bargaining. We derive the optimal top tax rate formula as a function of the three corresponding behavioral elasticities. The first elasticity (labor supply) is the sole real factor limiting optimal top tax rates. The optimal tax system should be designed to minimize the second elasticity (avoidance) through tax enforcement and tax neutrality across income forms. The optimal top tax rate increases with the third elasticity (bargaining) as bargaining efforts are zero-sum in aggregate. We provide evidence using cross-country times series macro-evidence and CEO pay micro-evidence. The macro-evidence from 18 OECD countries shows that there is a strong negative correlation between top tax rates and top 1% income shares since 1960, implying that the overall elasticity is large. However, top income share increases have not translated into higher economic growth. US CEO pay evidence shows that pay for luck is quantitatively more important when top tax rates are low. International CEO pay evidence shows that CEO pay is strongly negatively correlated with top tax rates even controlling for ″rm characteristics and performance, and this correlation is stronger in firms with poor governance. These results are consistent with bargaining effects playing a role in the link between top incomes and top tax rates. If bargaining effects in fact exist, optimal tax rates should be higher than commonly assumed.

Thomas Piketty
Paris School of Economics
48 Boulevard Jourdan
75014 Paris, France
piketty@ens.fr

Emmanuel Saez
Department of Economics
University of California, Berkeley
530 Evans Hall #3880
Berkeley, CA 94720
and NBER
saez@econ.berkeley.edu

Stefanie Stantcheva
MIT Department of Economics
50 Memorial Drive
Building E52
Cambridge, MA 02142-1347
stefanie@mit.edu

The share of total pre-tax income accruing to upper income groups has increased sharply in the United States. The top percentile income share has more than doubled from less than 10% in the 1970s to over 20% in recent years (Piketty and Saez, 2003). This trend toward income concentration has taken place in a number of other countries, especially English speaking countries, but is much more modest in continental Europe or Japan (Atkinson, Piketty, Saez, 2011 and Alvaredo et al. 2011). At the same time, top tax rates on upper income earners have declined sharply in many OECD countries, again particularly in English speaking countries.

While there have been many discussions both in the academic literature and the public debate about the causes of the surge in top incomes, there is not a fully compelling explanation. Most explanations can be classified into market driven changes vs. institution driven changes. The market driven stories posit that technological progress and globalization have been skilled-biased and have favored top earners relative to average earners (see e.g., Gabaix and Landier (2008) for CEOs and Rosen (1981) for Winner-Take-All theories for superstars). Those pure market explanations cannot account for the fact that top income shares have only increased modestly in a number of advanced countries (including Japan, Germany, or France) which are also subject to the same technological forces. The institution driven stories posit that changes in institutions, defined to include labor and financial market regulations, Union policies, tax policy, and more broadly social norms regarding pay disparity, have played a key role in the evolution of inequality. The main difficulty is that "institutions" are multi-dimensional and it is difficult to estimate compellingly the contribution of each specific factor.

Related, there is a wide empirical literature in public economics analyzing the effects of tax rates on pre-tax incomes (see Saez, Slemrod, and Giertz, 2012 for a recent survey) that reaches two broad conclusions. First, there is compelling evidence that upper incomes respond to tax rates whenever the tax code offers opportunities for tax avoidance. Such responses can sometime be quite large, especially in the short-run. Second however, when the tax base is broad and does not offer avoidance opportunities, the estimated elasticities are never large at least in the short or medium-run. To our knowledge, no study to date has been able to show convincing evidence in the short or medium-run of large real economic activity responses of upper earners to tax rates. However, it is difficult to provide compelling estimates of long-run elasticities. As we shall see, international evidence shows a strong correlation between top tax rate cuts and increases in top income shares in OECD countries since 1960.

There are three narratives of the link between top tax rates and upper incomes. First, after noting that top US incomes surged following the large top marginal tax rate cuts of the 1980s, Lindsey (1987) and Feldstein (1995) proposed a standard supply-side story whereby lower tax rates stimulate economic activity among top income earners (work, entrepreneurship, savings, etc.). Second, it has been pointed out–originally by Slemrod (1996)–that many of those dramatic responses were actually primarily due to tax avoidance rather than real economic behavior. Although this argument started as a critique of the supply-side success story, it has more recently been used to deny that any real increase in income concentration actually took place. Under this scenario, the real US top income shares were as high in the 1970s as they are today but a smaller fraction of top incomes was reported on tax returns in the 1970s than today. A third narrative contends that high top tax rates were part of the institutional set-up putting a brake on rent extraction among top earners. When top marginal tax rates are very high, the net reward to a highly paid executive for bargaining for more compensation is modest. When top tax rates fell, high earners started bargaining more aggressively to increase their compensation.

The first goal of this paper is to present a very simple model of optimal top labor income taxation that can capture all three avenues of response, the standard supply side response, the tax avoidance response, and the compensation bargaining response to assess how each narrative translates into tax policy implications. We therefore derive the optimal top tax rate formula as a function of the three elasticities corresponding to those three channels of responses. The first elasticity e_1 (supply side) is the sole real factor limiting optimal top tax rates. A large tax avoidance elasticity e_2 is a symptom of a poorly design tax system. A very high top tax rate within such a system offering many tax avoidance opportunities is counter-productive. Hence, the optimal tax system should be designed to minimize tax avoidance opportunities through a combination of tax enforcement, base broadening, and tax neutrality across income forms. In that case, the second elasticity (avoidance) becomes irrelevant. The optimal top tax rate *increases* with the third elasticity e_3 (bargaining) as bargaining efforts are wasteful and zero-sum in aggregate. If a substantial fraction of the behavioral response of top earners comes from bargaining effects and top earners are not paid less than their economic product, then the optimal top tax rate is much higher than the conventional formula and actually goes to 100% if the real supply-side elasticity is very small.[1] If bargaining effects are moderately large,

[1] The optimal top tax rate is moderate if the supply elasticity is fairly large and top earners are underpaid relative to their product, a situation that is theoretically possible in our model and might exist in countries with

the quasi-confiscatory top marginal tax rates–80%-90% or more–applied in the United States and the United Kingdom between the 1940s and the 1970s, might have been consistent with a sensibly-specified optimal tax model.

The second goal of the paper is to provide empirical evidence on the decomposition of the total behavioral response of top incomes to top tax rates into those three channels. We consider both macro-level cross-country/times series evidence and CEO pay micro-level evidence.

The macro-evidence uses time series on top income shares from the World Top Incomes Database, top income tax rates, and real GDP per capita data. We obtain three main results. First, we find a very clear correlation between the drop in top marginal tax rates and the surge in top income shares since 1960. This suggests that the long-run total elasticity of top incomes with respect to the net-of-tax rate is large, around 0.5. Second, examination of the US case suggests that the tax avoidance response cannot account for a significant fraction of the long-run surge in top incomes because top income shares based on a broader definition of income (that includes realized capital gains and hence a significant part of avoidance channels) has increased virtually as much as top income shares based on a narrower definition of income subject to the progressive tax schedule. Third, we find no evidence of a correlation between growth in real GDP per capita and the drop in the top marginal tax rate in the period 1960 to the present. This evidence is consistent with the bargaining model whereby gains at the top come at the expense of lower income earners. This suggests that the first elasticity is modest in size and that the overall effect comes mostly from the third elasticity.

The micro-evidence uses data on CEO pay in the United States since 1970 and international CEO pay data for 2006. We obtain two main results. First, the US evidence shows that pay for firm's performance outside of the control of the CEO (due to industry–wide performance as in Bertrand and Mullainathan, 2001) is quantitatively more important when top tax rates are low. This suggests that low top tax rates have induced CEOs to increase the component of their pay not directly related to their own performance. The main channel may have been the development of stock-options in the 1980s and 1990s which do not filter out performance unrelated to CEOs' actions (Hall and Murphy, 2003). Second, international CEO pay evidence for 2006 shows that CEO pay is strongly negatively correlated with top tax rates even controlling for firm's characteristics and performance, and that this correlation is stronger in firms with

very low income concentration.

poor governance. This suggests that the link between top tax rates and CEO pay does not run through firm performance but is likely due to bargaining effects as the bargaining position of the CEO is stronger when top rates are low and in firms with poorer governance.

All those results suggests that bargaining effects play a role in the link between top incomes and top tax rates implying that optimal top tax rates could be higher than commonly assumed. Bringing together the model and the empirical evidence, in our preferred estimates, we find an overall elasticity $e = 0.5$, which can be decomposed into $e_1 = 0.2$ (at most), $e_2 = 0$ and $e_3 = 0.3$ (at least). This corresponds to a socially optimal top tax rate $\tau^* = 83\%$ - as compared to $\tau^* = 57\%$ in the standard supply-side case with $e = e_1 = 0.5$ and $e_2 = e_3 = 0$. This illustrates the critical importance of this decomposition into three elasticities.

Our paper is related to a large body of theoretical work in optimal income taxation and empirical work on estimating behavioral responses to taxation. Previous work has focused mostly on the traditional supply-side channel and the tax avoidance/evasion channels.[2]

There is much less work in optimal taxation using models where pay differs from marginal product. A few studies have analyzed optimal taxation in models with labor market imperfections such as search models, Union models, efficiency wages models (Sorensen, 1999 provides a survey). The main focus of those papers has been on efficiency issues rather than redistributive issues, with most of the focus on the employment vs. unemployment margin. Fewer papers have addressed redistributive optimal tax policy in models with imperfect labor markets.[3] Motivated by recent events, a few papers have proposed models of optimal taxation with rent-seeking. Lockwood, Nathanson, and Weyl (2012) consider a model where each profession creates externalities that can only be targeted indirectly through a nonlinear income tax. If high earning professions generate larger negative externalities then progressive taxation is desirable on pure efficiency grounds (i.e., solely for correcting externalities). Rothschild and Scheuer (2012) consider a model with a rent-seeking sector and a traditional sector and solve for the (sector blind) optimal nonlinear income tax. They obtain optimal tax formulas than include the standard Mirrleesian term as well as an additional externality correcting term. The externality correcting term is naturally positive but it can be smaller or larger than the pure Pigouvian correction term

[2] Piketty and Saez (2013) and Saez, Slemrod, and Giertz (2012) provide recent surveys of the optimal tax and empirical literatures. Slemrod and Yitzhaki (2002) review specifically the tax avoidance/evasion literature.

[3] Hungerbuhler et al. (2006) analyze a search model with heterogeneous productivity, and Stantcheva (2011) considers optimal redistribution in a labor market screening setting where firms cannot observe perfectly the productivity of their employees.

depending on whether the within-sector or the across-sector externality dominates. In our simpler model, the correcting term is always equal to the Pigouvian term. As we shall discuss, our optimal top rate formula also can be connected to their more general analysis. Finally, Besley and Gathak (2013) show that the possibility of bailouts to financial intermediaries distorts the supply price of capital and creates an argument for taxing financial bonuses separately from other sources of income, in addition to the standard redistributive argument. Our theoretical value added is to bring together in a single framework the three channels of behavioral responses and show how optimal tax top tax rate formulas can be expressed in terms of the estimable elasticities corresponding to each response channel. Our empirical value added is to attempt to gauge the importance of these three channels, most notably the rent-seeking channel, and to calibrate our theoretical formulas accordingly.

The remainder of the paper is organized as follows. Section I presents our theoretical model. Section II presents macro-level empirical results. Section III presents micro-level evidence using CEO pay. Section IV synthesizes the results, and provides a brief conclusion. Extensions and data construction details are gathered in the web appendix. All data are available online.

I Theoretical Model

A Standard Model: Supply-Side and Tax Avoidance Responses

In the paper, we denote by z taxable earnings and by $T(z)$ the nonlinear tax schedule. We assume a constant marginal tax rate τ in the top bracket above a given income threshold $\bar{z}$. We assume without loss of generality that the number of taxpayers in the top bracket has measure one at the optimum. We refer to this group as top bracket taxpayers. We focus on the determination of the optimal top tax rate τ, taking $\bar{z}$ as given.

The government maximizes a standard social welfare function of the form:

$$W = \int G(u_i) d\nu(i), \quad \text{subject to} \quad \int T(z_i) d\nu(i) \geq T_0$$

where $G(.)$ is increasing concave, u_i is the utility of individual i, and $d\nu(i)$ is the density mass of people of individuals of type i, and $T_0 \geq 0$ is an exogenous tax revenue requirement.

Denoting by p the multiplier of the government budget constraint, we define the social marginal welfare weight on individual i as $g_i = G'(u_i)u_{ci}/p$. We assume that the average social marginal welfare weight among top bracket income earners is zero.[4] In that case, the government sets τ to maximize tax revenue raised from top bracket taxpayers. Considering a zero marginal welfare weight allows us to obtain an upper bound on the optimal top tax rate.[5]

Supply-side responses. We start with the standard model with only supply-side responses as in Saez (2001). See Piketty and Saez, 2013 for a detailed presentation and survey of this classic case. We assume away income effects for simplicity and tractability, and consider utility functions of the form $u_i(c,z) = c - h_i(z)$ where z is pre-tax earnings, $c = z - T(z)$ is disposable income, and $h_i(z)$ denotes the labor supply cost of earning z which is increasing and convex in z. Optimal effort choice is given by the first order condition $h_i'(z) = 1 - \tau$ where τ is the marginal tax rate so that individual earnings $z_i(1-\tau)$ are solely a function of the net-of-tax rate $1 - \tau$. Aggregating over all top bracket taxpayers, we denote by $z(1-\tau)$ the *average* income reported by top bracket taxpayers, as a function of the net-of-tax rate. The aggregate elasticity of income in the top bracket with respect to the net-of-tax rate is therefore defined as

$$e_1 = \frac{1-\tau}{z} \frac{dz}{d(1-\tau)}. \tag{1}$$

This is the standard *first elasticity* that reflects real economic responses to the net-of-tax rate, which can be labeled as labor supply effects, broadly defined (more hours of work, more intense effort per hour worked, occupational choice, etc.)

The optimal top tax rate maximizing tax revenue is given by:

$$\tau^* = \frac{1}{1 + a \cdot e_1}, \tag{2}$$

where $a = z/(z - \bar{z}) > 1$ is the Pareto parameter of the top tail of the distribution.[6]

The proof of formula (2) is straightforward and well known. The government chooses τ to maximize top bracket tax revenue $T = \tau[z(1-\tau) - \bar{z}]$. The first order condition is $[z - \bar{z}] - \tau \frac{dz}{d(1-\tau)} = 0$ which can be immediately re-arranged as (2) using the definition of e_1 from (1).

Adding tax avoidance responses. As shown by many empirical studies (see Saez, Slemrod, and Giertz, 2012 for a recent survey), responses to tax rates can also take the form of tax

[4] If the social welfare function $G(.)$ has curvature so that $G'(u) \to 0$ when $u \to \infty$, this will be the case when $\bar{z} \to \infty$ and will hence approximately be true for large $\bar{z}$.

[5] As we shall discuss, formulas can be easily adapted if we instead put a positive social welfare weight g on the marginal consumption of top bracket earners (relative to average).

[6] If a positive social weight $g > 0$ is set on top earners marginal consumption, then the optimal top tax rate is $\tau = (1-g)/(1-g+ae_1)$.

avoidance. We can define tax avoidance as changes in reported income due to changes in the form of compensation but not in the total level of compensation. Tax avoidance opportunities arise when taxpayers can shift part of their taxable income into another form or another time period that is treated more favorably from a tax perspective.[7]

The main distinction between real and tax avoidance responses is that real responses reflect underlying, deep individual preferences for work and consumption while tax avoidance responses depend critically on the design of the tax system and the avoidance opportunities it offers. While the government cannot drastically change underlying deep individual preferences and hence the size of the real elasticity, it can change the tax system to reduce avoidance opportunities. Naturally, this distinction is one of degree as some forms of tax avoidance cannot be easily eliminated due to technological constraints (see our discussion below) and, symmetrically, some real responses could be somewhat dampened by government policies.

We can extend the standard model as follows to incorporate tax avoidance.[8] Let us denote by y real income and by x sheltered income so that ordinary taxable income is $z = y - x$. The latter is taxed at marginal tax rate τ in the top bracket, while sheltered income x is taxed at a constant and uniform marginal tax rate t lower than τ.[9] The utility function of individual i takes the form $u_i(c, y, x) = c - h_i(y) - d_i(x)$, where $c = y - \tau z - tx + R = (1-\tau)y + (\tau - t)x + R$ is disposable after tax income and $R = \tau\bar{z} - T(\bar{z})$ denotes the virtual income coming out of the nonlinear tax schedule. $h_i(y)$ is the utility cost of earning real income y, and $d_i(x)$ is the cost of sheltering an amount of income x. There is a cost to sheltering, since sheltered income such as fringe benefits or deferred earnings are less valuable than cash income. We assume that both $h_i(.)$ and $d_i(.)$ are increasing and convex, and normalized so that $h_i'(0) = d_i'(0) = 0$. This model nests the standard model when the sheltering cost $d_i(x)$ is infinitely large for any $x > 0$.

Individual utility maximization implies that $h_i'(y) = 1 - \tau$ and $d_i'(x) = \tau - t$, so that y_i is an increasing function of $1 - \tau$ and x_i is an increasing function of the tax differential

[7] Examples of such avoidance/evasion are (a) reductions in current cash compensation for increased fringe benefits or deferred compensation such as stock-options or future pensions, (b) increased consumption within the firm such as better offices, vacation disguised as business travel, private use of corporate jets, etc. (c) changes in the form of business organization such as shifting profits from the individual income tax base to the corporate tax base, (d) re-characterization of ordinary income into tax favored capital gains, (e) outright tax evasion such as using off-shore accounts.

[8] This follows and extends Saez (2004) and Saez, Giertz, and Slemrod (2012). A broad literature surveyed by Slemrod and Yitzhaki (2002) and Piketty and Saez (2013) has introduced tax avoidance in optimal tax models.

[9] For example, in the case of non-taxable fringe benefits, $t = 0$. In the case of shifting ordinary income into tax favored capital gains, we have $t > 0$ but with t significantly less than τ.

$\tau - t$. Aggregating over all top bracket taxpayers, we have $y = y(1-\tau)$ with real elasticity $e_1 = \frac{1-\tau}{y}\frac{dy}{d(1-\tau)} > 0$ as in (1) and $x = x(\tau - t)$ increasing in $\tau - t$. Note that $x(0) = 0$ as there is sheltering only when $\tau > t$.

Hence $z = z(1-\tau, t) = y(1-\tau) - x(\tau - t)$ is increasing in $1-\tau$ and t. We denote by $e = \frac{1-\tau}{z}\frac{dz}{d(1-\tau)} > 0$ the total elasticity of taxable income z with respect to $1-\tau$ when keeping t constant. We denote by s the fraction of the behavioral response of z to $d\tau$ due to tax avoidance, and by $e_2 = s \cdot e$ the *tax avoidance elasticity component*:

$$s = \frac{dx/d(\tau - t)}{dy/d(1-\tau) + dx/d(\tau - t)} = \frac{dx/d(\tau - t)}{\partial z/\partial(1-\tau)} \quad \text{and} \quad e_2 = s \cdot e = \frac{1-\tau}{z}\frac{dx}{d(\tau - t)}. \tag{3}$$

By construction, we have $(1-s)e = (y/z)e_1$, or equivalently $e = (y/z)e_1 + e_2$. If we start from a situation with no tax avoidance ($y = z$), then we simply have $e = e_1 + e_2$, i.e. the total elasticity is the sum of the standard labor supply elasticity and the tax avoidance elasticity component. We can prove the following two results.[10]

Partial optimum: For a given t, the optimal top tax rate τ on taxable income is

$$\tau^* = \frac{1 + t \cdot a \cdot e_2}{1 + a \cdot e}, \tag{4}$$

where $e = (y/z)e_1 + e_2$ is the elasticity of taxable income (keeping t constant), $e_1 = \frac{1-\tau}{y}\frac{dy}{d(1-\tau)}$ is the real labor supply elasticity, and $e_2 = \frac{1-\tau}{z}\frac{dx}{d(\tau-t)}$ is the tax avoidance elasticity component.

Full optimum: If sheltering occurs only within top bracket earners and t can be changed at no cost to the government, the optimal global tax policy is to set t and τ equal to

$$t^* = \tau^* = \frac{1}{1 + a \cdot e_1}. \tag{5}$$

Proof: As there is a measure one of top bracket earners, the government chooses τ to maximize $T = \tau[z(1-\tau, t) - \bar{z}] + tx(\tau - t)$. The first order condition for τ is

$$0 = [z - \bar{z}] - \tau\frac{\partial z}{\partial(1-\tau)} + t\frac{dx}{d(\tau - t)} = [z - \bar{z}] - \tau\frac{\partial z}{\partial(1-\tau)} + ts\frac{\partial z}{\partial(1-\tau)},$$

where the second expression is obtained using the definition of s from (3). The first two terms are the same as in the standard model. The third term captures the "fiscal externality" as a fraction s of the behavioral response translates into sheltered income taxed at rate t. Using the

[10]Our results easily extend to the more general case with utility $c - d_i(x, y)$, which generates aggregate supply functions of the form $z(\tau, t), y(\tau, t), x(\tau, t)$. We used the separable case for simplicity of presentation.

definition of $e = \frac{1-\tau}{z}\frac{dz}{d(1-\tau)}$, we can rewrite the first order condition as $\frac{\tau - ts}{1-\tau} e = \frac{z-\bar{z}}{z} = \frac{1}{a}$, which can be rearranged into formula (4) using the fact that $e_2 = s \cdot e$ from (3).

The second part of the proof can be obtained by taking the first order condition with respect to t. As $z(1-\tau, t) = y(1-\tau) - x(\tau - t)$, the first order condition is $\frac{dT}{dt} = x + [\tau - t]\frac{dx}{d(\tau - t)} = 0$.[11] As $x \geq 0$ and $\tau \geq t$ and $dx/d(\tau - t) \geq 0$, this first order condition can only hold for $t = \tau$ and $x(\tau - t = 0) = 0$. Setting $t = \tau$ in equation (4), and noting that $x = 0$ implies that $z = y$ and hence $e - e_2 = e_1$, we immediately obtain (5). Intuitively, as x is completely wasteful, it is optimal to deter x entirely by setting $t = \tau$. **QED.**

Three comments are worth noting about these results.

First, if $t = 0$ then $\tau = 1/(1 + a \cdot e)$ as in the standard model. In the narrow framework where the tax system is taken as given (i.e. there is nothing the government can do about tax evasion and income shifting), and where sheltered income is totally untaxed, then whether e is due to real responses vs. avoidance responses is irrelevant, a point made by Feldstein (1999). Second however, if $t > 0$, then sheltering creates a "fiscal externality," as the shifted income is taxed at rate t and $\tau > 1/(1 + a \cdot e)$. Third and most important, the government can improve efficiency and its ability to tax top incomes by closing tax avoidance opportunities (setting $t = \tau$ in our model). Sheltering then becomes irrelevant and the real elasticity e_1 is the only factor limiting taxes on upper incomes. Kopczuk (2005) shows that the Tax Reform Act of 1986 in the United States, which broadened the tax base and closed loopholes did reduce the elasticity of reported income with respect to the net-of-tax rate. Kleven and Schultz (2012) finds small yet very compellingly identified elasticities for large top tax rate changes in Denmark, a very high tax country where tax avoidance opportunities are indeed very limited.

Actual tax avoidance opportunities come in two varieties. Some are pure creations of the tax system, such as exemption of fringe benefits or differential treatment of different income forms and hence could be eliminated by reforming the tax system. In that case, t is a free parameter that the government can change at no cost as in our model. Yet other tax avoidance opportunities reflect real enforcement constraints that are costly–sometimes even impossible–for the government to eliminate.[12] Slemrod and Kopczuk (2002) present a model with costs

[11] Note that we have used the assumption stated in the proposition that sheltering happens only within top bracket taxpayers so that a change in t has no effect on individuals below the top bracket.

[12] For example, it is very difficult for the government to tax profits from informal cash businesses. Fighting off-shore tax evasion requires international cooperation.

of enforcement. The government might also want to use differential taxes on different income sources for redistributive reasons or for efficiency reasons.[13] Our simple model also ignores that there might be political hurdles to setting $t = \tau$, for example if some types of tax sheltering are fiercely defended by special interests or lobbying groups. The important policy question is then what fraction of the tax avoidance elasticity can be eliminated by tax redesign and tax enforcement. In a developing country with most economic activity taking place in small informal businesses, the tax avoidance elasticity cannot be reduced to zero. But in a modern economy and with international cooperation, the tax avoidance elasticity could likely be substantially reduced as most economic transactions, especially at the top end, are recorded and hence verifiable (Kleven, Kreiner, and Saez, 2009). We come back to this issue below.

B Compensation Bargaining Responses

Motivation and previous work. Pay may not equal marginal economic product for top income earners. In particular, executives can be overpaid if they are entrenched and use their power to influence compensation committees (Bebchuk and Fried (2004) survey the wide corporate finance literature on this issue). In principle, executives could also be underpaid relative to their marginal product if there are social norms against high compensation levels. In that case, a company might find it more profitable to under-pay its executives to buy peace with its other employees, customers, or the public in general.[14] To the extent that top income earners generally have more opportunities to set their own pay than low and middle income earners, the first case seems more likely. But from a theoretical perspective both cases are interesting.

More generally, pay can differ from marginal product in any model in which compensation is decided by on-the-job bargaining between an employer and an employee, as in the classic search model of Diamond-Mortensen-Pissarides. In that framework, there is a rent to be shared on the job because of frictions in the matching process and inability to commit to a wage before the match has occurred. Indeed, in such models, the wage rate is not pinned down and can be set to any value within the outside options of the worker and his marginal product (Hall, 2005).[15] Typically, the wage is then determined by the relative bargaining powers of

[13]The Ramsey model recommends to tax relatively less the most elastic goods. In the presence of income shifting, the gap between tax rates should be reduced (see our earlier working paper version).

[14]Recent examples have arisen in the case of the 2008 and 2009 bailouts of financial firms in the United States–although the ultimate effects on executive compensation are unclear.

[15]In such simple models, pay is typically below marginal product if and only if the outside option of the

the employer and employee, for example through Nash bargaining with exogenous weights. In general, the wage rate is not efficient, unless the so-called Hosios condition is met.[16] In more general models, given the substantial costs involved in replacing workers who quit in most modern work environments, especially for management positions where specific human capital is important, as well as imperfect information between firm and employee, it seems reasonable to think that there would be a band of possible compensation levels. In such a context, bargaining efforts on the job can conceivably play a significant role in determining pay. Marginal tax rates affect the rewards to bargaining effort and can hence affect the level of such bargaining efforts.[17]

Yet another reason why pay may differ from marginal product is imperfect information. In the real world, it is often very difficult to estimate individual marginal products, especially for managers working in large corporations. For tasks that are performed similarly by many workers (e.g. one additional worker on a factory line), one can approximately compute the contribution to total output brought by an extra worker. But for tasks that are more or less unique, this is much more complex: one cannot run a company without a chief financial officer or a head of communication during a few years in order to see what the measurable impact on total output of the corporation is going to be. For such managerial tasks, it is very unlikely that market experimentation and competition can ever lead to full information about individual marginal products, especially in a rapidly changing corporate landscape. If marginal products are unknown, or are only known to belong to relatively large intervals, then institutions, market power and beliefs systems can naturally play a major role for pay determination (see Rotemberg 2002). This is particularly relevant for the recent rise of top incomes. Using matched individual tax return data with occupations and industries, Bakija, Cole and Heim (2010) have recently

employee is lower than his product on the job. In more complex settings, with the outside option and productivity on the job evolving over time, as well as switching costs for both employer and employee, pay can be also above marginal product.

[16]Those standard search models stand in contrast to newer "directed search" models where the wage is negotiated ex-ante and in which case efficiency is restored (see e.g., Moen, 1997).

[17]To take an example familiar to most readers, academic faculty pay is often determined by outside options taking the form of competitive offers from outside institutions. Because personal moving costs are difficult to observe by the upper administration of one's home University, a formal competitive offer letter is often sufficient to trigger a pay increase in one's current job. Obtaining an outside offer for the sole purposes of getting a pay raise is costly and time consuming (both for the academic and to potential recruiters). If the pay raise in the home institution does not translate into higher productivity, then this is a pure compensation bargaining response. Obviously, lower tax rates make the pay raise more valuable and might encourage such type of behavior. If it can be raised by competitive outside offers, faculty pay will typically have to be below marginal product (for the home University). Faculty pay can also be above marginal product (if productivity declines) as pay is downward rigid and tenured faculty cannot be laid off.

shown that executives, managers, supervisors, and financial professionals account for 70 percent of the increase in the share of national income going to the top 0.1 percent of the US income distribution between 1979 and 2005.[18]

Theoretical model. We consider the simplest model that can capture bargaining compensation effects. Individual i receives a fraction η of his/her real product y and can put productive effort both into increasing y and bargaining effort into increasing η. Both types of effort are costly and utility is given by:

$$u^i(c, \eta, y) = c - h_i(y) - k_i(\eta),$$

where c is disposable after-tax income, $h_i(y)$ is the cost of producing output y as in the standard model, and $k_i(\eta)$ is the cost of bargaining necessary in order to receive a share η of the product. Both h_i and k_i are increasing and convex.[19] We again rule out income effects for simplicity.[20]

Let $b = (\eta - 1)y$ be bargained earnings defined as the gap between received earnings ηy and actual product y. Note that the model allows both overpay (when $\eta > 1$ and hence $b > 0$) and underpay (when $\eta < 1$ and hence $b < 0$). Let us denote by $E(b)$ the average bargained earnings in the economy. In the aggregate, it must be the case that total product is equal to total compensation. Hence, if $E(b) > 0$, so that there is overpay on average, $E(b)$ must come at the expense of somebody. Symmetrically, if $E(b) < 0$, then the average underpay $-E(b)$ must benefit somebody. For simplicity, we assume that any gain made through bargaining comes uniformly at the expense of everybody else in the economy. Hence, individual incomes are all reduced by a uniform amount $E(b)$ (or increased by a uniform amount -$E(b)$ if $E(b) < 0$).[21] In reality, bargaining pay likely comes at the expense of other employees or shareholders in the same company or sector. In Appendix A.1, we discuss in detail how and in which class of models this uniformity assumption can be relaxed without affecting our results (we summarize those results below).

[18]Including about two thirds in the non-financial sector, and one third in the financial sector. In contrast, the combined share of the arts, sports and medias sub-sectors, usually used to illustrate winner-take-all theories, is only 3.1% of all top 0.1% taxpayers. See Bakija, Cole, and Heim (2012, Table 1).

[19]We could consider a general non separable cost of effort function $h_i(y, \eta)$ to allow for example for substitution between productive vs. bargaining effort. The optimal tax formula would be identical but the comparative statics would be less transparent and would require additional assumptions.

[20]This model nests the standard model if the cost function k is such that $k(1) = 0$ and there is infinite disutility cost of pushing η above 1.

[21]A simple but admittedly unrealistic scenario in which our uniformity assumption holds would be a situation where firms are owned equally in the population and bargaining for pay comes at the expense of profits.

Because the government uses a nonlinear income tax schedule, it can adjust the demogrant $-T(0)$ to fully offset $E(b)$. Effectively, the government can always tax (or subsidize) $E(b)$ at 100% before applying its nonlinear income tax. Hence, we can assume that the government absorbs one-for-one any change in $E(b)$. Therefore, we can simply define earnings as $z = \eta y = y + b$ and assume that those earnings are taxed nonlinearly. This simplification is possible because of our key assumption that $E(b)$ affects all individuals uniformly (or, alternatively, in the class of models presented in Appendix A.1).

Individual i chooses y and η to maximize $u^i(c, \eta, y) = \eta \cdot y - T(\eta \cdot y) - h_i(y) - k_i(\eta)$, so that

$$(1-\tau)\eta = h_i'(y) \quad \text{and} \quad (1-\tau)y = k_i'(\eta),$$

where $\tau = T'$ is the marginal tax rate. This naturally defines y_i and η_i as *increasing* functions of the net-of-tax rate $1-\tau$. Hence $z_i = \eta_i \cdot y_i$ and $b_i = (1-\eta_i) \cdot y_i$ are also functions of $1-\tau$.

Let us consider as in the previous section the optimal top tax rate τ for incomes above a threshold level $\bar{z}$ and assume again that there is a measure one of taxpayers with incomes above $\bar{z}$. Let us denote by $z(1-\tau)$, $y(1-\tau)$, and $b(1-\tau)$ average reported income, productive earnings, and bargained earnings across all taxpayers in the top bracket. We can then define, as above, the real labor supply elasticity e_1 and the total compensation elasticity e to be:

$$e_1 = \frac{1-\tau}{y}\frac{dy}{d(1-\tau)} \geq 0 \quad \text{and} \quad e = \frac{1-\tau}{z}\frac{dz}{d(1-\tau)}$$

We define s, the fraction of the marginal behavioral response due to bargaining and by $e_3 = s \cdot e$ the *bargaining elasticity component*:

$$s = \frac{db/d(1-\tau)}{dz/d(1-\tau)} = \frac{db/d(1-\tau)}{db/d(1-\tau) + dy/d(1-\tau)} \quad \text{and} \quad e_3 = s \cdot e = \frac{1-\tau}{z}\frac{db}{d(1-\tau)}. \tag{6}$$

This definition immediately implies that $(y/z)e_1 = (1-s) \cdot e$. By construction, $e = (y/z)e_1 + e_3$. If we start from a situation where top taxpayers are paid their marginal product ($y = z$), then we simply have $e = e_1 + e_3$. Importantly, s (and hence e_3) can be either positive or negative but it is always positive if individuals are overpaid (i.e., if $\eta > 1$). If individuals are underpaid (i.e., $\eta < 1$) then s (and hence e_3) can be negative. More precisely, we can easily prove:

$$s = 1 - \frac{e_1}{\eta\,(e_\eta + e_1)} = 1 - \frac{y \cdot e_1}{z \cdot e} \leq 1 \quad \text{with} \quad e_\eta = \frac{1-\tau}{\eta}\frac{d\eta}{d(1-\tau)} = e - e_1 \geq 0.$$

$$s \leq 0 \quad \text{if and only if} \quad \eta \leq \frac{e_1}{e_1 + e_\eta}. \quad \text{If} \quad \eta > 1 \quad \text{then} \quad s > 0.$$

We can now state our main proposition.

Proposition 1 *The optimal top tax rate is*

$$\tau^* = \frac{1 + a \cdot e_3}{1 + a \cdot e} = 1 - \frac{a(y/z)e_1}{1 + a \cdot e}, \tag{7}$$

where $e = (y/z)e_1 + e_3$ *is the elasticity of taxable income,* $e_1 = \frac{1-\tau}{y}\frac{dy}{d(1-\tau)} = \frac{z(1-s)e}{y}$ *the real labor supply elasticity, and* $e_3 = s \cdot e = \frac{1-\tau}{z}\frac{db}{d(1-\tau)}$ *the compensation bargaining elasticity.*

- τ^* *decreases with* e *(keeping* e_3 *constant) and increases with* e_3 *(keeping* e *constant).*
- τ^* *decreases with the real elasticity* e_1 *(keeping* e *and* y/z *constant) and increases with the level of overpayment* $\eta = z/y$ *(keeping* e_1 *and* e *constant)*
- *If* $e_1 = 0$ *then* $\tau^* = 1$.
- *If* $z \geq y$ *(top earners are overpaid) then* $e_3 \geq 0$ *and* $\tau^* \geq 1/(1 + a \cdot e_1)$

Proof: The government aims to maximize taxes collected from taxpayers in the top bracket. Taxes collected from the latter are $\tau[z - \bar{z}]$ but the tax τ also impacts $E(b)$ and hence the government's budget (as the government absorbs any change in $E(b)$ through the demogrant). Since the total size of the population is N, the government chooses τ to maximize $T = \tau[z(1-\tau) - \bar{z}] - N \cdot E(b)$. If $d\tau$ triggers a change in b in the top bracket, that change is then reflected one-for-one in $NE(b)$. Hence we have $NdE(b)/d(1-\tau) = db/d(1-\tau)$ and the first order condition for τ is:

$$[z - \bar{z}] - \tau\frac{dz}{d(1-\tau)} + \frac{db}{d(1-\tau)} = 0, \quad \Rightarrow \quad [\tau - s]\frac{dz}{d(1-\tau)} = z - \bar{z}, \quad \Rightarrow \quad \frac{\tau - s}{1 - \tau} \cdot e = \frac{z - \bar{z}}{z} = \frac{1}{a},$$

which leads to (7) using $e_3 = s \cdot e$. The rest of the proposition is straightforward. **QED.**

Proposition 1 shows that it is possible to obtain a simple optimal tax formula that nests the standard model in the case $e_3 = 0$ (no bargaining elasticity). Implementing the formula requires knowing the total elasticity e and the bargaining elasticity component e_3 (or equivalently the fraction s of the behavioral response at the margin due to bargaining effects). e_3 can also be indirectly be obtained by substraction from e using the real labor supply elasticity e_1 and the ratio of product to pay y/z. Hence, implementing the formula requires knowledge of not only the compensation response (i.e., the taxable income elasticity e), but also of the real economic product responses to tax changes, which is considerably more difficult.

Trickle-up. In the case where top earners are overpaid relative to their productivity ($z > y$), we have $s > 0$ and hence $e_3 > 0$ and the optimal top tax rate is higher than in the standard model (i.e., $\tau^* > 1/(1 + a \cdot e)$). This corresponds to a "trickle-up" situation where a tax cut on upper incomes shifts economic resources away from the bottom and toward the top. Those effects can be quantitatively large, as we will discuss in Section IV.

Trickle-down. In the case where top earners are underpaid relative to their productivity ($z < y$) and it is possible that $s < 0$ and hence $e_3 < 0$, in which case the optimal top tax rate is lower than in the standard model (i.e., $\tau < 1/(1 + a \cdot e)$). This corresponds to a "trickle-down" situation where a tax cut on upper incomes also shifts economic resources toward the bottom, as upper incomes work in part for the benefit of lower incomes.

Pigouvian interpretation. Economically, the extra-term in formula (7) relative to the standard formula $\tau = 1/(1 + a \cdot e)$ can therefore be interpreted as the Pigouvian correction term for the rent-seeking externality. A \$1 reduction in z due to a small increase in τ creates an \$$s = e_3/e$ positive externality. The optimal tax rate formula (7) takes the standard additive form of the conventional Mirrlees term plus the Pigouvian term.[22]

Regulation vs. taxation. We have taken as given the bargaining opportunities in the economy. Conceivably, the government can affect bargaining opportunities through regulations. A large literature in corporate finance analyzes whether regulations can impact executive compensation (see e.g., Frydman and Jenter 2010 and Murphy 2012 for recent discussions). In a reduced form way, regulations would impact the cost of bargaining $k_i(\eta)$ but our analysis of the optimal tax would remain valid taking regulations are given. Ideally, as bargaining is a wasteful effort that shifts resources without any real productive effect, the government would want to completely discourage it, so that pay would always be equal to real economic product. In that case, bargaining effects disappear and we naturally revert to the standard model. However, as long as some bargaining effects exist, our analysis remains relevant.

Differentiated taxation. Some economic sectors or industries might be more prone to bargaining effects than others. For example, less competitive industries have higher rents and hence more scope for bargaining effects. In that case, differentiated tax rates across industries could be desirable. The same argument calls for differentiated tax rates in the standard model if some sectors have a higher labor supply elasticity. In practice, there are two important arguments

[22]The additive form can be written as $\frac{\tau - s}{1-\tau} = \frac{1}{a \cdot e}$ where s is the externality and $\frac{1}{a \cdot e}$ is the conventional Mirrlees term. This additive decomposition in optimal taxation with externalities is well-known since at least Sandmo (1975). Similarly, formula (4) in the case with tax avoidance in the sum of the conventional Mirrlees term and the corrective fiscal externality term $t \cdot e_2/e$ so that $\frac{\tau - te_2/e}{1-\tau} = \frac{1}{a \cdot e}$.

against differentiated taxation. First, it would be difficult to measure bargaining effects for each sector. This uncertainty might allow the better paid lobbyists to argue in favor of preferential tax rates for their industry. Second, differentiated tax rates create additional distortions if there are opportunities to shift income from one sector to another. Lockwood, Nathanson, and Weyl (2012) make this point and consider nonlinear income taxation in a multi-sector model with different externalities across sectors.

Non uniform external effects and link with Rothschild and Scheuer (2012). We have made the strong assumption that aggregate external effects $E(b)$ are spread in a uniform and lumpsum fashion among all individuals, i.e., rent seekers reduce everybody else's earnings uniformly. That simplifies the formula because the government can exactly undo the external effect by simply shifting the schedule and adjusting the demogrant. Realistically, the external effects will not be uniformly distributed. If the government can still adjust the nonlinear tax system to undo the external effect, then our formula carries over unchanged. We provide an example in appendix A.1 showing that this is possible in the case of the discrete version of the Mirrlees model (with a finite number of possible occupations) if we assume that bargaining takes place solely at the top and comes at the expense of lower occupations. This extension shows that our basic formula has wider applicability.

However, if the government cannot undo the external effect, then formulas have to be modified. Rothschild and Scheuer (2012) consider such a model where external effects take place through sector level wages so that rent-seeking effects are proportional to earnings. They allow for both occupational choice across the productive and rent-seeking sectors and intensive responses within sector. They characterize the full optimal nonlinear in such a model (and not solely the optimal top tax rate as we do). Because the nonlinear tax system cannot undo external effects in their model, the formula they obtain is no longer the simple sum of the standard Mirrleesian term and the Pigouvian term. Instead, the externality correction term in their model can be either smaller or larger than the pure Pigouvian correction term depending on whether within-sector or across-sector externalities dominate. Rothschild and Scheuer (2012) also consider the optimal top tax rate and obtain a more general formula as the corrective term is not necessarily equal to the Pigouvian term but it is equal to our formula in the special case of their model where the corrective term equals to Pigouvian term.[23]

[23]This happens when there is a single rent-seeking sector in their model (Section 3.5) or in the case where

One case of interest is when rent-seekers gain solely at the expense of other top earners.[24] In that case, bargaining effects are irrelevant in aggregate among top earners and hence $e = e_1$ and the optimal tax formula boils down to the standard formula $\tau = 1/(1 + a \cdot e_1)$. Effectively, if top earners steal from top earners, decreasing the top tax rate stimulates stealing but this has no effect on the top income share as this is a wash across top earners. Hence, only e_1 matters.

C Putting the Three Elasticities Together

We can put the three elasticities together in a single formula. If there are both avoidance effects and compensation bargaining effects, then we can write the total elasticity of taxable income e as the sum of three terms: $e = (y/z)e_1 + e_2 + e_3$. In case we start from a situation where there is no tax avoidance activity and incomes are equal to marginal products, then $y = z$ and we simply have: $e = e_1 + e_2 + e_3$. For a given tax rate t on sheltered income, we have:

$$\tau^* = \frac{1 + t \cdot a \cdot e_2 + a \cdot e_3}{1 + a \cdot e}. \tag{8}$$

If the government can choose t to fully eliminate tax avoidance, we have $\tau^* = t^* = \frac{1+a \cdot e_3}{1+a \cdot e}$. If government puts a social welfare weight $0 \leq g < 1$ on marginal consumption of top earners (relative to the average), then the optimal top rate formula (8) generalizes to $\tau^* = \frac{1-g+t \cdot a \cdot e_2+a \cdot e_3}{1-g+a \cdot e}$.

II Macro-Level Empirical Evidence

In this section, we use our model to account for the evolution of top tax rates and top incomes in OECD countries. We first analyze US evidence and then turn to international evidence.

A US evidence

US evidence is depicted in graphical form in Figure 1 and key estimates are presented in Table 1. Panel A of Figure 1 depicts the top 1% income shares including realized capital gains (pictured with full diamonds) and excluding realized capital gains (the empty diamonds).[25] Both top income share series display a U-shape over the century. Panel A also displays (on the right

within- and across-sector externalities just cancel out.

[24] E.g., an academic department with a fixed compensation budget in our previous illustration and assuming that all academics are top earners.

[25] Those series are taken from Piketty and Saez (2003). They are based on the family unit (and not the individual adult). Income includes cash market income before individual taxes and credits, and excludes government transfers (such as Social Security benefits, unemployment insurance benefits, or means-tested transfers) as well as non-cash benefits (such as employer or government provided health insurance).

y-axis) the top marginal tax rate for the Federal individual income tax for ordinary income (dashed line) and for long-term realized capital gains (dotted line). Two lessons emerge.

First, considering the top income share excluding realized capital gains which corresponds roughly to income taxed according to the regular progressive schedule, there is a clear negative overall correlation between the top 1% income share and the top marginal tax rate: (a) the top 1% income share was high before the Great Depression when top tax rates were low (except for a short period from 1917 to 1922), (b) the top 1% income share was consistently low between 1932 to 1980 when the top tax rate was uniformly high, (c) the top 1% income share has increased significantly since 1980 after the top tax rate has been greatly lowered. If this correlation is due to a causal relationship from top tax rates to top income shares as in our theoretical model, the overall elasticity of reported incomes is high. For the recent period, the top 1% income share more than doubled from around 8% in 1960-4 to around 18% in last five years, while the net-of-tax (retention) rate increased from 15% (the top marginal tax rate was 85% on average in 1960-4) to 65% (when the top tax rate is 35%). If we attribute the entire surge in the top income share to the decline in the top tax rate, this translates into an elasticity of top incomes with respect to the net-of-tax rate around .5, as shown in column (1), Panel A of Table 1. Column (1) of Panel B in Table 1 also shows a strong correlation between the net-of-tax rate and the top income share with a basic time series regression of the form

$$\log(\text{Top 1\% Income Share}) = \alpha + e \cdot \log(\text{1-Top MTR}) + \varepsilon.$$

This link remains the same when including a linear time trend in the regression.[26] The implied elasticity is around 0.25-0.30 and very significant. Importantly, as the average marginal tax rate faced by the top 1% was smaller than the statutory top rate before the 1970s, our elasticity estimate is a lower bound. The solution would be to instead use the actual average marginal tax rate faced by the top 1% instrumented with the top marginal tax rate (as in Saez, 2004).[27] Importantly, Piketty and Saez (2003) show that the surge in US top income shares since the 1970s is higher in the upper part of the top percentile (top .1% and especially top .01%). The marginal tax rate cuts are also much larger in the upper part of the top percentile so that the resulting elasticities are actually quite similar across sub-groups within the top 1% (Saez, 2004,

[26]Naturally, the correlation disappears when additional polynomials in time are added as identification is based solely on time series variation.

[27]Unfortunately, actual top 1% marginal tax rate series are not available before 1960 and would be very time consuming to construct.

Table 7). It is also conceivable that very high incomes have more *opportunities* to respond to tax rates through avoidance or bargaining effects. This could explain why estimated elasticities below the top 1% are much lower than in the top 1% (Saez, 2004, Table 7).

Second, the correlation between the top shares and the top tax rate also holds for the series including capital gains. Realized capital gains have been traditionally tax favored (as illustrated by the gap between the top tax rate and the tax rate on realized capital gains in the figure) and have constituted the main channel for tax avoidance of upper incomes.[28] Under the tax avoidance scenario, taxable income subject to the progressive tax schedule should be much more elastic than a broader income definition that also includes forms of income that are tax favored. Indeed, in the pure tax avoidance scenario, total real income should be completely inelastic. However, both the graphical analysis of Panel A and the estimates presented in Table 1, column (2) show that the link between the top tax rate is as strong for income including realized capital gains as it is for income excluding capital gains. The time series regressions also generate virtually identical estimates as the series excluding capital gains. This suggests that income shifting responses do not account for much of the long-term evolution in top income shares documented in Figure 1. In future work, it would be useful to sharpen this test by (a) subtracting deductions–such as charitable giving or interest paid on debt–from the narrow income definition to come closer to taxable income, (b) adding forms of income that are non-taxable–such as tax exempt interest, capital gains unrealized till death, or fringe benefits to further broaden the broader income definition. There is no easy route to do this as most of those items are not reported consistently and continuously in income tax statistics. In the short-run, to be sure, there is strong evidence on Panel A of large tax avoidance responses in various tax reform episodes with clear differential responses for top incomes including vs. excluding realized capital gains.[29] But in the long run the income shifting elasticity e_2 (as estimated along

[28]When the individual top tax rate is high (relative to corporate and realized capital gains tax rates), it is advantageous for upper incomes to organize their business activity using the corporate form and retain profits in the corporation. Profits only show up on individual returns as realized capital gains when the corporate stock is eventually sold (see Gordon and Slemrod (2000) for an empirical analysis).

[29]For example, in 1986, realized capital gains surged in anticipation of the increase in the capital gains tax rate from 20 to 28% (Auerbach, 1988), creating a clear spike in the series including capital gains. From 1986 to 1988, income excluding realized capital gains surged as closely held businesses shifted from the corporate form to the individual form, and as many business owners paid themselves accumulated profits as wages and salaries (Slemrod, 1996, Saez, 2004). Such shifting increased reported ordinary income at the expense of realized capital gains, explaining why there is a big discontinuity in income excluding realized capital gains but not in income including realized capital gains. Finally, there is a clear surge in incomes in 1992 in anticipation of the increase in the top tax rate on ordinary income in 1993 due to re-timing in the exercise of stock-options for executives

the ordinary income vs. capital gains margin) appears to be small (say, $e_2 < 0.1$).

Clearly, capital gains are not the only channel through which tax avoidance can occur. Our estimates of e_2 would be biased downward if those alternative tax avoidance channels, such as off-shore accounts or perquisites had sharply declined since the 1960s. However, if anything, it seems that those have increased at the same time as top rates have declined. For the former channel, Zucman (2013) for example shows that a growing fraction of Swiss fiduciary deposits are recorded as belonging to tax havens since the 1970s. For the latter, it is notoriously hard to find historical data, as disclosure rules for perquisites have only recently been imposed[30] but perquisites would have had to be huge pre-1970 to generate a high elasticity of avoidance through that channel.[31]

This analysis has been predicated on the assumption that the link between top tax rates and top income shares is causal. Reverse causality remains a possibility. For example, higher top income shares provide more political power to top earners to influence policy (via lobbying or campaign funding) and leads to lower top tax rates. This would lead to an upward bias in our elasticity estimates (but would not necessarily invalidate the tax avoidance analysis just presented). We come back to this important issue when we consider international evidence.

The even more difficult question to resolve is whether this large responsiveness of top incomes to tax rates is due to supply side effects generating more economic activity as in the standard model or whether it is due to a zero-sum game transfer from the bottom 99% to the top 1% as in the bargaining model. This is critical in order to decompose the total elasticity e into its real (e_1) and bargaining (e_3) components. Panel B of Figure 1 tackles this issue by plotting the evolution of top 1% incomes and bottom 99% incomes adjusting for price inflation.[32] The graph shows clearly that income growth for the bottom 99% was highest in the 1933 to 1973 period when top income tax rates were high and the growth of top 1% was modest. Conversely,

(Goolsbee, 2000). See Saez, Slemrod, and Giertz (2012) for a much more detailed discussion.

[30]Regulation introduced in December, 1978 required firms to disclose only the total amount of remuneration distributed in the form of securities or property, insurance benefits or reimbursement, and personal benefits. Only in 1993 were perquisites and other personal benefits (above a minimum threshold) separately reported. Even then, the data poses problems in terms of transparency and accuracy.

[31]According to Yermack (2006), Grinstein, Weinbaum and Yehuda (2008), and Frydman and Saks (2010), today's perks are significantly larger than even the total taxable pay of top executives pre-1970s, casting doubt upon the idea that perks could have been even larger pre 1970.

[32]To control for changes in the number of adults per family, we plot income per adult (aged 20 and over) assuming that the top 1% income share at the individual adult level is the same as at the family level. This assumption holds true in countries such as Canada where top income shares can be constructed both at the individual and family levels (Saez and Veall, 2005).

the growth of bottom 99% incomes has slowed down since the 1970s when top tax rates came down and top 1% incomes grew very fast. Those findings can be captured by a basic regression analysis of the form:

$$\log(\text{Real Income}_{gt}) = \alpha + \beta \cdot \log(\text{1-Top MTR}_t) + c \cdot t + \varepsilon_t,$$

where g indexes either the Bottom 99% or the Top 1% or the overall average income and t denotes the year. We naturally control for time to capture overall exogenous growth independent of tax policy. The estimates for β, reported in Table 2 Panel C, are positive and highly significant for the top 1% incomes, with a magnitude around .25 very similar to the time series elasticity estimation of Panel B. In contrast, the estimates for β are negative (and just significant at the 5% level with a t-statistics around 2) for the bottom 99%, and close to zero and insignificant for the overall average income. Again, the estimates are very similar for income excluding capital gains in column (1) and for income including capital gains in column (2).

This evidence is consistent with the bargaining model where gains at the top have come at the expense of the bottom. In principle, the estimate β obtained for the overall average income can be used to compute e_1. I.e. if the model is well identified we have: $\beta = \pi \cdot e_1$, where π is the initial income share of top marginal tax rate taxpayers. That is, if we take $\pi = 10\%$,[33] then a doubling of the net-of-top-marginal-tax-rate should lead to a $\beta = 5\%$ rise in the average real income of the economy if the real supply side elasticity e_1 were .5. Since we find that β is close to zero and insignificant for the overall average income, under our identification assumptions, e_1 is also small and insignificant, and that the overall elasticity e comes mostly from bargaining effects through e_3.

This evidence can also be used to rule out the possibility of significant unrecorded tax avoidance effects. That is, assume that in the 1950s-1970s top income earners were escaping high top rates via consumption within the firm or tax havens. Many of those tax avoidance schemes are not recorded in GDP.[34] If such tax avoidance had declined significantly in the recent period, then this should show up as extra economic growth. I.e. in presence of such unrecorded

[33]The exact fraction of taxpayers falling in the top marginal rate bracket varies over time. In recent decades, it is generally larger than the top 1% (it is often closer to the top 2%-3%), so $\pi = 10\%$ should be viewed as a lower bound (implying that the estimates for e_1 should be viewed as upper bounds).

[34]For example, consumption within corporations such as fancy offices or restaurants, corporate jets, etc. are intermediate costs of production and hence unrecorded in GDP estimates. Incidentally, we know of no evidence showing that such intermediate consumption has declined since the 1960s (anecdotal evidence suggests that it might have risen, along with the rise of cash compensation). Importantly, other forms of tax avoidance such as deferred compensation or legal income shifting toward fringe benefits are recorded in GDP.

tax avoidance activities, the estimate β should actually be equal to: $\beta = \pi \cdot (e_1 + e_2)$. This suggests that the overall elasticity e comes mostly from e_3 effects.

However, this evidence relies on the strong OLS assumption that any deviation of growth from trend (captured by the error term ε_t) is uncorrelated with the top marginal tax rate. It is conceivable that economic growth could have slowed down in the 1970s for reasons unrelated to the top tax rate decreases. This could have driven down the bottom 99% income growth as well. In that case, the cut in top tax rates could have increased top incomes growth as in the supply side scenario without negatively impacting bottom 99% incomes. Indeed, growth slowed down in many OECD countries after the oil shocks of the 1970s. Therefore, this evidence based on a single country is at best suggestive. Hence, we next turn to international evidence.

B International Evidence

Effects of top tax rates on top income shares. To analyze international evidence, we use data on the income shares of the top 1% from 18 OECD countries, gathered in the *World Top Incomes Database* (Alvaredo et al., 2011) combined with top income tax rate data since 1960. We focus on the period since 1960 because this allows us to include more countries (a number of countries in the top income database have data only for recent decades) and to be able to obtain top tax rate data for all countries. In addition, focus on the recent period is interesting because of the very divergent trends across countries in both top income shares and top tax rates.[35] Top incomes are defined as cash market income excluding capital gains and subject to the regular income tax. Our top income tax rates series include both the central and local government top tax rates on ordinary income. We do not include payroll taxes as those taxes apply only to wage earnings which constitute only a fraction of top 1% incomes and are often capped. We do not include consumption taxes either. Details on the construction of top tax rates and other variables, together with data sources, are in Appendix A.2.

We start in Figure 2 by showing the link across countries between the top tax rate and the top 1% income share for the periods 1960-64 (Panel A) and 2005-09 (Panel B). If the country does not have top income share data for those years, we select the first available five years after 1960 and the most recent 5 years.[36] Panel A shows that there was a very wide dispersion in

[35] Roine, Vlachos, and Waldenstrom (2009) have used the database to explain the long-run determinants of inequality over the full century, including the top tax rate as an explanatory variable among many others. They present overall regressions without focusing specifically on the recent decades as we do here.

[36] For the following 5 countries, the data start after 1960: Denmark (1980), Ireland (1975), Italy (1974),

top tax rates across OECD countries in the early 1960s with rates as low as 45% for Spain and Switzerland and above 80% for the United Kingdom, or the United States. The graph shows that there is virtually no correlation between top income shares and top tax rates in the early 1960s. The implied elasticity, obtained from a simple OLS regression of the log of the retention rate (one minus the top marginal tax rate) on the log of the top 1% share based on those 18 observations is very small (.07) and insignificant.

Panel B of Figure 2 shows a dramatic shift by 2005-9. Top tax rates are much lower than they were in the 1960s with no country above 60% and a number of countries clustering around 40% including the United States, or the United Kingdom. Those two countries have moved from being the highest top tax rate countries in the 1960s to the lowest today. There is also much more heterogeneity in top income shares which vary from a low of 4% for Denmark to a high of almost 18% for the United States. Importantly, there is also a very strong negative correlation between top tax rates and top income shares in 2005-9. As reported in the graph, the implied elasticity for 2005-9 is extremely large at around 1.90 and highly significant.

In order to extend the 1960s vs. today comparison we did for the United States to our 18 OECD countries, Figure 3 plots the change in top income shares from 1960-4 to 2005-9 against the change in the top marginal tax rate for all the countries. The figure shows a very clear and strong correlation between the cut in top tax rates and the increase in the top 1% income share, with some interesting heterogeneity. Countries such as Germany, Spain, Denmark or Switzerland which did not experience any significant top rate tax cut did not experience increases in top income shares. Among the countries which experienced significant top rate cuts, some experience a large increase in top income shares (all 5 English speaking countries but also Norway and Portugal) while others experience only modest increases in top income shares (Japan). Interestingly, no country experiences a significant increase in top income shares without large top rate tax cuts. The implied elasticity from the OLS regression of the change in the log of the top retention rate on the change of the log of the top 1% share is .47 and highly significant.

Panel A in Table 2 reports estimates from regressions of the form

$$\log(\text{Top 1\% Income Share}_{it}) = \alpha + e \cdot \log(\text{1-Top MTR}_{it}) + \varepsilon_{it}$$

Portugal (1976), Spain (1981). For Switzerland, the data end in 1995 (they end in 2005 or after for all the other countries).

on the complete time series. Column (1) considers the full period 1960-2010, column (2) the early 1960-1980 period and column (3) the 1981-2010 most recent period. Three lessons emerge.

First, full period regressions generate estimates around .3-.4, highly significant, and robust to the introduction of an overall time trend or country fixed effects.[37] Second, the implied elasticity varies significantly across countries with strong effects in English speaking countries, and particularly the United States and the United Kingdom where the elasticity is around 0.5, and much more modest effects in other countries such as Japan, Sweden, or Italy, where the elasticity is close to zero. This suggests that the elasticity likely depends on the institutional set-up of each country. Third, the elasticity was smaller (below .2) in the early period 1960-1980 but increased sharply to .6-.8 in the period 1981-2010, again suggesting that behavioral response varies over time perhaps as institutions changes.

Columns (4) to (6) perform a robustness check on our results, by drawing 500 times a random period between 1960 and 2010 (spanning at least 17 of the 51 years) and a random subset of countries (including at least 6 of the 18 countries) from the full sample to generate the distribution of the elasticity.[38] This exercises serves two purposes. First, it tries to examine how sensitive our results are to the choice of the period and set of countries. In our case, our full period estimates are very close to the median (column (4)) and even the 5th and 95th percentile would be consistent with our message that top tax rates affect top 1% income shares. Secondly, it highlights the wide range of results that could be obtained if one strategically mined the data. This is why we prefer reporting the full range of possible estimates.

Naturally, the strong correlation between top tax rates and top income shares does not prove a causal link. Reverse causality scenarios, e.g., gains in top income shares lead to more political power among the rich and ability to lower the top tax rate, remain a possibility. A striking feature of the evidence however is that, in all countries which experience both a large top tax rate cut and a large increase in top income shares, the surge in top incomes tend to follow the

[37]Estimates using both country and time fixed effects generate smaller elasticities as they rely on year-to-year variation for identification. Our analysis focuses instead on long-run effects of top tax rates.

[38]More precisely, we implement the following algorithm. First, we randomly select a start and an end year (y_0 and y_T respectively) from uniform distributions but imposing the constraint that $y_T - y_0 \geq 17$ to allow for a sufficiently large sample. Then, we randomly select a number x of countries between 6 and 18 (from a uniform distribution). Finally, we randomly draw x countries among our 18 countries. This leaves us with x randomly selected countries for a period y_0 to y_T (the same period for all countries). We then perform all regressions for all 5 specifications (presented in each row) on each selected sample and repeat this procedure 500 times. From the 500 coefficients, we compute the 5th, 50th and 95th percentile of coefficients, standard errors and sample sizes.

top rate cuts, consistent with our elasticity interpretation.[39]

Effects of top tax rates on growth. To distinguish between the supply side vs. the bargaining scenario, we examine the links between top tax rates and economic growth using real GDP per capita (deflated using the GDP deflator and expressed in 2010 US dollars using Purchasing Power Parity as of 2010). We have complete series of GDP per capita and top tax rates from 1960 to 2010 for all 18 OECD countries.

Figure 4 plots the annual real GDP per capita growth from 1960-4 to 2006-10 against the change in the top marginal tax rate for all 18 countries. Panel A plots the raw growth rate while Panel B controls for initial GDP per capita as of 1960 (since most growth theories would suggest that poorer countries tend to grow faster as they catch up with richer ones).[40] Under the supply-side scenario, a cut in top rates translates into additional economic activity among upper incomes, hence higher top income shares but also higher economic growth. In contrast, under the bargaining scenario, a cut in top tax rates generates a "trickle-up" transfer from lower to upper incomes with an increase in top income shares but no additional economic activity.

Both graphs display no visible correlation between the change in top tax rates and growth rates. The countries experiencing the largest increases in top income shares (the US and the UK) have growth rates that are comparable to those of Germany, or Denmark who did not experience large top rate cuts and top income share increases. We show in appendix Figure A1 that there is no correlation between cuts in top marginal tax rates and growth rates for the sub periods 1960-4 to 1976-80 and 1976-80 to 2006-10 either. Panel B of Table 2 provides systematic regression evidence using the complete time series and specifications of the form:

$$\log(\text{Real GDP per Capita}_{it}) = \alpha + \beta \cdot \log(\text{1-Top MTR}_{it}) + c \cdot t + \varepsilon_{it} \quad \text{with} \quad b = \pi \cdot e_1$$

All regressions include a time trend to account for growth. Regressions include the same 18 OECD countries as in panel A for three time periods: 1960 to 2010 in column (1), 1960 to 1980 in column (2), 1981 to 2010 in column (3).[41] The second regression includes country fixed effects. The third regression includes initial GDP per capita. The fourth regression includes

[39] For example, in the United States, top income shares increase sharply after 1981, when the Reagan tax cuts take place. In the United Kingdom, top income shares fall until 1978, and then increase sharply exactly when the Thatcher top rate cuts start in 1979.

[40] Formally, adjusted growth rates are obtained by regressing log(GDP per capita) on log(1-MTR), country fixed effects, a time trend and a time trend interacted with log(GDP). We then estimate the adjusted log(GDP) by removing the interaction component.

[41] In contrast to Panel A, the series are complete for all countries.

initial GDP per capita and the interaction of initial GDP per capita with a time trend (to capture catch-up effects). The fifth regression includes country fixed effects and the interaction of initial GDP per capita with a time trend. Finally, columns (4) to (6) again perform the replication exercise of those results on randomly selected subsamples of countries and periods.

The regressions consistently display negative coefficients across the full period, suggesting that low top tax rates are detrimental to growth. The estimates however are not fully robust to the choice of time period, as seen in columns (2) to (6). Therefore, we can conservatively conclude that low top tax rates do not have any detectable positive impact on GDP per capita. Our preferred bottom row specification including the largest set of controls shows insignificant effects for all three periods. Given the magnitude of our estimate obtained using the largest number of control variables for the most recent period 1981-2010, increasing the retention rate by 100% (as done in the US for top incomes since 1981) would only increase GDP per capita by .8 percent. If we take $\pi = 10\%$, as in the previous sub-section, then $\beta = 0.008$ translates into a real supply side elasticity $e_1 = .08$. Given that the overall elasticity e is about 0.5, this implies that the compensation bargaining elasticity e_3 is larger than e_1.[42] Importantly and as mentioned above, as the top statutory rates in 1960s and 1970s sometimes applied to less than the top 1% of incomes, the average marginal tax rate effectively faced by the top 1% was likely smaller than the statutory top rate. This means that we should scale up our elasticity estimate e and the growth effect β by the same factor $\log(1 - \text{Actual MTR})/\log(1 - \text{Top MTR})$. Hence, this does not affect the ratio of the two estimates. All this tends to imply that e_1 is at most 40% of the total elasticity. Note that the standard error around this estimate is large as the growth regressions are not precisely estimated.

As an important caveat, those regressions rely on a very strong identifying assumption, namely that any deviation of GDP growth from its trend not caused by top tax rates is uncorrelated with the evolution of top tax rates. Many potential factors could invalidate this assumption. For example, if countries cut top tax rates when their growth is expected to slow down (for example if Anglo-Saxon countries during the 1970s feared to be overtaken by Continental Europe and Japan, and opted for the Thatcher-Reagan revolution as a way to fight relative decline), that would generate a spurious negative correlation between growth and the

[42]As was noted above, in the presence of unrecorded tax avoidance effects e_2 we would have $\beta = \pi(e_1 + e_2)$, in which case $e_1 + e_2 = .08$. In any case, this corresponds to an elasticity e_3 of at least 0.4.

net-of-tax top rate - thereby implying that the elasticity e_1 is underestimated.[43] The goal of this analysis is show that, a basic macro-level analysis appears to be more consistent with our non-conventional bargaining model than with the standard model used in tax analysis. To provide more compelling evidence, we next turn to micro evidence using CEO pay.

III Empirical Micro Level Evidence from CEO Pay

A wide literature in corporate finance suggests that CEOs may be able to influence their pay through bargaining effects (see Frydman and Jenter, 2010, Section 4, for a summary of the debate). First, parts of compensation packages are deliberately hidden from shareholders which should not be the case if pay were set competitively (Bebchuk and Fried, 2004, Kuhnen and Zwiebel, 2009). Second, CEOs are frequently rewarded for good outcomes that are not the result of their own effort (such as a booming stock-market) and are not symmetrically punished for unlucky events (Bertrand and Mullainathan, 2001, Garvey and Milbourn, 2006). CEO compensation also decreases after regulatory changes aimed at improving board control (Chhaochharia and Grinstein, 2009). Third, there is widespread malpractice in compensation setting which seems to indicate rent-extraction. For example, 30% of firms from 1996 to 2005 seem to have used 'options backdating' (which consists in choosing the "grant dates" ex-post to allow for the minimal strike price of at-the-money options) (Heron and Lie, 2009, Narayanan and Seyhun, 2008).[44] While evidence for rent-seeking among CEOs is substantial, the link between this practice and top tax rates that is central to our model has not been investigated. Therefore, in this section, we use micro level data on CEO pay to directly investigate whether CEO pay responds to top tax rates and whether this response is due to bargaining effects rather than productive effort. To obtain substantial variation in top tax rates, we consider first US evidence since 1970 and then cross-sectional international evidence for 2006.

A US Evidence since 1970

One of the most compelling pieces of evidence in favor of CEOs being rewarded for luck, as opposed to performance, is from Bertrand and Mullainathan (2001). They show that observed

[43] Factors going in the other direction include the voluntary reduction in working hours that took place in a number of Continental European countries since the 1970s-1980s - which in principle should have reduced their growth performance relatively to Anglo-Saxon countries with long working hours.

[44] Bebchuk, Grinstein and Peyer (2010) further show that this practice occurred more frequently in firms with weak boards.

industry wide performance has large effects on individual CEO pay, even though this cannot be the result of an individual CEO's effort. CEOs might hence be able to bargain in order to extract rents which take the form of being rewarded for lucky outcomes. We repeat the methodology pioneered by Bertrand and Mullainathan (2001) (BM, hereafter), but apply it to a larger sample and longer period spanning both high and low top tax rate periods to test whether rewards for luck are higher when top tax rates are lower.

The BM methodology. Let p be any observable performance measure that a firm can use to set CEO compensation. Suppose that realized firm performance is a function of imperfectly observed effort a of the CEO, an observable luck component p_{luck}, (e.g., industry wide performance) and a random noise component ε, such that $p = a + p_{luck} + \varepsilon$. With an efficient pay contract, observable components independent of labor effort should be filtered out before rewarding the CEO (Holmstrom and Milgrom 1987). Using the notation from our bargaining model, total pay z should be equal to $z = \alpha + \beta(p - p_{luck}) = \alpha + \beta(a + \varepsilon) = y$, where y is the real product of the CEO and β is the optimal strength of incentives. In this case, the bargaining component b is zero ($\eta = 1$). With bargaining however, the CEO pay contract may only partially filter out the luck component, total pay could differ from real product and instead be equal to $z = y + \beta_{luck} p_{luck} = y + b$.

To test empirically for the presence of luck income, we follow BM's two-stage analysis.

First, we estimate the effect of firm performance on CEO pay using the OLS regression

$$\log(pay_{it}) = \beta p_{it} + \gamma_i + \chi_t + X_{it}\alpha + \varepsilon_{it} \tag{9}$$

where pay_{it} is total CEO pay in firm i in year t, p_{it} is a performance measure, described in more detail below, γ_i and χ_t are firm and time fixed effects respectively, and X_{it} are CEO controls, namely a second-order polynomial in age, tenure, and tenure as a CEO. The corresponding estimate of β, denoted β_{OLS}, measures how strongly CEO pay is tied to general firm performance.

Second, to determine whether pay reflects luck, we repeat the same regression (9) by instrumenting performance p_{it} with the instrument $p_{luck,it}$–a measure of the observable firm performance unrelated to the CEO's actions, that is, due to luck from the CEO's perspective. As in BM, we use the asset-weighted mean of the relevant performance measure taken over all other firms in the same 2-digit SIC industry.[45] Finding $\beta_{IV} > 0$ means that there is pay for luck, that

[45] The firm under consideration is excluded from that average. Any firm-year level observation which is such that there is only one reported firm in an industry in that year is dropped.

is, incomplete filtering of observable components which are not the result of CEOs' efforts. If $\beta_{IV} \geq \beta_{OLS}$, there is no filtering at all of the luck component and luck is rewarded at least as much as general performance. As in BM, we use two measures of performance: the growth of net income and the stock market performance (that is, the growth of shareholder wealth i.e., of market value of the firm).[46] We consider total CEO pay which captures all salaries, bonuses, restricted stock grants, long term interest participation payouts, the value of option grants and all other payments made. Our data comes from several sources, described in detail in Appendix A.2. For CEO compensation data, we append the Forbes 800 compensation data (available for 1970 to 1991) to the Execucomp data (since 1992). For firm-level information, we use the COMPUSTAT-CRSP database. Overall, we cover the years from 1970 to 2010, and our sample -after accounting for missing variables as described in the Appendix- includes on average 550 firms per year before 1992 and 700 thereafter.

The effect of high vs. low top tax rates. Our bargaining model predicts that top tax rates should reduce wasteful bargaining effort by reducing the returns to it, and, inversely, lower top tax rates should increase wasteful rent-seeking. Hence, higher top tax rates should also decrease luck income. To determine the effect of top retention rates on luck income, we perform the aforementioned OLS and IV regressions separately in the high tax period (pre-1987 when top tax rates were at or above 50%) and in the low tax period (post-1986 when top tax rates were below 40%) as depicted in Figure 1. Note that higher top tax rates could also reduce pay for (real) performance; this is why β_{IV} needs to be considered *relative* to β_{OLS} for each time period.

Table 3 shows our results for the high top tax rate period before 1987 (Panel A) and for the low top tax rate period after 1986 (Panel B). Columns (1) and (2) use the log of net income as the performance measure while columns (4) and (5) use the log of shareholder wealth.

Interestingly and consistent with BM's findings, there is strong evidence of pay-for-luck, with $\beta_{IV} > 0$ for both performance measures and both periods. As expected, CEO pay is strongly related to performance, with $\beta_{OLS} > 0$ too. But in the low tax recent period, we even find that $\beta_{IV} > \beta_{OLS}$ for both performance measures (in the earlier, high tax period, this only holds for the log of net income) suggesting that pay is particularly sensitive to industry-wide performance and that pay contracts do not filter out such luck effects at all. For example, in the low tax

[46] Since all regressions include firm fixed effects, we use simply the logs of these two variables instead of their changes.

period, post 1986, an overall 1 percent increase in shareholder wealth increases pay by 0.37 percent, but a 1 percent increase in shareholder wealth due to luck increases CEO pay by 0.95 percent.

The comparison between Panels A and B highlights that pay for luck has been stronger in the low tax period than in the high tax period. Panel C reports the difference in the OLS and IV coefficients between Panels B and A, together with the p-value of the test for significance of the difference.[47] For both performance measures, pay for luck has not only significantly increased in the recent, low-tax period, but it has increased by more than pay for performance.

A simple explanation is that the use of stock-options has exploded in the post-1986 period, i.e., after top tax rates went down. As stock-option compensation automatically rewards for industry wide luck, it is not surprising that pay is more sensitive to luck in the high stock-option period. This fits with Hall and Murphy (2003) who argue that stock-options have been a device that has allowed CEOs to increase their pay because boards and shareholders fail to perceive the real costs of granting them. This intuition is confirmed by our analysis (not reported in the table) that salaries and bonuses, excluding stock-option grants, exhibit less pay for luck.[48]

Discussion and alternative explanations. Our results seem to suggest that CEOs are rewarded for luck and that the prevalence of pay for luck is reduced by top tax rates. One concern that remains, and which was raised for the original BM analysis as well, is that pay for luck might be optimal and consistent with the traditional contracting model. In this view, pay for luck is not the result of rent extraction but rather of a contract which optimally bases pay on measures of luck. There are several arguments that can be brought up in favor of the contracting view, which are also considered and rejected in BM's original paper. First, it could be that CEOs are paid for luck in order to incentivize them to predict luck shocks better, if timely knowledge of these shocks is valuable for a firm. However, as BM explain, the IV approach does not use any between-firm variation: hence all that is captured is the effect of the luck shock on the pay of a CEO with average predictive ability. It is also hard to find a good rationale for why CEOs should be more incentivized to predict luck shocks when retention

[47]For the OLS regressions, we use a simple F-test on the SURE system composed of the equations for both periods, while for the IV we use a χ-squared test on the stacked regressions system.

[48]As robustness checks, we have performed the same analysis using only those Execucomp firms which were already present in the Forbes 800 sample. We have also experimented with dropping the top 0.5% and bottom 0.5% CEOs in terms of pay per year, as well as the 0.5% tails in terms of net income and shareholder wealth. The results are qualitatively consistent, even if the exact magnitudes vary somewhat.

rates at the top increase. Second, pay for luck might not be due to rent-seeking, but simply due to the inability to filter out luck shocks. There are three replies to this concern though. First, as shown by BM, badly governed firms exhibit more pay for luck, which means that there is no fundamental technological barrier to filtering out luck shocks. Secondly, if filtering was not feasible, top MTRs should not affect its prevalence. Finally, the fact that it is so hard to reward based on performance would only strengthen our point that pay is not equal to marginal product and not set optimally.

Third and most importantly however, a CEO's productivity might increase with the industry's performance and drive the response of pay for luck. BM perform several checks to ensure that this is not the case. We provide an additional and novel one.[49] If the industry wide luck shock was increasing the value of the CEO's productivity, it is likely that it would also increase the human capital value of other workers in the industry, and hence their wages. In contrast, if the CEO bargains in order to be rewarded for luck, without a corresponding increase in the value of his human capital, then we would expect other workers' wages to not be affected by the luck shock, or be affected negatively if increased CEO pay comes at their expense. If some workers, besides the CEO, are also able to extract rents, there might still be some luck income detectable in wages. However, it should be less prevalent than for CEOs, if top earners are the ones mainly engaging in rent-seeking activities.

To test this, we use average wages and salaries at the 2-digit SIC industry level since 1970 from the Bureau of Economic Analysis' National Accounts and regress them on our two industry wide performance measures. The results are reported in Table 3, columns (3) and (6). First, workers' wages exhibit much less pay for luck than CEO pay. Panels A and B show that there is essentially no relation at all between industry performance and workers' wages.[50] CEOs seem to be more rewarded for luck than the average worker. Second, the comparison of Panels A and B (together with the formal tests in Panel C) confirms that workers' pay for luck was not sensitive at all to the decline in top tax rates. This is consistent with our CEO rent-extraction theory and inconsistent with the productivity explanation. Naturally, this test is not definitive as luck shocks could affect differently the productivity of CEOs vs. other workers. Hence, the possibility remains that CEOs true marginal product of effort could vary with the performance of the industry.

[49]This idea was kindly suggested to us by Marco Bassetto.

[50]Note that these regressions control for time and industry fixed effects.

B Microevidence: International CEO Compensation

We now turn to international evidence on CEO compensation and top tax rates, using a novel dataset constructed by Fernandes, Ferreira, Matos and Murphy (2013). It combines detailed information on CEO pay from 14 countries in 2006, built from the BoardEx and Execucomp databases, information on stock ownership (from LionShares), firm performance (from Worldscope and Datastream), and firm governance. We can further quantify the importance of our bargaining channel using those data by asking two additional questions: First of all, does controlling for firm performance still leave CEO pay dependent on top tax rates? Secondly, how does the effect of top tax rates on CEO pay depend on the quality of a firm's governance? Identification is based on cross-sectional variation in top tax rates across countries.

As regards the first question, a standard supply side story would imply that top tax rates should have no effect on CEO pay after controlling for firm performance. This is because if tax rates affect labor supply or effort, this would be captured by firm performance. In a bargaining story though, there would still remain an independent, additional negative effect of top tax rates on CEO pay. Naturally, this argument hinges on correctly and comprehensively capturing firm performance. This is why we use an extended set of different performance measures: firm sales, stock market return and its standard deviation, leverage, and Tobin's q. This gives us a quite complete picture of firm performance and hence, of the margins that a CEO can affect through his productive effort. Appendix A.2 describes these variables in detail.

Figure 5, Panel A depicts the strong, negative correlation between log total pay and top marginal tax rates across countries. Table 4 column (1) confirms this simple correlation. Retention rates at the top strongly increase CEO pay, with an elasticity of 1.97. Panel B of Figure 5 repeats Panel A, but controlling for the aforementioned firm performance measures, as well as for CEO characteristics, namely a second-order polynomial in age and tenure, and education. It is striking that the relation remains starkly negative. Column (2) of Table 4 adds the aforementioned firm performance, CEO characteristics controls and industry fixed effects. The coefficient on the top retention rate remains virtually unaffected at 1.90. As already alluded to, if a CEO only reacted to top tax rates by adjusting his productive effort, as in the standard labor supply model, then his reduced effort would translate into a reduced firm performance, and this would be captured directly by the performance measures in the regression. By contrast, if a CEO also adjusted his rent-seeking effort, then we would expect the effect of retention rates to remain strongly positive, despite controlling for the change in firm performance resulting from the adjustment of productive effort. The similarity of the elasticities in columns (1) and (2) seems to indicate that almost none of the effect of taxes on pay goes through productive effort, that is, through firm performance.[51] Those results also imply that different industrial composition across countries (such as large financial sectors in the United States and the United Kingdom) cannot explain away the link between top tax rates and CEO pay.

Turning to the second question, we are interested in exploring the effect of governance measures on tax rate elasticities. Our bargaining model implies that, if the rent-seeking channel were important, the effect of taxes in well-governed firms would be weaker, as CEOs would only adjust along the productive, supply side margin instead of along both the labor supply and rent-seeking dimensions. To test this, we construct a governance index, based on five governance measures, described in detail in Fernandes et. al. (2013) and in our Appendix A.2: Insider ownership, institutional ownership, whether the CEO is also the chairman of the board, the average number of positions that boards members hold at other companies' boards, and the fraction of independent board directors. The governance index is normalized to have mean zero and standard deviation equal to 1. A higher index represents better governance.[52] Column (3) shows that in better governed firms, total CEO pay is lower, as we might expect, if good governance puts a brake on excessive compensation. More interestingly, column (4) shows the interaction of the governance index with the retention rate. Consistent with our bargaining theory, the retention rate increases CEO pay, but less so in well-governed firms. Note that even the better governed firms seem to suffer from managerial rent extraction. This suggests that better regulation and stricter governance rules are not enough: one needs taxation.

Finally, columns (5) and (6) examine the effect of the retention rate and the governance index on log salary and log equity and bonus pay respectively. The elasticity for bonuses and equity pay is extremely large at 4.68, while the elasticity for salaries is small at .35. The fact that bonuses and equity pay adjust much more than salaries is again at least a weak support

[51] One might make two objections to this argument. First, that CEO effort could be affecting other measures of firm performance that we are not controlling for. Our measures are however very complete and likely very highly correlated with any omitted measures. For example, if CEO effort also affects future firm performance, this should be captured in the stock market return as well as in Tobin's q. Another argument is that CEOs may not be paid for performance, if there are other contracting problems, such as multitasking (Holmstrom and Milgrom, 1991) for example. Then, however, the supply side story of a fair reward becomes more shaky and it is hard to see how pay would be equal to marginal product.

[52] We also explored other possible indices, as well as the effect of each of the governance measures separately. The results are usually in the right direction, but not always significant.

for our rent-seeking theory. Indeed, it seems easier to extract money in an occasional bonus or through stock in the company than it is to explicitly change one's base salary, presumably written in a formal contract and potentially subject to closer and easier shareholder scrutiny.

Overall, those additional results using international CEO compensation data support our rent-seeking and bargaining channel hypothesis and are also consistent with our macro evidence.

IV Policy Implications and Conclusions

We can now bring together our theoretical and empirical analysis to evaluate the plausibility and policy consequences of each of the key scenarios that have been put forward to explain the surge in top incomes in recent decades. The tax implications are summarized in Table 5.

(0) Skill-biased technological change. This scenario posits that technological progress has been skill-biased and has favored top earners relative to average earners. In the case of top earners, this hypothesis takes the form of "Winner-Takes-All" theories whereby highly talented individuals can deploy their skills on a broader and worldwide market, hence increasing the marginal product of any given unit of talent. The theory of skill-biased technological progress is largely independent of behavioral responses to taxation.

This scenario cannot explain why only some OECD countries have experienced a surge in top income shares and why that surge has been highly correlated with the drop in top marginal tax rates. Indeed, all OECD countries have been subject to similar technological and globalization forces and hence should have experienced the same change in income concentration under the basic skill-biased technological change scenario. It also cannot explain why CEO pay is so correlated with the top tax rate (Figure 5) even controlling for firms' and CEOs' characteristics.

(1) Supply side tax effects ($\mathbf{e}_1$). This scenario posits that the drop in top tax rates has led to an increase in top income shares through a standard supply side effect whereby highly skilled individuals work and earn more. In this case, the standard model is valid and there is no avoidance nor bargaining elasticity. If this scenario is correct, then we can interpret the overall cross-country elasticity $e = 0.5$ as deriving from standard supply side effects: $e = e_1 = 0.5$ and $e_2 = e_3 = 0$. With $a = 1.5$ (the Pareto coefficient currently prevailing in the U.S.), the top revenue maximizing tax rate would then be $\tau^* = 1/(1+a \cdot e) = 57\%$ (see Table 5). With $a = 2.0$ (prevailing in many European countries), the top tax rate maximizing tax revenue would only be $\tau^* = 50\%$. This is less than the effective top tax rate (taking into consideration all taxes) currently applied in a number of European countries (Figure 2B). Hence, decreasing top tax rates would be a desirable policy both from the point of view of top earners but also from the point of view of the bottom 99% as taxes collected from upper incomes would increase. This would also imply that the high top US tax rates of the 1970s were set well above the revenue maximizing rate.

However, this scenario creates three major difficulties. First, it somewhat strains credibility to believe that the top 1% earners in the U.S. had enough leeway to be able to drastically increase their work effort. Any objective measure of labor supply such as hours of work or based on retirement behavior does not show any such large increase.[53]

Second, the link between the surge in top income shares and top rate cuts is not perfect. Some countries, such as Japan, have cut their top tax rates about as much as the United States and yet have experienced no surge in top shares. It is also equally difficult to understand why top rates are not correlated with top income shares in the 1960s. This suggests that the behavioral response to taxes might depend on the tax system and institutions rather than on some universal preferences on work and leisure.

Third, and most importantly, the supply side scenario implies that the surge in top incomes is due to additional economic activity and does not come at the expense of lower incomes. Therefore, countries who cut their top tax rates should have experienced more economic growth than other countries a prediction that is not borne out by our simple cross-country analysis. Accordingly, a large e_1 does not seem plausible.

(2) Tax avoidance effects ($\mathbf{e_2}$). In this scenario, the link between top income shares and top tax rates is due to a large avoidance elasticity. When tax rates are high, top income earners find ways to exploit loopholes and report less of their taxable income.

Under this scenario, the avoidance elasticity is large while the standard supply side elasticity is modest. Under the current US tax regime with its existing loopholes, the optimal tax rate should be $\tau = (1 + t \cdot a \cdot e_2)/(1 + a \cdot e)$. It is difficult to estimate t accurately but tax avoidance exploits primarily deferral rules and the favorable treatment of capital gains, so that a marginal tax rate t of 20% is perhaps reasonable. If we assume $e = 0.5$, $e_1 = 0.2$, $e_2 = 0.3$, $e_3 = 0$,

[53] For example, Moffitt and Wilhelm (2000) show that, while top incomes surged after the Tax Reform Act of 1986, hours of work of top earners measured in the Survey of Consumer Finances did not change.

$a = 1.5$ and $t = 20\%$, then we obtain a revenue maximizing top tax rate $\tau^* = 62\%$ which is somewhat larger than the 57% from the pure supply side scenario due to the "fiscal externality" (see Table 5). More importantly however, the deeper policy implication is that one needs to first close tax avoidance opportunities, in order to reduce the shifting elasticity and only then increase the top tax rate. As shown in Table 5, if the government can broaden the base and reduce the avoidance elasticity from 0.3 to 0.1 (keeping $e_1 = .2$ constant), then the optimal top tax rate increases to 71%.

(3) Compensation bargaining effects ($\mathbf{e_3}$). Under this scenario, the high top tax rates of the 1960s were part of the institutional setup putting a brake on top compensation through bargaining or rent extraction effects. Lower top tax rates induces top earners to bargain more aggressively for higher pay.

The main difficulty with this scenario is that it is difficult to obtain compelling direct evidence that the surge in top incomes did come at the expense of lower earners. The US evidence over a century is consistent with this scenario. International evidence since 1960 is also consistent with this scenario. Our CEO pay evidence is also suggestive of bargaining effects. From this evidence, reasonable estimates could be $e = 0.5$, $e_1 = 0.2$ (at most), $e_2 = 0.0$, $e_3 = 0.3$ (at least), which together with $a = 1.5$ would imply $\tau^* = 83\%$ (see Table 5). Naturally, these estimates are not sharply identified but they illustrate the critical importance of the decomposition of the overall elasticity into three elasticities.

Our paper has focused on the revenue maximizing top tax rate, which provides an upper bound on top tax rates. However, the marginal welfare weight put on top earners by society is likely to depend on perceptions of whether top pay is fair or not. In the supply side scenario, pay is fair by definition and hence a zero weight can only be justified by strong redistributive motives. In the tax avoidance scenario, the public might perceive upper incomes as taking unfair advantage of the tax system, which might lower the marginal welfare weight that society puts on top earners. In the bargaining model, if top earners are overpaid at the expense of lower paid workers, then top pay would naturally be considered as unfair, which could translate into a very low social welfare weight on top earners. Therefore, social views are likely to further widen the differences in the socially desired level of top earners taxation across the different scenarios relative to our analysis with zero welfare weights.[54] Historically, top pay fairness perceptions

[54]Stantcheva and Saez (2013) develop a systematic optimal tax theory founded on such generalized social

have certainly played a key role in the development of very progressive taxation in the first part of the 20th century in most advanced countries, and will likely play a key role in the future of top tax rates. It is also possible that higher income shares raise the ability of top income groups to influence social perceptions (e.g. by funding think tanks or media outlets that are more pro-rich), thereby creating some reverse causality between income inequality, perceptions and policies. Economists can play a key role in enlightening those perceptions by evaluating empirically which economic model of top incomes determination accounts best for the facts.

References

Alvaredo, Facundo, Anthony Atkinson, Thomas Piketty, and Emmanuel Saez. 2011. *The World Top Incomes Database*, `http://g-mond.parisschoolofeconomics.eu/topincomes/`

Atkinson, Anthony, Thomas Piketty, and Emmanuel Saez. 2011. "Top Incomes in the Long-Run of History", *Journal of Economic Literature* 50(1), 3-71.

Auerbach, Alan. 1988. "Capital Gains Taxation in the United States." *Brookings Papers on Economic Activity*, 2: 595-631.

Bakija, Jon, Adam Cole, and Bradley Heim. 2012. "Jobs and Income Growth of Top Earners and the Causes of Changing Income Inequality: Evidence from U.S. Tax Return Data," Working Paper Williams College, April 2012.

Bebchuk, Lucian, and Jesse Fried. 2004. *Pay without Performance: The Unfulfilled Promise of Executive Compensation,* Harvard University Press: Cambridge.

Bebchuk, Lucian, Grinstein, Yaniv, and Urs Peyer. 2010. "Lucky CEOs and Lucky Directors," *Journal of Finance,* 65(6), 2363–2401.

Bertrand, Marianne, and Sendhil Mullainathan. 2001. "Are CEOs Rewarded for Luck? The Ones without Principals Are," *Quarterly Journal of Economics,* 116(3), 901–932.

Besley, Timothy and Maitreesh Ghatak. 2013. "Bailouts and the Optimal Taxation of Bonus Pay," *American Economic Review: Papers and Proceedings* 103(3).

Chhaochharia, Vidhi, and Yaniv Grinstein. 2009. "CEO Compensation and Board Oversight", *Journal of Finance* 64(1), 231–261.

Feldstein, Martin. 1995. "The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1986 Tax Reform Act." *Journal of Political Economy*, 103(3): 551-572.

Feldstein, Martin. 1999. "Tax Avoidance and the Deadweight Loss of the Income Tax." *Review of Economics and Statistics*, 81(4), 674-680.

Fernandes, Nuno, Miguel A. Ferreira, Pedro Matos, Kevin J. Murphy. 2013. "Are US CEOs Paid More? New International Evidence," *Review of Financial Studies*, 26(2), 323–367.

Frydman, Carola, and Dirk Jenter. 2010. "CEO Compensation," *Annual Review of Financial Economics* 2, 75-102.

marginal welfare weights.

Frydman, Carola, and Raven Saks. 2010. "Executive Compensation: A New View from a Long-Term Perspective, 1936-2005", *Review of Financial Studies* 23, 2099-2138.

Gabaix, Xavier, and Augustin Landier. 2008. "Why Has CEO Pay Increased So Much?," *Quarterly Journal of Ecoonomics* 123(1), 49-100.

Garvey, Gerald, and Todd T. Milbourn. 2006. "Asymmetric Benchmarking in Compensation: Executives are Rewarded for Good Luck but not Penalized for Bad," *Journal of Financial Economics* 82(1), 197-225.

Goolsbee, Austan. 2000. "What Happens When You Tax the Rich? Evidence from Executive Compensation." *Journal of Political Economy*, 108(2): 352-378.

Gordon, Roger, and Joel Slemrod. 2000. "Are 'Real' Responses to Taxes Simply Income Shifting Between Corporate and Personal Tax Bases?" In *Does Atlas Shrug? The Economic Consequences of Taxing the Rich*, ed. Joel Slemrod, New York: Russell Sage Foundation.

Grinstein, Yaniv, David Weinbaum, and Nir Yehuda. 2008. "Perks and Excess: Evidence from the New Executive Compensation Disclosure Rules." Working paper, Cornell University, Johnson Graduate School of Management, Cornell University

Hall, Robert E. 2005. "Employment Fluctuations with Equilibrium Wage Stickiness," *American Economic Review*, 95 (1), 50–65.

Hall, Brian, and Kevin Murphy. 2003. "The Trouble with Stock Options," *Journal of Economic Perspectives* 17(3), 49–70.

Heron Randall A., and Erik Lie. 2009. "What Fraction of the Stock Option Grants to Top Executives Have Been Backdated or Manipulated?" *Management Science,* 55(4), 513–525.

Holmstrom, Bengt, and Paul Milgrom. 1987. "Aggregation and Linearity in the Provision of Intertemporal Incentives", *Econometrica* 55(2), 303–328.

Holmstrom, Bengt, and Paul Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design", *Journal of Law, Economics, and Organization* 7, 24–52.

Hungerbuhler, Mathias, Lehmann, Etienne, Parmentier, Alexis, and Van Der Linden, Bruno. 2006. "Optimal Redistributive Taxation in a Search Equilibrium Model," *Review of Economic Studies,* 73, 743–767.

Kleven, Henrik, Claus Kreiner, and Emmanuel Saez. 2009. "Why Can Modern Governments Tax So Much? An Agency Model of Firms as Fiscal Intermediaries," NBER Working Paper No. 15218.

Kleven, Henrik and Esben Anton Schultz. 2012. "Estimating Taxable Income Responses using Danish Tax Reforms", LSE Working Paper.

Kopczuk, Wojciech. 2005. "Tax Bases, Tax Rates and the Elasticity of Reported Income." *Journal of Public Economics*, 89(11-12): 2093-2119.

Kuhnen, Camelia, and Jeffrey Zwiebel. 2008. "Executive Pay, Hidden Compensation and Managerial Entrenchment." Working Paper, Northwestern University

Lindsey, Lawrence. 1987. "Individual Taxpayer Response to Tax Cuts: 1982-1984, with Implications for the Revenue Maximizing Tax Rate," *Journal of Public Economics* 33, 173-206.

Lockwood, Benjamin, Charles G. Nathanson, and Glen Weyl. 2012. "Taxation and the Allocation of Talent," University of Chicago Working Paper.

Moen, Espen R. 1997. "Competitive Search Equilibrium" *Journal of Political Economy* 105(2), 385-411.

Moffitt, Robert, and Mark Wilhelm. 2000. "Taxation and the Labor Supply Decisions of the Affluent." In *Does Atlas Shrug? The Economic Consequences of Taxing the Rich*, ed. Joel Slemrod, 193-234. New York: Russell Sage Foundation and Harvard University Press.

Murphy, Kevin J. 2012. "The Politics of Pay: A Legislative History of Executive Compensation", in Jennifer Hill and Randall Thomas (eds.), *Research Handbook on Executive Pay*, Edward Elgar Publishers.

Narayanan, M.P. and S. H. Nejat. 2008. "The Dating Game: Do Managers Designate Option Grant Dates to Increase their Compensation?," *Review of Financial Studies* 21, 1907–45.

Piketty, Thomas, and Emmanuel Saez. 2003. "Income Inequality in the United States, 1913-1998," *Quarterly Journal of Economics* 118(1), 1-39, series updated to 2008 in July 2010.

Piketty, Thomas, and Emmanuel Saez. 2013 "Optimal Labor Income Taxation," in A. Auerbach, R. Chetty, M. Feldstein, and E. Saez (eds.) *Handbook of Public Economics,* Volume 5, Amsterdam: Elsevier-North Holland.

Roine, Jesper, Jonas Vlachos, and Daniel Waldenstrom. 2009. "The Long-Run Determinants of Inequality: What Can We Learn from Top Income Data?" *Journal of Public Economics*, 93(7-8): 974-988.

Rosen, Sherwin. 1981. "The Economics of Superstars," *American Economic Review* 71(5), 845-858.

Rotemberg, Julio. 2002. "Perceptions of Equity and the Distribution of Income," *Journal of Labor Economics* 20, 249-288.

Rothschild, Casey, and Florian Scheuer. 2012. "Optimal Taxation with Rent-Seeking" NBER working paper No. 17035, revised version December 2012.

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

Saez, Emmanuel. 2004. "Reported Incomes and Marginal Tax Rates, 1960-2000: Evidence and Policy Implications." in James Poterba, ed., *Tax Policy and the Economy*, 18: 117–174.

Saez, Emmanuel, Joel Slemrod, and Seth Giertz. 2012. "The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review," *Journal of Economic Literature*, 50(1), 3–50.

Saez, Emmanuel, and Stefanie Stantcheva. 2013. "Generalized Social Welfare Weights for Optimal Tax Theory," NBER Working Paper No. 18835.

Saez, Emmanuel, and Michael Veall. 2005. "The Evolution of High Incomes in Northern America: Lessons from Canadian Evidence," *American Economic Review,* 95(3), 831-849.

Sandmo, Agnar. 1975. "Optimal Taxation in the Presence of Externalities," *Swedish Journal of Economics,* 77, 86–98.

Slemrod, Joel. 1996. "High Income Families and the Tax Changes of the 1980s: The Anatomy of Behavioral Response." In *Empirical Foundations of Household Taxation*, eds. Martin Feldstein and James Poterba, University of Chicago.

Slemrod, Joel, and Wojciech Kopczuk. 2002. "The Optimal Elasticity of Taxable Income." *Journal of Public Economics*, 84(1), 91-112.

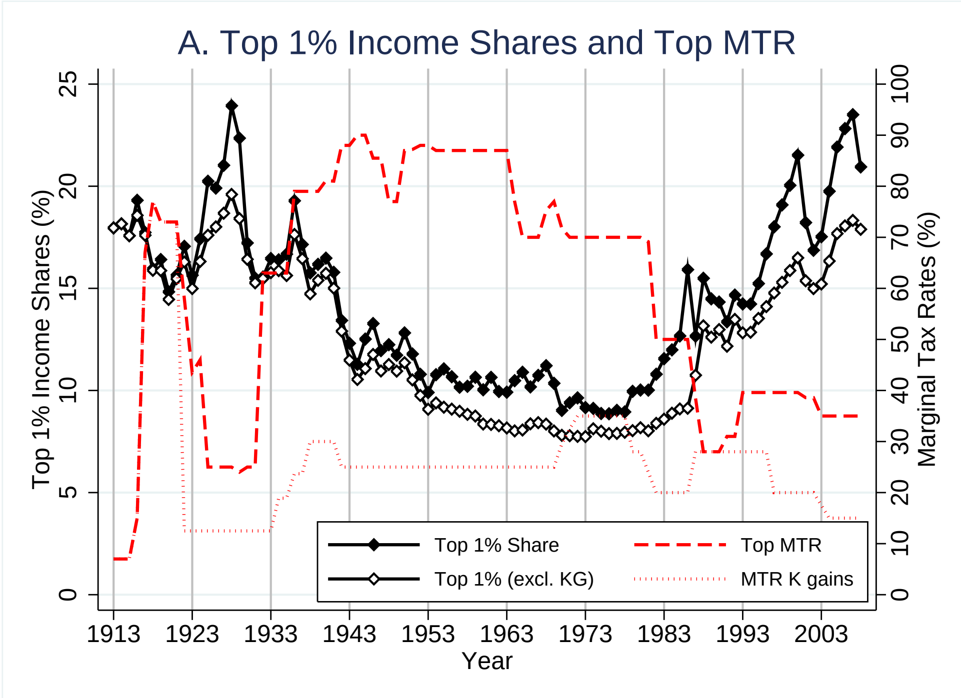
Slemrod, Joel, and Shlomo Yitzhaki. 2002. "Tax Avoidance, Evasion and Administration." in Alan Auerbach and Martin Feldstein, ed., *Handbook of Public Economics*, Volume 3, 1423-1470, Amsterdam: North-Holland.

Sorensen, Peter B. 1999. "Optimal Tax Progressivity in Imperfect Labour Markets," *Labour Economics* 6, 435-452.

Stantcheva, Stefanie. 2011. "Optimal Taxation with Adverse Selection in the Labor market," MIT Working Paper.

Yermack, David. 2006. "Flights of Fancy: Corporate Jets, CEO Perquisites, and Inferior Shareholder Returns," *Journal of Financial Economics* 80, 211-242.

Zucman, Gabriel. 2013. "The Missing Wealth of Nations: Are Europe and the U.S. net Debtors or net Creditors?", forthcoming *Quarterly Journal of Economics*.



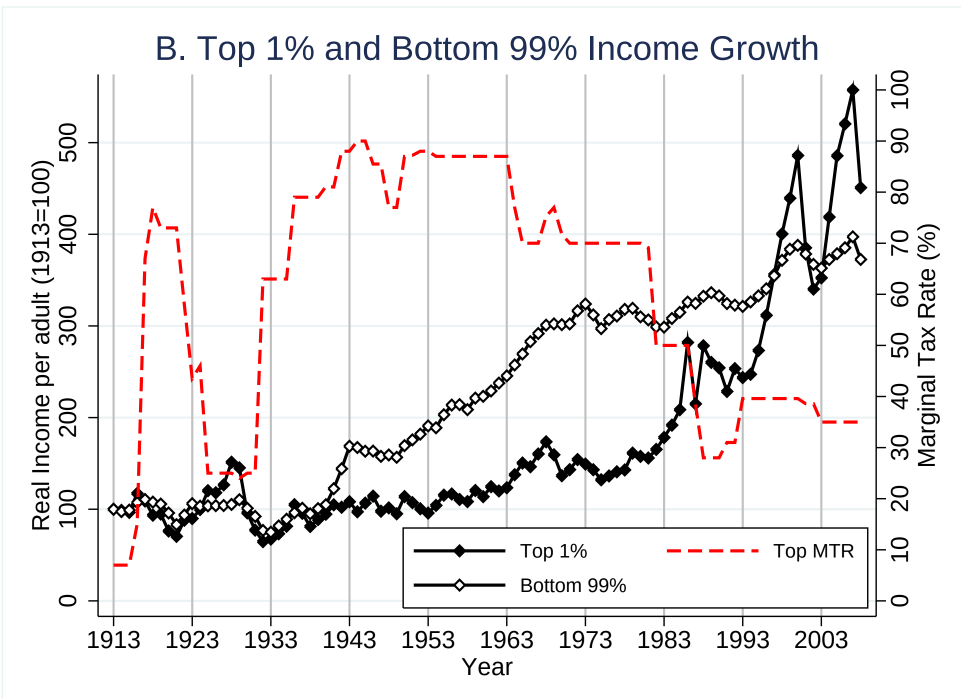


Figure 1: Top Marginal Tax Rates, Top Incomes Shares, and Income Growth: US Evidence

Panel A depicts the top 1% income shares including realized capital gains in full diamonds and excluding realized capital gains in empty diamonds. Computations are based on family market cash income. Income excludes government transfers and is before individual taxes (source is Piketty and Saez, 2003, series updated to 2008). Panel A also depicts the top marginal tax rate on ordinary income and on realized long-term capital gains (source is Tax Policy Center). Panel B depicts real cash market income growth per adult of top 1% incomes and bottom 99% incomes (base 100 in 1913), assuming that individual adult top 1% and bottom 99% shares are the same as top 1% and bottom 99% family based shares.

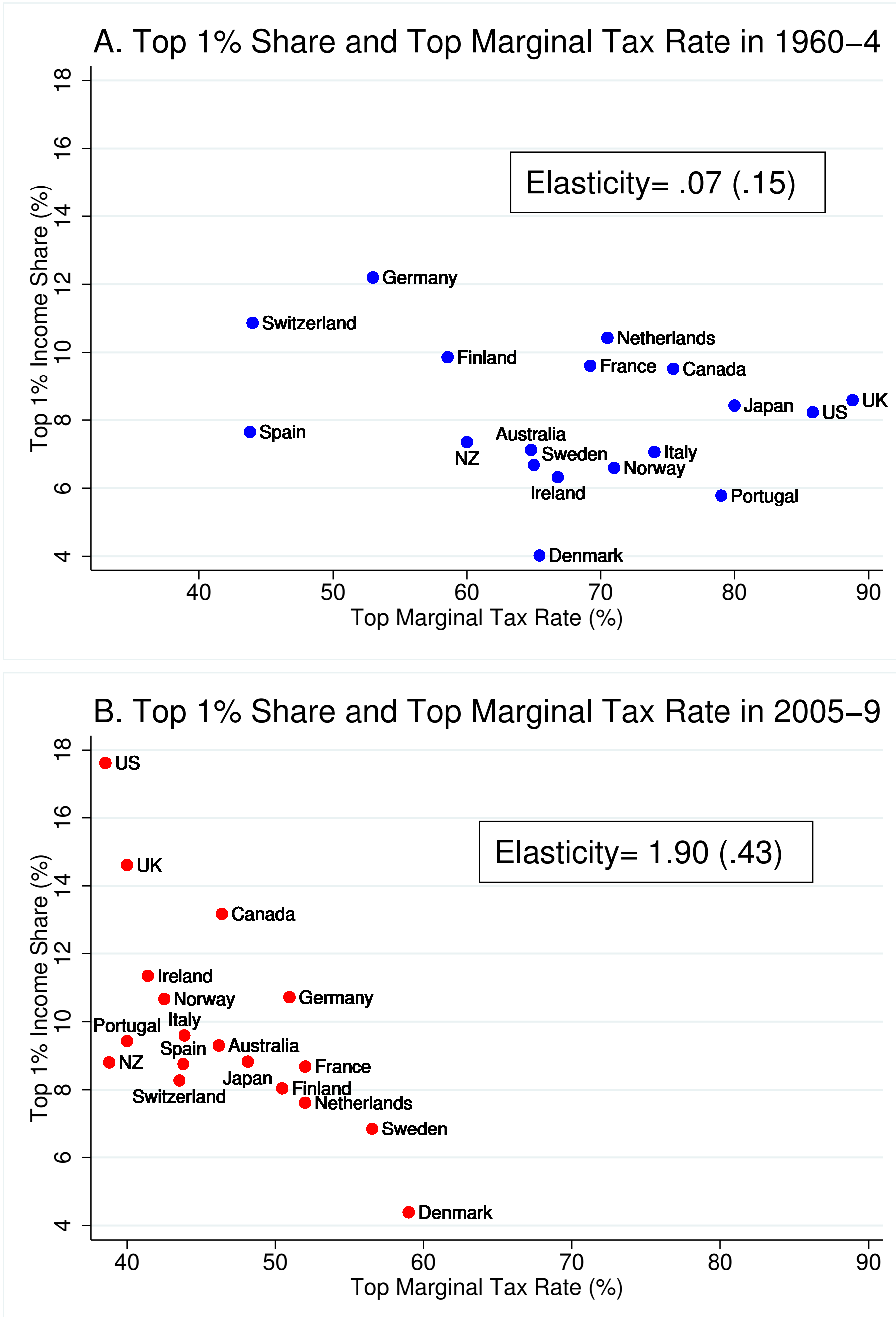


Figure 2: Top Income Shares and Top Marginal Tax Rates: International Evidence

The figure depicts the top 1% income shares and top income tax rates (including both central and local government individual income taxes) across 18 OECD countries in 1960-64 (Panel A) and 2005-09 (Panel B). Source for top income shares is the *World Top Incomes Database*. Source for top income tax rates is OECD and country specific sources. If the country does not have top income share data for those years, we select the first five years after 1960 available and the most recent 5 years (full details in appendix A.2). For the following 5 countries, the data start after 1960: Denmark (1980), Ireland (1975), Italy (1974), Portugal (1976), Spain (1981). For Switzerland, the data end in 1995 (they end in 2005 or after for all the other countries). The figures report the elasticity estimate of the OLS regression of log(top 1% share) on log(1-MTR) based on the depicted dots. The correlation between top tax rates and top income shares is much stronger in 2005-09 than in 1960-64.

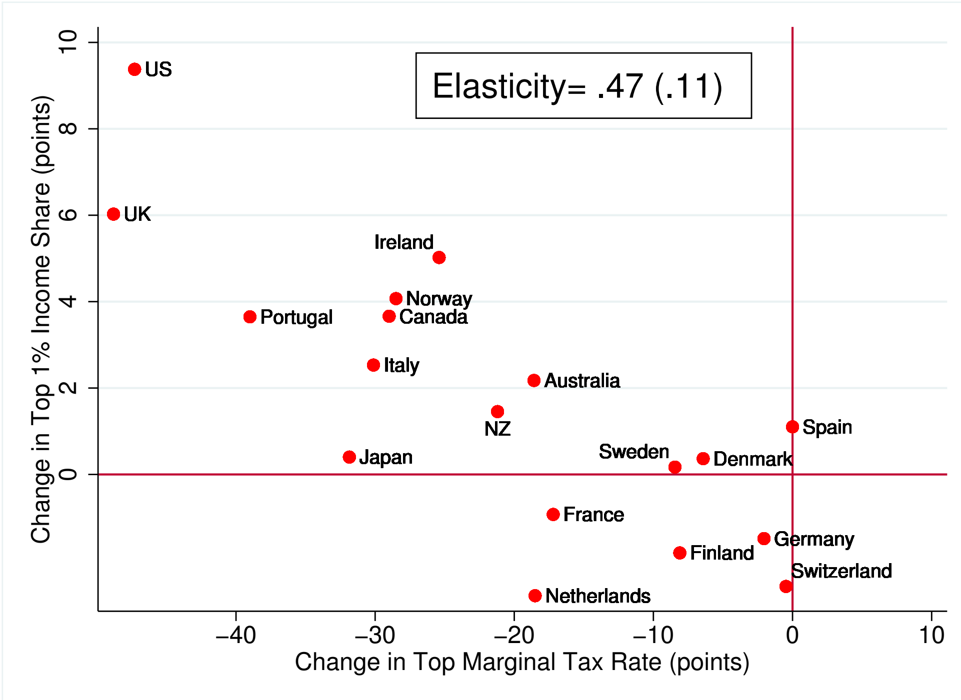


Figure 3: Changes in Top Income Shares and Top Marginal Tax Rates

The figure depicts the change in top 1% income shares against the change in top income tax rate from 1960-64 to 2005-09 based on Figure 2 data for 18 OECD countries. The correlation between those changes is very strong. The figure reports the elasticity estimate of the OLS regression of Δlog(top 1% share) on Δlog(1-MTR) based on the depicted dots.

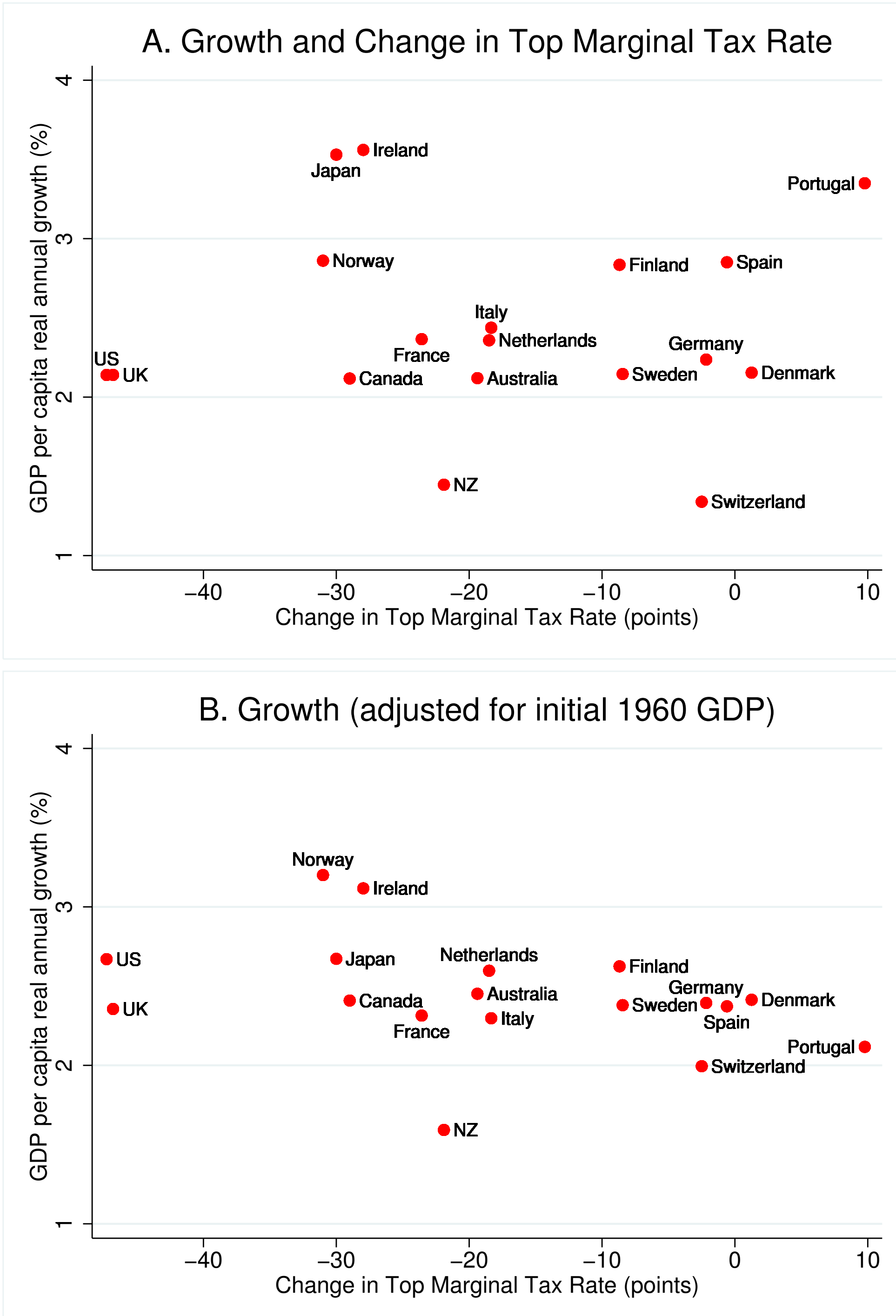


Figure 4: Top Marginal Tax Rates and Growth from 1960-4 to 2006-10

The figure depicts the average real GDP per capita annual growth rate from 1960-64 to 2006-10 against the change in top marginal tax rate. Panel A considers the raw growth rate while Panel B adjusts the growth rate for initial real GDP per capita as of 1960. Formally, adjusted growth rates are obtained by regressing log(GDP) on log(1-MTR), country fixed effects, a time trend and a time trend interacted with demeaned log(GDP). We then estimate adjusted log(GDP) by removing the estimated interaction component time×log(GDP). In both panels, the correlation between GDP growth and top tax rates is insignificant suggesting that cuts in top tax rates do not lead to higher economic growth. Table 2 reports estimates based on the complete time series.

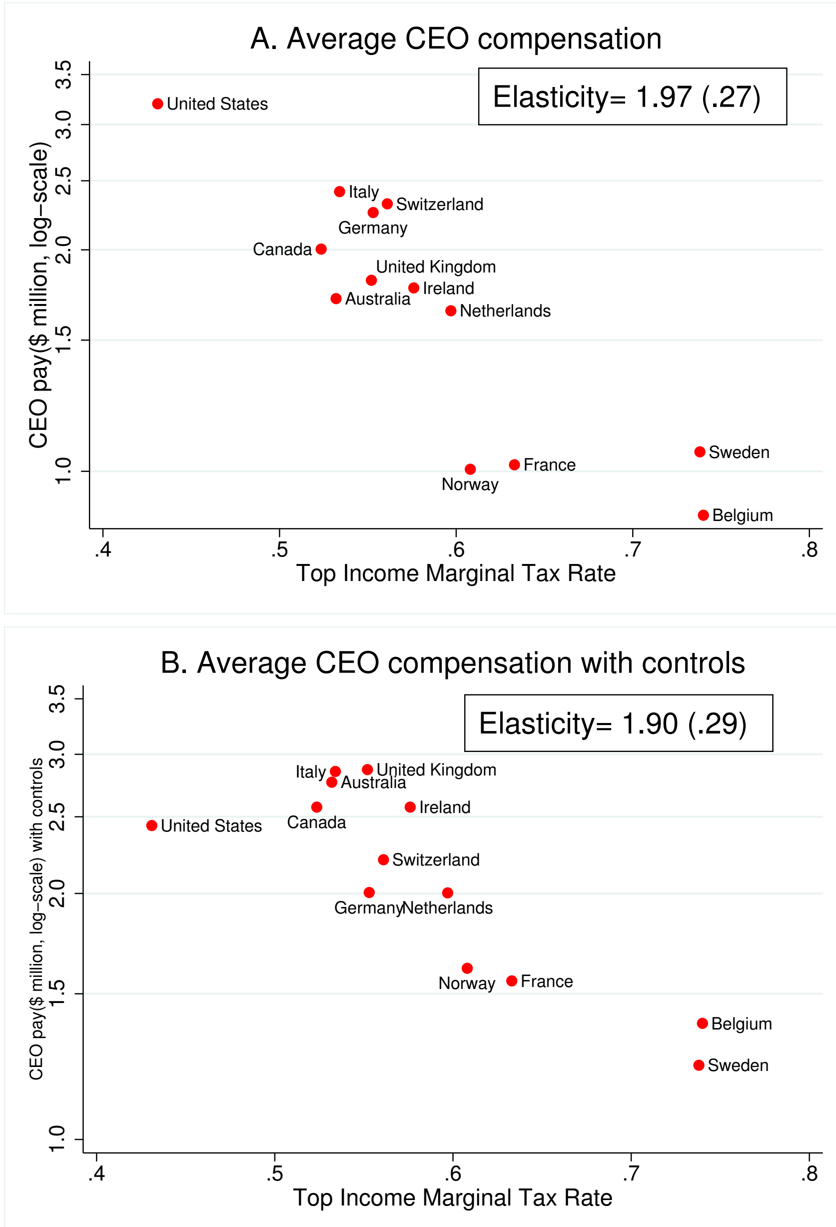


Figure 5: International CEO Pay and Top Tax Rates

Panel A depicts the average CEO pay (in millions of 2006 $) from the Ferreira et al. (2012) dataset against the top individual income tax rate across 13 countries in 2006. Panel B controls for the following variables. (a) firm level variables: log of sales, Tobin's q, return on asset and stock return as well as their standard deviations, leverage, industry dummies. (b) CEO level variables: age of the CEO, age squared, tenure as CEO, and a dummy for college education. (c) governance variables: insider ownership, institutional ownership, the ratio of independent board directors, whether the CEO is also chairman of the board, the average number of board positions held by board members, the number of board members. We run a regression of log(CEO pay) on those (demeaned) control variables, and then plot the average of the residuals for each country. In both panels, the correlations between CEO pay and the top marginal tax rate is very strong. The implied elasticity of CEO pay with respect to the net-of-tax top rate is reported on each panel as well as in Table 4, columns (1) and (2).

Table 1: US Evidence on Top Income Elasticities

		Income excluding capital gains	Income including capital gains (to control for tax avoidance)
		(1)	(2)
A. 1975-1979 vs. 2004-2008 Comparison			
Top Marginal Tax Rate (MTR)	1960-4	85%	85%
	2004-8	35%	35%
Top 1% Income Share	1960-4	8.2%	10.2%
	2004-8	17.7%	21.8%
Elasticity estimate:			
Δ log (top 1% share) / Δ log (1-Top MTR)		0.52	0.52
B. Elasticity estimation (1913-2008): log(share) = α + e*log(1-Top MTR) + c*time + ε			
No time trend		0.25	0.26
		(0.07)	(0.06)
Linear time trend		0.30	0.29
		(0.06)	(0.05)
Number of observations		96	96
C. Effect of Top MTR on income growth (1913-2008): log(income) = α + β*log(1-Top MTR) + c*time + ε			
Top 1% real income		0.265	0.261
		(0.047)	(0.041)
Bottom 99% real income		-0.080	-0.076
		(0.040)	(0.039)
Average real income		-0.027	-0.027
		(0.018)	(0.034)
Number of observations		96	96

Estimates from Panel A are obtained using series from Figure 1 (source is Piketty and Saez, 2003 for top income shares and Tax Policy Center for top marginal tax rate). If the surge in top income shares since 1960 is explained solely by the reduction in the top marginal tax rate, then the elasticity is large, around 0.5. The elasticity is the same for income excluding capital gains and income including capital gains. As capital gains are treated more favorably and are the main channel of avoidance for top incomes, this implies that tax avoidance plays no role in the surge of top incomes in the long-run.

Estimates from Panels B and C are obtained by time-series regressions over the period 1913-2008 (96 observations) and using standard errors from Newey-West with 8 lags. Panel B shows significant elasticities of top 1% income shares with respect to the net-of-tax rate (using the top MTR). Elasticities are virtually the same when excluding or including capital gains and are robust to including a linear time trend in the regression. This shows that there is a strong link in the time-series between top income shares and top MTR as evidenced in Figure 1A.

Panel C shows that real income growth of top 1% is strongly related to the net-of-tax rate (using the top MTR), confirming the results of Panel B. Bottom 99% incomes are negatively related to the net-of-tax rate (using the top MTR) suggesting that top 1% income gains came at the expense of bottom 99% earners. Average incomes (including both the top 1% and bottom 99%) are not significantly related to the net-of-tax rate. Those results suggest that most of the elasticity of top incomes is due to bargaining effects and not real supply side effects.

Table 2: International Evidence on Top Income Elasticities

	All 18 countries and fixed periods			Bootstrapping period and country set		
	1960-2010	1960-1980	1981-2010	Median	5th percentile	95th percentile
	(1)	(2)	(3)	(4)	(5)	(6)
A. Effect of the Top Marginal Income Tax Rate on Top 1% Income Share						
Regression: log(Top 1% share) = α + e*log(1-Top MTR) + ε						
No controls	0.324	0.163	0.803	0.364	0.128	0.821
	(0.034)	(0.039)	(0.053)	(0.043)	(0.085)	(0.032)
Time trend control	0.375	0.182	0.656	0.425	0.191	0.761
	(0.042)	(0.030)	(0.056)	(0.045)	(0.091)	(0.032)
Country fixed effects	0.314	0.007	0.626	0.267	0.008	0.595
	(0.025)	(0.039)	(0.044)	(0.035)	(0.070)	(0.026)
Number of observations	774	292	482	286	132	516
B. Effect of the Top Marginal Income Tax Rate on real GDP per capita						
Regression: log(real GDP per capita) = α + β*log(1-Top MTR) + c*time + ε						
No country fixed effects	-0.064	-0.018	-0.097	0.002	-0.214	0.173
	(0.033)	(0.041)	(0.043)	(0.042)	(0.080)	(0.026)
Country fixed effects	-0.029	-0.082	0.037	-0.004	-0.087	0.071
	(0.014)	(0.016)	(0.019)	(0.016)	(0.031)	(0.011)
Initial GDP per capita	-0.095	-0.025	-0.023	-0.054	-0.149	0.022
	(0.019)	(0.016)	(0.014)	(0.017)	(0.030)	(0.011)
Initial GDP per capita, time*intial GDP per capita	-0.088	0.004	-0.037	-0.060	-0.160	0.012
	(0.017)	(0.011)	(0.014)	(0.016)	(0.030)	(0.011)
Country fixed effects, time*initial GDP per capita	-0.018	0.000	0.008	-0.015	-0.069	0.040
	(0.011)	(0.014)	(0.017)	(0.013)	(0.031)	(0.009)
Number of observations	918	378	540	317	152	576

Panel A presents regression elasticity estimates to the top 1% income share with respect to the net-of-tax top rate. Those estimates are obtained by regressing log(top 1% income share) on the log(1-top MTR). Columns (1)-(3) use the complete panel of top 1% income share series from the World Top Income Database for 18 OECD countries for three time periods: 1960 to 2010 in column (1), 1960 to 1980 in column (2), 1981 to 2010 in column (3). Estimates are not sensitive to the inclusion of a time trend or of country fixed effects. For the following 5 countries, the data start after 1960: Denmark (1980), Ireland (1975), Italy (1974), Portugal (1976), Spain (1981). For Switzerland, the data end in 1995 (they end in 2005 or after for all other countries).

Panel B presents regressions of the log real GDP per capita (2010 PPP) on the log net-of-tax rate. All regressions include a time trend to account for growth. Regressions include the same 18 OECD countries as in panel A for three time periods: 1960 to 2010 in column (1), 1960 to 1980 in column (2), 1981 to 2010 in column (3). In contrast to Panel A, the series are complete for all countries. The second regression include country fixed effects. The third regression includes initial GDP per capita. The fourth regression includes initial GDP per capita and the interaction of initial GDP per capita with a time trend (to capture catching up effects). The fifth regression includes country fixed effects and the interaction of initial GDP per capita with a time trend. Negative numbers imply that high top MTR lead to more growth (in contrast with the standard supply-side scenario). The effect of the top MTR on GDP per capita growth is small and insignificant when using the widest set of controls (last row).

Columns (4) to (6) perform a robustness check by repeating the same regression 500 times on 500 randomly selected samples. More precisely, we randomly select a time period (with a minimum of 17 years, i.e., 1/3 of our 51 year span) common to all countries, a subset of countries (between 6 and 18, i.e., at least 1/3 of our sample). We then compute the 500 coefficients and their standard deviations and report the median (column 4), 5th percentile (column 5), and 95th percentile (column 6). In Panel A, all estimates are positive (highly significant for the median and 95th percentile and mostly insignificant for the 5th percentile but still positive), implying that the correlation between top tax rates and top income shares is robust. In panel B, median estimates are either negative or insignificant. 5th percentile estimates are always negative, while 95th percentile estimate are positive. Overall, there is no systematic evidence that GDP growth is related to top tax rates.

Table 3: US CEO Pay Evidence, 1970-2010

Firm performance measure	Log(net income)			Log(stock-market value)		
Outcome (LHS variable)	Log(CEO pay)	Log(CEO pay)	Log(industry level workers pay)	Log(CEO pay)	Log(CEO pay)	Log(industry level workers pay)
OLS vs. IV	OLS	Industry luck IV	Industry level OLS regression	OLS	Industry luck IV	Industry level OLS regression
	(1)	(2)	(3)	(4)	(5)	(6)
A. Effect of firm performance on log-pay in high-top tax rate period (1970-1986)						
Firm performance (RHS variable)	0.23***	0.34***	0.00	0.28***	0.22*	0.00
	(0.013)	(0.072)	(0.010)	(0.022)	(0.123)	(0.015)
Number of observations	8,632	8,503	890	9,005	8,865	898
B. Effect of firm performance on log-pay in low-top tax rate period (1987-2010)						
Firm performance (RHS variable)	0.27***	0.70***	-0.02	0.37***	0.95***	-0.02
	(0.012)	(0.148)	(0.020)	(0.021)	(0.309)	(0.023)
Number of observations	14,914	14,697	1,422	17,775	17,593	1,443
C. Test for difference between low- and high- top tax rate periods						
Difference Panel B - Panel A	0.04***	0.36*	-0.019	0.09***	0.72**	-0.023
p-value of difference	0.01	0.06	0.440	0.00	0.05	0.46

Notes: The table uses micro-level CEO pay data for the United States from 1970 to 2010 by combining the Forbes 800 CEO data from 1970 to 1991 and the Execucomp CEO data for 1992 to 2010. See Appendix A.2 for details on data sources and variables construction. In columns (1), (2), (4) and (5), we consider total CEO pay including salaries, bonuses, and equity pay (stock-options and stock grants). In columns (3) and (6), we consider the log of the average annual wages and salaries per full time employee at 2-digit SIC level industry level from National Accounts.

The performance measure for the firm is the log of net income in columns (1)-(2) and log of shareholder wealth (defined as log of market value capitalization) in columns (4)-(5). Columns (1)-(2) exclude firms with negative net income. In the IV luck regressions in columns (2) and (5), the performance measure is instrumented with the asset-weighted mean industry performance at the 2-digit SIC level, excluding the firm itself. In columns (3) and (6), the right-hand-side variable in the industry wide (2-digit SIC level) average performance of firms in our sample. Regressions (1), (2), (4), (5) include year and firm fixed effects and a quadratic in CEO age, tenure and tenure as a CEO. Regressions (3) and (6) include year and industry fixed effects. Standard errors reported in parenthesis are clustered at the firm level in columns (1), (2), (4), (5) and the industry level in columns (3) and (6).

Panel A reports regressions based on the 1970-1986 period (when top tax rates were at or above 50%) while Panel B reports results based on the 1987-2010 period (top tax rates below 40%). Panel C tests for the significance of the difference in the coefficients between the 1970-1986 and 1987-2010 periods using an F-test (for OLS) and a chi-squared test (for IV). We report the p-values of the test.

The results highlight two points. First, there is pay for luck among CEOs, as CEOs are rewarded for industry wide performance. This cannot be due to an increase in the human capital value of CEOs during times when the industry performs well since workers' wages do not exhibit any pay for luck. Second, pay for luck has increased in the more recent period since 1987 when top tax rates are lower. In contrast, the sensitivity of workers' wages to industry performance does not seem to have been affected by the change in top tax rates. This is consistent with our bargaining model: CEOs bargain in order to be rewarded for luck and the attractiveness of such rent-seeking increases when top tax rates are lower.

Table 4: International CEO Pay Evidence

Outcome (LHS variable)	Log(CEO pay)	Log(CEO pay)	Log(CEO pay)	Log(CEO pay)	Log(CEO salary)	Log(CEO bonus and equity pay)
	(1)	(2)	(3)	(4)	(5)	(6)
Explanatory variables (RHS variables)						
log(1-Top MTR)	1.97***	1.90***	1.92***	1.90***	0.35*	4.68***
	(0.27)	(0.286)	(0.336)	(0.328)	(0.189)	(0.782)
Governance index			-0.10***	-0.19***	-0.02	-0.26
			(0.020)	(0.038)	(0.072)	(0.201)
log(1-Top MTR)*Governance index				-0.13**	0.06	-0.03
				(0.057)	(0.089)	(0.281)
Firm and CEO controls	no	yes	yes	yes	yes	yes
Number of observations	2,959	2,844	2,711	2,711	2,691	2,711

This table uses the international CEO pay micro-data from Ferreira et al. (2012) (excluding South-Africa). All variables are for year 2006 and include 14 OECD countries (see Figure 5). We regress log(CEO pay) on log(1-Top MTR) in columns (1) and (2). In column (3), we add a firm governance index (normalized to have zero mean and standard deviation equal to one). The governance index combines the following five governance measures: insider ownership, institutional ownership, the ratio of independent board directors, whether the CEO is also chairman of the board, the average number of board positions held by board members. In columns (4)-(6), we add the interaction of log(1-MTR) and governance. Columns (1)-(4) use total CEO pay. Column (5) uses only the salary component of pay while column (6) uses CEO pay excluding the salary component (those other forms of pay are primarily bonuses and equity related pay such as stock grants and stock options). Column (1) does not add any control variables. Columns (2)-(6) control for the following variables: (a) firm level variables: log of sales, Tobin's q, stock return as well as its standard deviation, leverage, and industry dummies. (b) CEO level variables: age of the CEO, age squared, tenure as CEO, and a dummy for college education. Detailed data definition and sources are in Appendix A.3. Standard errors, clustered by country are in parenthesis. All regressions show a strong link between log(1-Top MTR) and log(CEO pay). This link is particularly strong for non-salary compensation. Importantly, the link is not affected for firm level controls (including firm performance) suggesting that the link is not due to observable CEO performance. The link between log(1-Top MTR) and log(CEO pay) is stronger in firms with poorer governance.

Table 5: Synthesis of Various Scenarios

Total elasticity $e = e_1 + e_2 + e_3 =$	0.5

Scenario 1: Standard supply side tax effects		Scenario 2: Tax avoidance effects		Scenario 3: Compensation bargaining effects	
		(a) current narrow tax base	(b) after base broadening		
$e_1 =$	0.5	$e_1 = 0.2$	$e_1 = 0.2$	$e_1 =$	0.2
$e_2 =$	0.0	$e_2 = 0.3$	$e_2 = 0.1$	$e_2 =$	0.0
$e_3 =$	0.0	$e_3 = 0.0$	$e_3 = 0.0$	$e_3 =$	0.3

Optimal top tax rate $\tau^* = (1+ tae_2 + ae_3)/(1+ae)$

Pareto coeffient a =	1.5
Alternative tax rate t =	20%

Scenario 1		Scenario 2		Scenario 3	
		(a) e_2=0.3	(b) e_2=0.1		
$\tau^* =$	57%	τ^* = 62 %	τ^* = 71 %	$\tau^* =$	83%

This table presents optimal top tax rates in the case where the overall elasticity of reported taxable income is e=0.5 in three scenarios depending on how this total elasticity breaks down into the standard labor supply elasticity (e_1), the tax avoidance elasticity (e_2), the compensation bargaining elasticity (e_3). In scenario 1, the only elasticity is e_1. In scenario 2, both e_1 and e_2 are present, income shifted away from the regular tax is assumed to be taxed at rate t=20%. Scenario 2a considers the case of the current narrow base with avoidance opportunities and scenario 2b considers the case where the base is first broadened so that e_2 falls to 0.1 (end hence e falls to 0.3). In scenario 3, both e_1 and e_3 are present. In all cases, top tax rates are set to maximize tax revenue raised from top bracket earners.

Web Appendix of

Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities

Thomas Piketty, Emmanuel Saez, and Stefanie Stantcheva

A.1 Bargaining Theory with Non Uniform External Effects

We show here that the assumption that bargaining spillover effects are uniformly distributed can be relaxed in an optimal tax model with a discrete set of occupations. We consider the discrete number of occupations version of the Mirrlees model developed by Piketty (1997) and Saez (2002) (see Piketty and Saez, 2013 for a detailed presentation).

There are $N+1$ occupations with marginal product $y_0 < y_1 < y_2 < .. < y_N$. Occupations $0, .., N-1$ are production workers requiring increasing skills, while occupation N represents managers. Actual pay in each occupation is denoted by $z_0 < z_1 < .. < z_N$. Actual pay may differ from marginal product because of bargaining effects. We assume that only managers can bargain for higher pay (no other occupation can), but the extra pay of managers extracted through bargaining effects can come at the expense of any other occupation in a fully general way. After-tax disposable incomes in each occupation are denoted by $c_0 < .. < c_N$.

To mimic the intensive behavioral response of the standard Mirrlees model (see appendix in Saez, 2002 for a detailed presentation), we assume that the population (normalized to be of measure one) is partitioned into N types with fractions $p_1, .., p_N$ so that $p_1 + .. + p_N = 1$.

An individual i of type $n < N$ with effort cost parameter θ_i can get into occupation n if she exerts effort at cost θ_i, otherwise she is in occupation $n-1$ (at no effort cost). We assume quasi-linear utilities for production workers $u = c - \theta \cdot l$ where $l = 0, 1$ is a dummy variable denoting effort choice. Individual i of type $n < N$ exerts effort and works in occupation n if and only if $\theta_i \leq c_n - c_{n-1}$. We assume that θ_i has distribution $P_n(\theta_i)$ among type n individuals. For $n < N$, the fraction of type n individuals working in occupation n is $P_n(c_n - c_{n-1})$.

We define the marginal tax rate τ_n on the transition from occupation $n-1$ to occupation n by $(1-\tau_n)(z_n - z_{n-1}) = c_n - c_{n-1}$ for $n = 1, .., N$.

An individual of type N can either put no effort and work in occupation $N-1$ or exert effort at cost θ_i and work in occupation N, i.e., become a manager. When in occupation N, the individual can put extra-bargaining effort η (at increasing and convex cost $k(\eta)$) to increase her compensation relative to adjacent occupation $N-1$, i.e., we assume that $z_N - z_{N-1} = \eta \cdot (y_N - y_{N-1})$. When $\eta = 1$, managers are paid their true marginal product (relative to occupation $N-1$). The utility of type N individuals is therefore $u_i = c - l \cdot [\theta_i + k(\eta)]$.

Conditional on being a manager, η is chosen to maximize $c_N - k(\eta) = c_{N-1} + \eta(1-\tau_N)(y_N - y_{N-1}) - k(\eta)$ taking $c_{N-1}, y_{N-1}, y_N, \tau_N$ as given. This leads to first order condition $k'(\eta^*) = (1-\tau_N)(y_N - y_{N-1})$ so that η^* increases with $y_{N-1} - y_N$ and with $1-\tau_N$. The individual of type N

with cost of productive effort θ_i decides to become a manager if and only if $c_{N-1} \leq c_N - \theta_i - k(\eta^*)$, i.e., if and only if $\theta_i \leq \eta^*(1-\tau_N)(y_N - y_{N-1}) - k(\eta^*)$. The fraction of type N individuals working in occupation N is $P_N(\eta^*(1-\tau_N)(y_N - y_{N-1}) - k(\eta^*))$.

We denote by h_n the fraction of individuals in occupation n in general equilibrium. Given the structure of the model, for $n < N-1$, h_n is a function of $c_n - c_{n-1}$ and $c_{n+1} - c_n$. For the top occupation, we have $h_N = p_N P_N(\eta^*(1-\tau_N)(y_N - y_{N-1}) - k(\eta^*))$.

Finally, we assume that the spillover effect of bargaining pay extracted by managers is shared across the different occupations $0, .., N-1$ in any arbitrary way that satisfies the global resource constraint $\sum_n h_n y_n = \sum_n h_n z_n$.

The government chooses $(c_0, \tau_1, .., \tau_N)$ to maximize a standard social welfare function of the form $W = \int_i G(u_i) d\nu(i)$ with $G(.)$ increasing and concave subject to a resource constraint $\sum_{n=0}^N h_n \cdot (z_n - c_n) \geq T_0$ (where T_0 is an exogenous government spending requirement).

First, let us note that choosing $(c_0, \tau_1, .., \tau_N)$ is equivalent to choosing $(c_0, .., c_{N-1}, \tau_N)$. This is because there is no bargaining choice in occupations $0, .., N-1$, and hence behavioral responses depends solely on $c_0, .., c_{N-1}$. In contrast, if the government could choose c_N directly, it could entirely eliminate bargaining issues (as η could not increase disposable income c_N). Therefore, we assume instead that the government chooses τ_N which leaves scope for bargaining.

Second, let us therefore derive the optimal τ_N taking $c_0, .., c_{N-1}$ as fixed and considering a small change $d\tau_N$. This implies that, although $d\tau_N$ might change bargaining and affect lower earnings $z_0, .., z_{N-1}$ through the bargaining spillovers, the government adjusts $\tau_1, .., \tau_{N-1}$ to keep $c_0, .., c_{N-1}$ constant. As a result, neither the utility nor the labor supply choices of individuals of types $1, .., N-1$ are affected.

As in the main text, we assume that occupation N has zero social marginal welfare weight so τ_N is chosen to maximize tax revenue and hence $d\tau_N$ has zero effect on revenue at the optimum,[55]

$$0 = d\left[\sum_{n=0}^{N} h_n \cdot (z_n - c_n)\right] = -h_N dc_N + dh_{N-1}[z_{N-1} - c_{N-1}] + dh_N[z_N - c_N] + \sum_{n=0}^{N} h_n dz_n. \quad \text{(A1)}$$

The second expression is obtained because $c_0, .., c_{N-1}$ stay constant and (hence) labor supply happens only between occupations $N-1$ and N so that $h_0, .., h_{N-2}$ stay constant. Naturally $dh_{N-1} + dh_N = 0$. Note that $\sum_n h_n z_n = \sum_n h_n y_n$ and hence (as $y_0, .., y_N$ are fixed),

$$\sum_{n=0}^{N} h_n dz_n + dh_N[z_N - z_{N-1}] = dh_N[y_N - y_{N-1}], \quad \text{i.e.,}$$

$$\sum_{n=0}^{N} h_n dz_n = dh_N[z_N - z_{N-1}]\left(\frac{1}{\eta^*} - 1\right),$$

using the fact that $z_N - z_{N-1} = \eta^*[y_N - y_{N-1}]$. As $c_N = c_{N-1} + \eta^*(1-\tau_N)(y_N - y_{N-1})$, we have $dc_N = -d\tau_N[z_N - z_{N-1}] + d\eta^*(1-\tau_N)(y_N - y_{N-1})$. Finally denoting $\Delta z = z_N - z_{N-1}$ and

[55] As in the main text, the extension to a positive weight is straightforward.

noting that $c_N - c_{N-1} = (1-\tau_N)\Delta z$, we can rewrite (A1) as

$$0 = h_N d\tau_N \Delta z - \frac{d\eta^*}{\eta^*} h_N (1-\tau_N)\Delta z + dh_N \tau_N \Delta z + dh_N \Delta z \left(\frac{1}{\eta^*} - 1\right). \tag{A2}$$

The first term is the mechanical fiscal effect (absent any behavioral response), the second term is the behavioral bargaining effect response, the third term is the behavioral labor supply response, the last term is the spillover bargaining effect. We can define the elasticities:

$$\varepsilon_1 = \frac{1-\tau_N}{h_N}\frac{dh_N}{d(1-\tau_N)}, \quad \varepsilon_\eta = \frac{1-\tau_N}{\Delta z}\frac{d\Delta z}{d(1-\tau_N)} = \frac{1-\tau_N}{\eta^*}\frac{d\eta^*}{d(1-\tau_N)} \quad \text{and} \quad \varepsilon = \varepsilon_1 + \varepsilon_\eta$$

ε_1 captures the real labor supply response (occupation changes) while ε_η captures the bargaining elasticity (changes in compensation on the job). ε is the total elasticity of top earners including both labor supply and bargaining responses. Dividing (A2) by $h_N \Delta z d\tau_N$, we can then rewrite the first order condition on as:

$$0 = 1 + \varepsilon_\eta - \frac{\tau_N}{1-\tau_N}\varepsilon_1 - \frac{\varepsilon_1}{1-\tau_N}\left(\frac{1}{\eta^*} - 1\right),$$

which can be re-arranged into the following optimal formula:

$$\tau_N = 1 - \frac{\varepsilon_1/\eta^*}{1+\varepsilon}. \tag{A3}$$

We can show that this formula is the discrete model equivalent to the main text formula (7). As discussed in Saez (2002) appendix, in the continuous model, the elasticities are defined for total earnings z instead of marginal earnings Δz. Therefore, the continuous elasticity e is related to the discrete model elasticity ε by $\varepsilon \cdot \Delta z = e \cdot z$, i.e., $\varepsilon = e \cdot z/\Delta z = a \cdot e$ with $a = z/\Delta z$ the Pareto parameter of the top tail. Similarly, for the real labor supply response, we have $\varepsilon_1 = e_1 \cdot \frac{y}{\Delta y} = e_1 \cdot \frac{z}{\Delta z} \cdot \frac{y}{z} \cdot \eta^*$ so that $\varepsilon_1/\eta^* = a(y/z)e_1$. This allows to rewrite (A3) as $\tau = 1 - a(y/z)e_1/(1+ae)$ exactly as the first formula (7) in the main text. In the main text, we define $e_3 = e - e_1/\eta = (\varepsilon - \varepsilon_1/\eta)/a$ which allows to rewrite (A3) as $\tau = (1+ae_3)/(1+ae)$ exactly as in the second formula (7) in the main text.

The key reason why the formula is unchanged (relative to the main text case where bargaining comes at the expense of everybody uniformly) is because the government can adjust the nonlinear tax to fully absorb any change in compensation due to the bargaining externality. This is possible in the discrete model with arbitrary bargaining externalities because there is no direct bargaining among lower level occupations.

Whenever an occupation n can do direct bargaining (as in the top occupation of the model just analyzed), the government cannot control c_n directly, and it is no longer possible for the government to fully offset an external bargaining effect on this occupation. Therefore, our simple formula carries over more generally in situations where the extra pay $z_n - z_{n-1}$ for all occupations n who can bargain for pay is never affected by bargaining externalities, i.e., the bargaining externalities leave $z_n - z_{n-1}$ unchanged. The model we presented had this simple property. A model where low occupations $z_0 < .. < z_K$ have no bargaining power while high occupations $z_{K+1} < .. < z_N$ have bargaining power and where bargaining comes solely at the expense of low occupations also has this property.[56]

The key property that allows us to write the optimal top rate formula as the sum of the standard Mirrleesian top rate formula and the Pigouvian corrective term is whether the distributional effects of marginal changes in rent-seeking can be undone using the nonlinear income tax. This is possible in the main text model as rent-seeking effects are uniform (and hence similar to a demogrant adjustment). This is also possible in the discrete occupation model presented in this appendix as each occupation corresponds to a pay level that can be directly affected by the nonlinear income tax. This is also possible in the special case of the Rothschild and Scheuer (2012) model when there is only a single sector rent-seeking (that they discuss in their Section 3.5). In that case, the externality acts as an atmospheric externality on the wage rate. The single sector model of Rothschild and Scheuer (2012) is a good model for a limited resource activity like fishing where each fisherman imposes an externality by reducing the stock of fish and hence the catch rate of other fishermen. In that case, the standard uniform Pigouvian correction, regardless of income, is naturally the efficient solution (on top of the standard Mirrlees formula). The models we have proposed in the main text and in this appendix aim instead at capturing situations where rent-seeking opportunities are concentrated among top earners. In that case, a large Pigouvian correction for the top tax rate does not necessarily imply that tax rates for low and middle earners should also carry this large Pigouvian correction. Hence, our model is not formally nested in the single sector case of Rothschild and Scheuer (2012). The fact that all models produce very similar top tax rate formulas is a testimony to the value of the sufficient statistics approach we have tried to emphasize in this paper.

In contrast, when the government cannot undo the distributional effects of marginal changes in rent-seeking, the simple additive decomposition on the optimal tax rate into the Mirrlees term and the Pigouvian term is lost. Instead, the Pigouvian term is replaced by a corrective term that maybe smaller or larger than the Pigouvian term. Rothschild and Scheuer (2012)'s two sector model analysis falls into this more complex case. They provide a thorough theoretical analysis of that complex case and precisely characterize when the correcting term is larger or smaller than the Pigouvian term. Hence, their theoretical contribution is more ambitious and general. We focus instead on a simpler case with simpler formulas expressed in estimable sufficient statistics that can be brought to the data. Naturally, it would be valuable in future research to estimate empirically the exact correcting term in Rothschild and Scheuer (2012) to refine and improve the empirical calibration.

This discussion on theoretical models parallels the analysis of optimal taxation in multi-

[56] The continuous model of the main text cannot be simply presented as a model with this property because it is difficult to make a clean/seamless link between bottom occupations with no bargaining and top occupations with bargaining.

sector models with no externalities. As is well known, Stiglitz (1982) showed that the standard Mirrlees formula breaks down in the two-skill type model with endogenous wages because the nonlinear income tax cannot undo wage changes across sectors. In contrast, Piketty (1997) and Saez (2004b) showed that if individuals make solely occupational choices as in the appendix model presented here, the standard (discrete version of the) Mirrlees formula carries over because the nonlinear income tax can then undo wage changes across sectors. Rothschild and Scheuer (2013) consider the general case with both occupational choice across sectors and intensive labor supply responses within sectors. The model of Rothschild and Scheuer (2013) has exactly the same structure as Rothschild and Scheuer (2012)'s rent-seeking model except that there is no externality across sectors. In that case, the general nonlinear income tax cannot undo wage effects across sectors and hence the Mirrlees formula needs to be corrected. The correction term is complex and has a very similar structure to the corrective term in Rothschild and Scheuer (2012). The key point of the model presented in this appendix is to show that the occupational model of Piketty (1997) and Saez (2004b) can be easily extended to accommodate rent-seeking externalities among top earners, while preserving the Mirrlees formula plus Pigouvian correction additive decomposition that is standard in optimal taxation with externalities.

A.2 Data Sources

A.2.1 Top Tax Rate Data

Top tax rates are based on the top statutory individual income tax rate including both central and local governments (when such local individual income taxes exist). The series for top tax rates cover the full period 1960-2010 for all 18 countries.

The primary source is the OECD annual "Taxing Wages" publication which covers the period from the early 1980s to the present. For the period 1975-1983 taxes are summarized in the publication "Personal income tax systems for the period 1975-1983." (OECD, 1986). Top tax rate statistics are also summarized in
`http://www.taxpolicycenter.org/taxfacts/Content/PDF/oecd_historical_toprate.pdf`

The tax rates for all the European countries since 1975 were also gathered in Kleven, Landais, and Saez (2013).

For specific countries, additional sources were used, especially to extend the tax rates back to 1960. These are listed below. Note that secondary sources such as OECD sometimes have typos so that we have tried in all cases to double check the numbers with country specific publications or consulting scholars from specific countries. Our data are available online. We naturally welcome input on any remaining typos to further improve data quality.

Australia: Source is Atkinson and Leigh (2010).

Canada: The tax rates series were taken from and described in Saez and Veall (2007), in their long version from Appendix table E1, which considers the case of Ontario, the largest province.

Denmark: The information for the years before 1975 was obtained from Esben Schultz from income tax statistics.
Finland: The top tax rate data was provided by Markus Jantti based on income tax statistics published annually in Finland.
France: Source is Roine, Vlachos and Waldenstrom (2009).
Germany: Source is Roine, Vlachos and Waldenstrom (2009).
Ireland: tax rates for 1964-1971 obtained from Brian Nolan based on his compilation of individual income tax statistics (top tax rates for 1960-3 are assumed the same as those in 1964 for lack of better information).
Italy: The source is the chapter on top income shares in Italy by Alvaredo in Atkinson and Piketty (2010).
Japan: Local taxes were taken from the National Tax Administration data, as well as Moriguchi and Saez chapter on Japan in Atkinson and Piketty (2010).[57] Local tax rates were assumed to be constant from 1960 to 1975 (due to lack of better information).
Netherlands: The top tax rate data before 1975 was provided by Floris Zoutman based on internal income tax statistics at the ministry of finance in the Netherlands.
New Zealand: Source is Atkinson and Leigh (2010).
Norway: The top tax rate data was provided by Rolf Aaberge based on income tax statistics published annually in Norway.
Portugal: Source is the chapter by Alvaredo on top income shares in Portugal in Atkinson and Piketty (2010), appendix table 11.A.2.
Spain: Source is the chapter by Alvaredo and Saez on top income shares in Spain in Atkinson and Piketty (2010), appendix table 10.A.1. We use the maximum average tax rate of 50% (and then reduced to 44%) for the period 1960-1975.
Sweden: Source is Roine, Vlachos and Waldenstrom (2009).
Switzerland: Numbers obtained from Swiss annual income tax statistics.
United Kingdom: Source is Atkinson and Leigh (2010).
United States: Source for Federal top tax rate is the Tax Policy Center. The average state tax rate is estimated using actual top statutory state income tax rates weighted by the fraction of high income tax returns in each state (as of 2007). We assume that state rates have not changed during the period 1960 to 1975.

A.2.2 GDP and Top Income Share Data

GDP per capital series in constant US dollars are taken from the Bureau of Labor Statistics (Division of International Labor Comparisons, available at `http://www.bls.gov/ilc/`). For years which were missing, we used the *International Historical Statistics* by Mitchell (1998)

[57]We thank Yusuke Narita for kind help with the translation of the Japanese files.

combining the real GDP and population series. The series of GDP per capita cover the full period 1960-2010 for all 18 countries.

Data on the Top 1% income shares comes from the *World Top Incomes Database* (Alvaredo et al. 2011). We use the top 1% pre-tax income shares based on income excluding realized capital gains.

The data for each country cover the following years: Australia (1960-2008), Canada (1960-2009), Denmark (1980-2005), Finland (1960-2009), France (1960-2006), Germany (1960-2007), Ireland (1975-2009), Italy (1974-2009), Japan (1960-2005), New Zealand (1960-2009), Netherlands (1960-2006), Norway (1960-2008), Portugal (1976-2005), Spain (1981-2009), Sweden (1960-2009), Switzerland (1960-1995), United Kingdom (1960-2009), United States (1960-2009).

In the rare case where there is a gap in the data, we use a linear interpolation to fill out the gap. For Germany, we used series including realized capital gains for 1997-2007 (adjusted to match series excluding capital gains in 1995) as series excluding capital gains are not available after 1995. For the Netherlands, we have used series after 1999 from Straathof, Groot, and Mohlmann (2010), spliced to match the World Top Incomes Database in 1999.

A.2.3 US CEO pay data

Data and variable construction for CEO pay and firm characteristics. For firm-level information, we use the COMPUSTAT-CRSP database (quarterly update for North America). We use annual measures (for the current fiscal year) for all variables.

Our measures of performance are (1) the log of net income of the company (COMPUSTAT variable "ni") and (2) the stock market return which is the log of the market value of the firm defined as common shares outstanding (COMPUSTAT variable "csho") multiplied by the annual closing price (COMPUSTAT item "prcc_c"). To capture other firm characteristics, we use firm fixed effects consistently in all our regressions. Because of the latter, using the logs of the variables in the regressions is akin to using as performance measures the growth in net income and the growth in shareholder value (stock market value of the firm).

For CEO pay, we append the Forbes 800 compensation data available for 1970 to 1991 (shared in electronic format by Kevin Murphy) to the Execucomp data (since 1992). The Forbes 800 contains the companies ranked in the top 500 along one of the following characteristics: revenues, total assets, net income and market capitalization. Around 800 companies per year fit those criteria. Execucomp contains around 1500 companies per year. We use the full universe of firm-year observations available, subject to the following restrictions: We drop observations with negative net income, to be able to use the log of net income. In the Execucomp data, the top 5 executives from each company are reported and there is an "annual ceo" variable (variable "ceoann") which flags the CEO in a given year. However, not all companies report a CEO in a given year. We hence drop company-year observations which do not report a CEO (around 1/8th of observations). We merge these two data series to the COMPUSTAT-CRSP database using the 6-digit firm CUSIP code. The match to the Forbes is less than perfect and we were unable to get a better match: per year we merge on average 550 Forbes firms to Compustat.

For both series (Forbes and Execucomp) we use the available "total compensation" measure, which captures all salaries, bonuses, restricted stock grants, long term interest participation payouts, the value of option grants and all other payments made (variable "totalpay" in Forbes and "tdc2" in Execucomp - Note that we also performed all the analyses with the "tdc1" variable instead, which values option grants in a different way, but the results were almost unaffected and the "tdc1" measure seems less consistent with the Forbes measure of total pay).

Demographic CEO variables such as age, tenure and tenure as CEO are already constructed in the Forbes 800 database. For Execucomp, we use the reported date at which the executive became CEO ("becameceo") to infer tenure as CEO and the date at which he joined the company ("joined_co") to infer the tenure in the company. Unfortunately, while tenure is a crucial variable for determining CEO pay, it is missing for a lot of the CEOs in the Execucomp sample. Since we do not want to omit tenure from the analysis, we are left with around 23,000 observations which have non-negative net income, non-missing CEO controls and a reported CEO, for our analysis. In final, we have around 550 firms per year pre 1992 and on average 700 firms per year after 1992.

Luck performance. To obtain a measure of luck, we compute the average, asset-weighted industry performance (for either the log of net income or the market return) across 2-digit SIC industries, for each year, and excluding the firm under consideration. This is then used as an instrument for firm performance in a given year.

Wages of workers. Average wages at the 2-digit SIC Industry level since 1970 come from the Bureau of Economic Analysis' National Accounts, more specifically from Table 6.6B. Wage and Salary Accruals Per Full-Time Equivalent Employee by Industry available on their website. For the wage regressions, we simply regress average industry wages on the asset-weighted average industry performance for both the log of net income and the market return. The wage regressions also contain industry fixed effects.

All nominal variables are deflated using the US CPI. All regressions also contain time fixed effects.

A.2.4 International US CEO pay data

We use the novel dataset constructed by Fernandes, Ferreira, Matos, and Murphy (2012), for CEO pay from 14 countries in 2006. Detailed information can be found in these authors' paper, but is reproduced more or less verbatim here for convenience. The data contains information on CEO pay and characteristics from the BoardEx and Execucomp databases, information on stock ownership (from LionShares), firm performance (from Worldscope and Datastream), and firm governance. We use several of their variables: The log of firm sales (in thousands of US$ (Worldscope item 01001)), the stock return (Datastream item RI) and its volatility (the annualized standard deviation of daily stock returns), leverage (Total debt divided by total assets (Worldscope item 03255 / item 02999)), and Tobin's q (Total assets (Worldscope item 02999) plus market value of equity (item 08001) minus book value of equity (item 03501) divided by total assets)

In terms of governance measures we use the following five variables. (1) A dummy equal to 1 if the percentage of closely owned shares (that is, owned by shareholders who hold at least 5% of the outstanding shares such as officers and directors and immediate families, other corporations, or individuals) as a proportion of the number of shares outstanding (Worldscope item 08021) is greater than the median in the sample; (2) a dummy equal to 1 if the percentage of institutionally owned shares (from Lionshare) is greater than the median in the sample; (3) a dummy equal to 1 if the CEO is also the chairman of the board; (4) a dummy equal to 1 if the average number of positions that boards members hold at other companies' boards is less than the median in the sample; (5) the fraction of independent board directors. We construct an index by first turning each variable into a 'z-score', by subtracting its mean and dividing by its standard deviation. We then directly add them to form the governance index and normalize the index to have mean zero and standard deviation equal to 1. In accordance with the arguments and analysis in Fernandes et al. (2012), we have coded all variables such that a higher index represents better governance.

We are well aware that all measures of governance are bound to be imperfect, somewhat imprecise and prone to different interpretations. Many papers consider some sort of index, or instead a single measure of governance. We have also explored different index constructions, such as based directly on the percentage of closely owned shares or institutionally owned shares (instead of our dummy variables). We have also tried using each of the governance measures independently and directly. Broadly speaking, the results are qualitatively consistent with the ones reported, although not always significant and the magnitudes differ.

Additional Appendix References

Atkinson, Anthony, and Andrew Leigh. 2010. "The Distribution of Top Incomes in Five Anglo-Saxon Countries over the Twentieth Century," IZA Discussion Paper No. 4937

Atkinson, Anthony, and Thomas Piketty. 2007. *Top Incomes over the Twentieth Century.* Oxford: Oxford University Press.

Atkinson, Anthony, and Thomas Piketty. 2010. *Top Incomes: A Global Perspective.* Oxford: Oxford University Press.

International Labor Office Labor Statistics Database. http://laborsta.ilo.org/

Kleven, Henrik, Camille Landais, and Emmanuel Saez. 2013. "Taxation and International Mobility of Superstars: Evidence from the European Football Market," forthcoming *American Economic Review.*

Japan National Tax Administation Data.
http://www.soumu.go.jp/main_sosiki/jichi_zeisei/czaisei/czaisei_seido/ichiran06_h17.html

Mitchell, B.R. (1998). *International Historical Statistics*, (Macmillan: London), forth edition.

OECD Statistics. http://stats.oecd.org

OECD. (annual). *Taxing wages*, OECD, Paris.

OECD. (1986). *Personal income tax systems*, OECD, Paris.

Piketty, Thomas. 1997. "La Redistribution Fiscale face au Chômage," *Revue Française d'Economie*, 12, 157-201.

Rothschild, Casey, and Florian Scheuer. 2013. "Redistributive Taxation in the Roy Model," forthcoming, *Quarterly Journal of Economics.*

Saez, Emmanuel. 2002. "Optimal Income Transfer Programs: Intensive Versus Extensive Labour Supply Responses." *Quarterly Journal of Economics*, 117(2): 1039-73.

Saez, Emmanuel. 2004b. "Direct or Indirect Tax Instruments for Redistribution: Short-Run versus Long-Run," *Journal of Public Economics* 88(3-4), 503–518.

Saez, Emmanuel, and Michael Veall. 2007. "The Evolution of High Incomes in Canada, 1920-2000", Chapter 6 in A. B. Atkinson and T. Piketty, eds. *Top Incomes over the Twentieth Century: A Contrast Between European and English Speaking Countries*, Oxford: Oxford University Press, 226-308.

Stiglitz, Joseph. 1982. "Self-selection and Pareto Efficient Taxation," *Journal of Public Economics* 17, 213-240.

Straathof, Bas, Stefan Groot en Jan Mohlmann. 2010. "Hoge bomen in de polder: Globalisering en topbeloningen in Nederland." CPB Document, No. 199 (in Dutch).

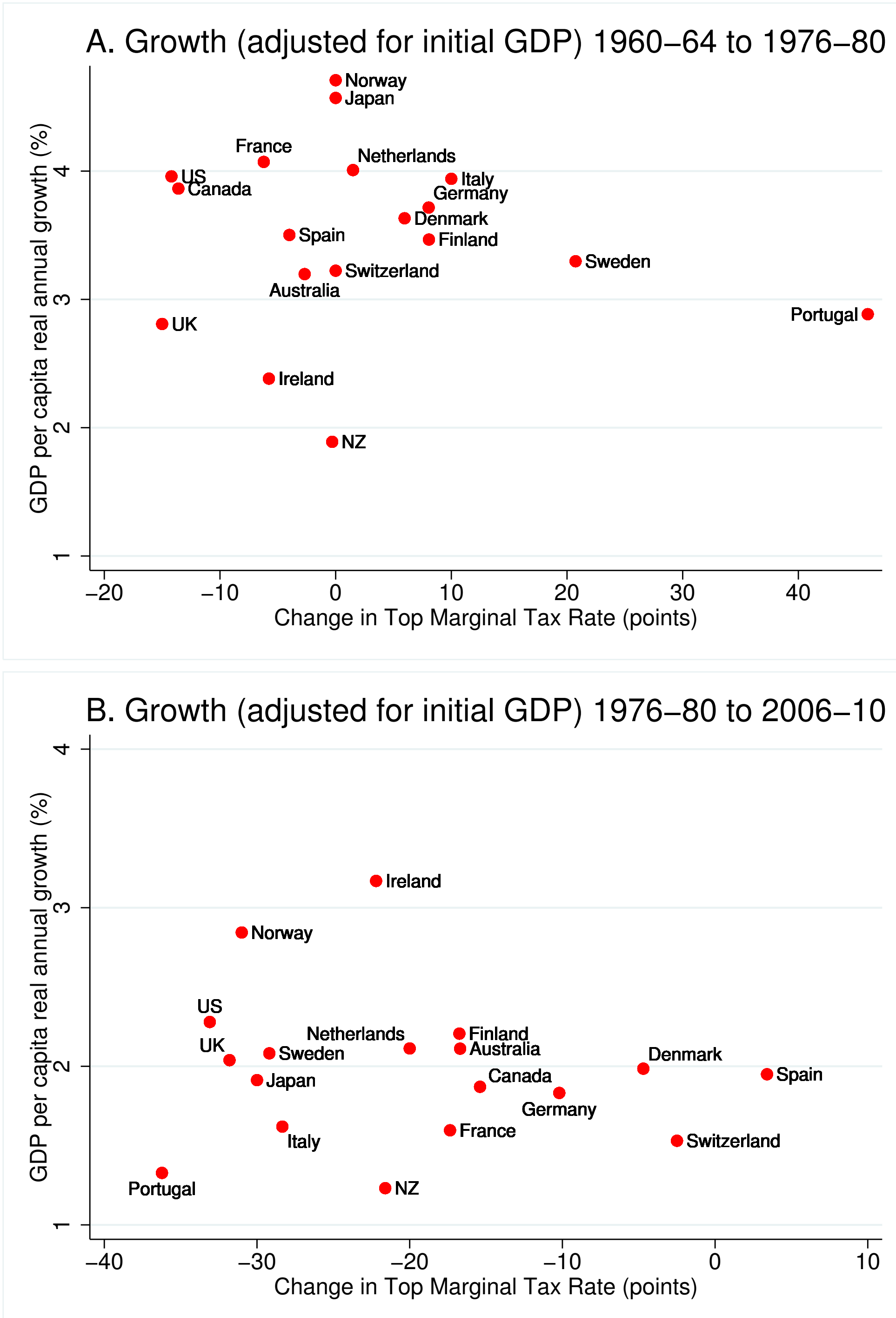


Figure A1: Top Marginal Tax Rates and Growth: 1960-4 to 1976-80 and 1976-80 to 2006-10

The figure depicts the average real GDP per capita annual growth rate (adjusted for initial GDP as in Figure 5, Panel B) against the change in top marginal tax rate for two sub-periods: 1960-4 to 1976-80 in panel A and 1976-80 to 2006-10 in panel B. In both sub-periods, there is no correlation between the change in top marginal tax rate and the average growth over the period. Panel B captures the period starting with the Thatcher and Reagan revolutions. While the US and the UK did cut top tax rates more and grew faster than France and Germany, this does not generalize to the 18 OECD countries. Some countries (such as Portugal) cut top tax rates sharply and did not grow fast. Other countries (such as Finland or Denmark) did not cut top tax rates much and yet grew as fast as the US or UK.