Web Appendix of

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

By **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 percent (and then reduced to 44 percent) 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 percent income shares comes from the *World Top Incomes Database* (Alvaredo et al. 2011). We use the top 1 percent 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 percent 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), fourth 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).

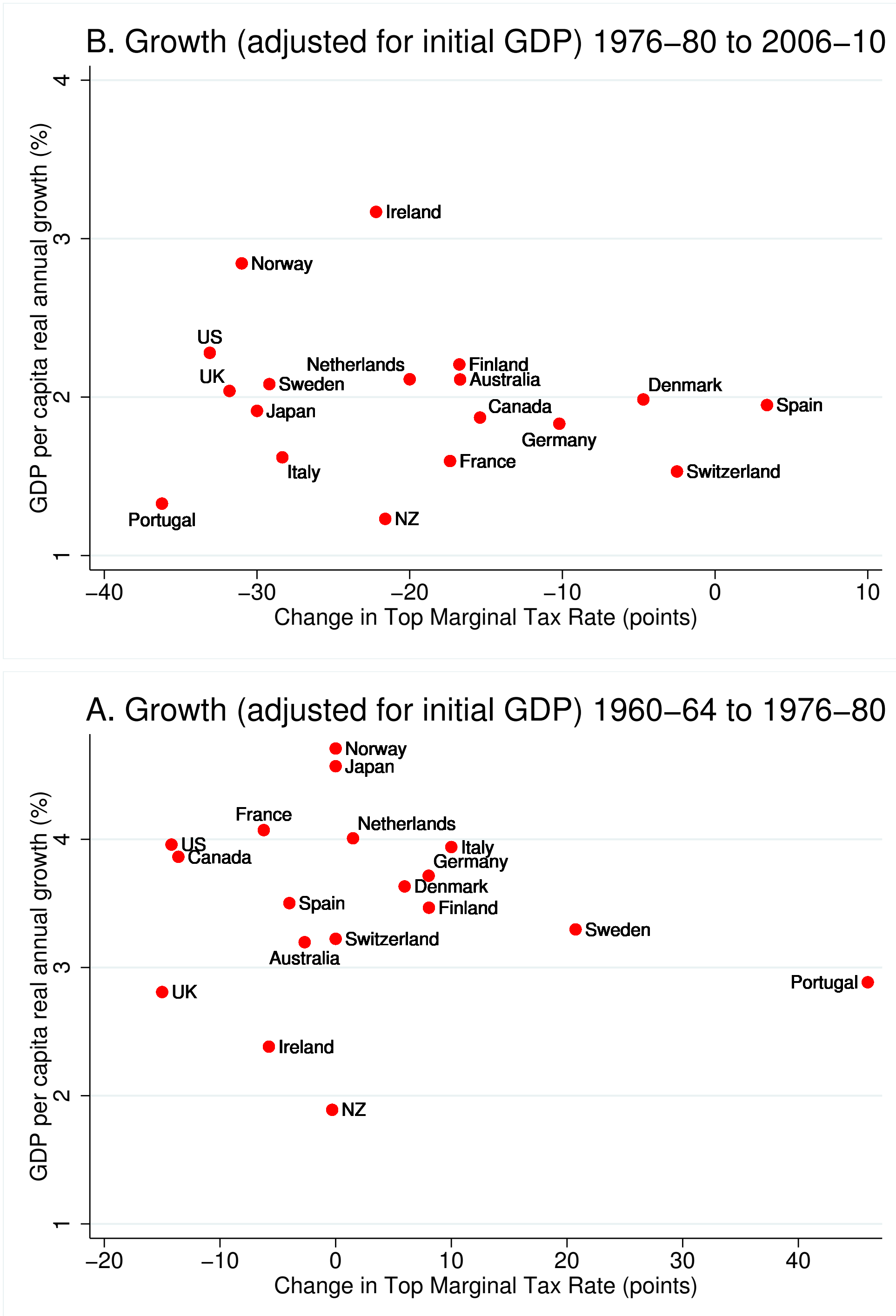

Figure A1: Top Marginal Tax Rates and Growth: 1960-4 to 1976-80 and 1976-80 to 2006-10

Notes: 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.