Using the elasticity of the Pareto parameter to derive top income optimal tax rates*

Shih-Ying Wu

Department of Economics, National Tsing Hua University

C.C. Yang†

Institute of Economics, Academia Sinica

Department of Public Finance, National Chengchi University

Department of Public Finance, Feng Chia University

Shenghao Zhu

Department of Economics, National University of Singapore

March 2016

Abstract

The Pareto parameter is a measure of the thickness of the income distribution at the top. Using the elasticity of the Pareto parameter with respect to top tax rates, this paper derives a formula for top income optimal tax rates. We empirically

*We thank Fabian Kindermann, Thomas Sargent, Pengfei Wang, Ye Yuan, and seminar participants of the 2015 SAET conference held in Cambridge. Shenghao Zhu acknowledges the research grant support from National University of Singapore (R-122-000-175-112).

†Corresponding author, C.C. Yang. Email: ccyang@econ.sinica.edu.tw

implement the formula by the narrative method of Romer and Romer (2010) and Mertens (2015). Our formula refines Saez's (2001), in that the elasticity of the Pareto parameter is a *sufficient statistic* for summarizing changes in the top tax base, whereas the well-known elasticity of taxable income is not.

Keywords: The elasticity of the Pareto parameter, Optimal top tax rates, Sufficient statistics

JEL Classification: D31, H21

1 Introduction

Mirrlees (1971) pioneered the modern study of optimal income taxation in a framework where income differences between agents are attributed to unobservable innate ability, and optimal taxes seek a trade-off between redistributing income for equity and dulling agents' incentives to work. Subsequent works have elaborated on Mirrlees' original idea in a variety of directions. Of these elaborations, the "zero top" result of Sadka (1976) and Seade (1977) is perhaps the most salient. In stark contrast to the real world in which progressivity in the marginal tax rate is typical, they showed that the marginal tax rate at the top should be zero as long as the distribution of income is bounded from above.

Saez (2001) forcefully argued that the "zero top" result is of little practical interest. He derived a formula for top income optimal tax rates in the Mirrlees framework. If a society is simply to maximize tax revenue rather than social welfare, the formula is given by

$$\tau^* = \frac{1}{1 + e \cdot \alpha}, \tag{1}$$

where τ^* is the revenue-maximizing top tax rate, e is the aggregate elasticity of taxable income in the top tax bracket with respect to the net-of-top-tax rate $(1-\tau)$, and α is the Pareto parameter of the income distribution at the top. Saez (2001) numerically implemented his derived formula, finding that τ^* are fairly high in general. This is in sharp contrast to the "zero top" result of Sadka (1976) and Seade (1977).[1] Recently formula (1) has been the focal point of studies on revenue-maximizing top tax rates, including Guner et al. (2014), Kindermann and Krueger (2014), and Badel and Huggett (2015a, 2015b).

Saez (2001) emphasized that the aggregate elasticity e in (1) is microfounded, built on some aggregation of individual elasticities of taxable income. His approach tracks how the taxable incomes of those in the top income group have varied in response to tax changes, but it fails to account for top earners' *extensive mobility* into and out of the top income group caused by tax changes. Saez (2001, Footnote 7) particularly assumed that top earners' extensive mobility is negligible.[2] However, using the 1986 Tax Reform Act (TRA 1986) in the United States as an illustration, we empirically show, in this paper, that the impact of a tax change on extensive mobility of top earners can be statistically significant.

Applying formula (1), researchers also face the difficulty of choosing the value of the Pareto parameter α. Saez (2001) claimed that the formula "can be applied using directly the empirical value of α" (p. 212).[3] However, from an empirical point of view, there are

[1]Saez (2001) also addressed welfare-maximizing top tax rates. The numerical results are qualitatively similar to revenue-maximizing top tax rates.

[2]See also Saez et al. (2012, Footnote 7), who made the same assumption.

[3]Saez (2001, p. 222) recognized the possibility of the dependence of α upon τ, but theoretically showed that in the Mirrlees model the Pareto parameter of the income distribution is equal to the Pareto parameter of the skill distribution divided by $1+\bar{\zeta}^u$, where $\bar{\zeta}^u$ is a weighted average of the uncompensated elasticity of the top income group above $\bar{z}$. Under the assumption that $\bar{\zeta}^u$ converges as $\bar{z}$ increases, one can show that the Pareto parameter of the income distribution is independent of the top income optimal tax rate as long as the top tax rate is less than one. This assumption is satisfied for the two utility functions used in numerical simulations in Saez (2001),

$$u = \log\left(c - \frac{l^{1+k}}{1+k}\right),$$

many values of α that can be chosen in data and the resulting value of τ^* in (1) is obviously sensitive to the choice. Figure 1 shows that the values for the Pareto parameter of top income groups vary widely during 1946-2012 in the United States, from far above 2 in the 1970s to close to 1.5 in the most recent years.[4] In fact, in his empirical implementation, Saez (2001) considered three values for α (1.5, 2, and 2.5) and found that the value of α exerts a big impact on top income optimal tax rates.

(Insert Figure 1 about here)

To address these two problems, we derive a new formula in the Mirrlees (1971) framework

$$\tau^* = \frac{1}{1+\zeta}, \tag{2}$$

where ζ is called "the elasticity of the Pareto parameter," defined as the elasticity of the Pareto parameter with respect to the net-of-top-tax rate $(1-\tau)$. We show that the elasticity ζ fully summarizes changes in the top tax base so that it is a *sufficient statistic* for the revenue-maximizing top tax rate τ^*. The new formula (2) resolves the two problems of formula (1) discussed above, in that: (i) the elasticity ζ incorporates the *extensive* as well as the *intensive* income mobility in the top income group, and (ii) formula (2) is immune from arbitrarily choosing values of the Pareto parameter α.

We empirically estimate the elasticity of the Pareto parameter for the top 1% income group based on the exogenous tax changes identified by the narrative method of Romer

and

$$u = \log(c) - \log\left(1 + \frac{l^{1+k}}{1+k}\right),$$

where c is consumption, l is labor supply, and k is a constant parameter. Nevertheless, the empirical results in Section 3.2 of our paper show that the Pareto parameter of the income distribution does depend on the top income tax rate.

[4] To estimate Pareto parameters, we use the extended data file attached to Piketty and Saez (2003) at http://eml.berkeley.edu/~saez/. Income here is the same as the so-called "broad income" defined in Gruber and Saez (2002), which includes wages, salaries and tips, interest income, dividends, and business income; capital gains are excluded because of their special tax treatment.

and Romer (2010) and Mertens (2015). We then numerically implement our derived formula (2) for top income optimal tax rates and compare our results with those of Saez (2001). Top income optimal tax rates resulting from our derived formula are still high, but they are lower than those in Saez (2001) to some extent.

Related literature

Besides Saez (2001), papers including Diamond and Saez (2011), Piketty and Saez (2013), Piketty et al. (2014), Badel and Huggett (2015b), and Scheuer and Werning (2015) are mostly related to our paper. Diamond and Saez (2011, Recommendation 1) and Piketty and Saez (2013, Section 5.1.1) mainly summarized the results of Saez (2001). Diamond and Saez (2011) also explored the path from tax theory to policy recommendations. Piketty et al. (2014) extended Saez's (2001) derived formula for top income optimal tax rates by taking into consideration top earners' activities other than their labor supply, such as tax avoidance and compensation bargaining. Scheuer and Werning (2015) extended the Mirrlees model to generating superstar effects and addressed the question of how top income optimal tax rates are affected by the presence of superstar phenomena. Badel and Huggett (2015b) extended Saez's (2001) derived formula to account for the impact of varying the top tax rate on the tax base other than the top one. We make a brief comparison between our approach and theirs in Section 2.1.5.

Saez et al. (2012) critically reviewed the literature on the elasticity of taxable income (ETI). The focus of the literature is on the estimation of ETI in general and e in formula (1) in particular. We argue that using panel data to estimate e, as is often done in practice, is problematic since it leaves out the extensive mobility of the tax base in the top tax bracket. Furthermore, we show that the elasticity of the Pareto parameter ζ in our derived formula (2) is a sufficient statistic for summarizing changes in the top tax base, whereas e in formula (1) is not.

Our paper is also related to a class of simulation works. Guner et al. (2014) considered a life cycle model to investigate how the progressivity of income taxes influences government revenues, finding that a tilt of the income-tax scheme toward high earners leads to endogenous responses of agents in the long run and small effects on the overall revenue collected. Badel and Huggett (2015a) considered the role of human capital in taxing top earners, arguing that the top tax rate should be lower if compared with Saez's (2001) static setting. Kindermann and Krueger (2014) allowed for the possibility of an extremely high skill shock in a life cycle model, arguing that the top tax rate should be higher if compared with Saez's (2001) static setting. These papers mainly find the optimal tax scheme by numerical computations, while our paper derives a new formula for top income optimal tax rates.

The rest of the paper is organized as follows. Section 2 revisits the derivation of top income optimal tax rates, showing the "sufficient statistic" feature of the elasticity of the Pareto parameter. Section 3 estimates the elasticity of the Pareto parameter and then numerically implements our derived formula (2). Section 4 concludes.

2 Derivation of top income optimal tax rates revisited

We follow the sufficient statistic approach as in Saez (2001) and Piketty and Saez (2013). Saez (2001) considered a stylized model *à la* Mirrlees (1971), in which $u(c, z)$ is a well-behaved individual utility function, depending positively on consumption c and negatively on earnings z. An agent in the top tax bracket faces the budget constraint $c = (1-\tau)z+R$, where R is virtual income and τ is the top tax rate imposed. Utility maximization leads to an earnings function $z = z(1-\tau, R)$. Following Feldstein (1999) and the ETI literature, z can be extended to represent taxable income, recognizing that hours of work are only one component of the individual behavioral response to income taxation; other components include intensity of work, form and timing of compensation, and tax avoidance/evasion.[5]

In deriving the top income optimal tax rate, Saez (2001) broke effects from a small perturbation of the top tax rate into two components: one regarding changes in the tax revenue (the revenue effect), and the other regarding changes in the welfare of the agents in the top tax bracket (the welfare effect). The optimal top tax rate can be derived by exploiting the fact that, after evaluating the welfare effect in terms of the value of public funds, the resulting change in these two components must sum to zero at the optimum, that is, there will be no first-order effect from the perturbation at the optimum. We follow the same methodology but make some important refinements.

2.1 Revenue effect

Suppose that T is the top tax base associated with the top tax rate τ. Obviously, the tax revenue collected from the top tax bracket is given by $\tau \cdot T$. Changes in the tax revenue collected can then be divided into two pieces: (i) those stemming from changes in the *tax rate* τ given that the tax base T is unchanged, and (ii) those stemming from changes in the *tax base* T evaluated at the pre-change tax rate τ. Saez (2001) called the first the "mechanical effect," in that this effect represents the projected change in tax revenue

[5] In empirical studies (see, for example, Gruber and Saez, 2002), income definitions include broad income and taxable income. For the definition of broad income, see Footnote 4. Taxable income is defined as broad income minus the adjustments that are made to arrive at taxable income, which is close to the taxable income defined in the income tax law. For a critical review of the literature on ETI, see Saez et al. (2012).

if there is no change in agents' taxable income; while he referred to the second as the "behavioral response," in that it accounts for agents' response to tax changes evaluated at the pre-change tax rate. The mechanical effect in our analysis is exactly the same as the one in Saez (2001), but the behavioral response will be replaced by the so-called "mobile response," which generalizes Saez's behavioral response. We also compare our approach with that of using the so-called "aggregate elasticity of taxable income" as defined in Saez et al. (2012).

2.1.1 Mechanical effect

Let z denote taxable income and $f(z)$ the corresponding density function. As emphasized by Saez (2001), $f(z)$ itself is not exogenously given but endogenous in response to changes in tax codes.

Consider an increase in the top income tax rate, denoted $d\tau$. The mechanical effect M represents an increase in tax revenue if there is no change in the tax base,

$$M \equiv (z_m - \bar{z})\, d\tau,$$

where $(z_m - \bar{z})$ is the tax base in the top tax bracket, and z_m denotes the mean of taxable incomes above the threshold $\bar{z}$ of the top tax bracket. By definition we have

$$z_m = \frac{\int_{\bar{z}}^{\infty} z f(z) dz}{\int_{\bar{z}}^{\infty} f(z) dz} = \int_{\bar{z}}^{\infty} z h(z) dz, \tag{3}$$

where $h(z) = f(z) / \int_{\bar{z}}^{\infty} f(z) dz$ with $z \geq \bar{z}$ and the denominator $\int_{\bar{z}}^{\infty} f(z) dz$ of $h(z)$ serves to normalize the population with taxable income above $\bar{z}$ to one. Saez (2001) made the same normalization. The mechanical effect M here is identical to that in Saez (2001, Eq.

(5)).

2.1.2 Mobile vs. behavioral response

When there is an increase $d\tau$, the resulting change in tax revenue at the top evaluated at the pre-change tax rate τ is given by

$$D \equiv \tau dz_m,$$

which indicates that this change in tax revenue is completely attributed to the change in the top tax base, i.e., dz_m.

Saez (2001) measured dz_m by a weighted average of individuals' elasticities of taxable income, where $h(z)$ in (3) serves as the weight; see his Eq. (7). He called this change the "behavioral response," which calculates changes in z, given $h(z)$. We measure dz_m by changes in $h(z)$ itself given $z \geq \bar{z}$ and call it the "mobile response." We illustrate possible differences between these two measures by the following examples.

Illustrative examples There are four agents in an economy, with $z_1 = 50$, $z_2 = z_3 = 100$, $z_4 = 300$ when the top tax rate is τ and the threshold $\bar{z} = 80$. Thus, we have[6]

$$h(z;\tau) = \begin{cases} z_2 = z_3 = 100, \\ \quad z_4 = 300. \end{cases}$$

Note that $z_1 = 50$ is not part of $h(z)$ since $z_1 < \bar{z} = 80$, that is, z_1 is not in the rich club. The top tax base T at τ is given by $2(100 - \bar{z}) + (300 - \bar{z}) = 2(100 - 80) + (300 - 80)$.

Example 1.

Suppose that the government lowers the top tax rate from τ to τ', and that the resulting outcome is: z_2 increases from 100 to 300 and all else remains unchanged.

[6] For ease of exposition, we do not express $h(z)$ in terms of frequencies. Our examples are illustrative. We do not assume that the incomes of the rich group follow a Pareto distribution in our examples.

Behavioral response: keep track of how the taxable incomes of those in the rich club at the pre-change τ have changed, that is, the change in z given $h(z;\tau)$.

Since z_2 is in the rich club at τ and his taxable income has increased from 100 to 300 as the top tax rate is lowered to τ' (all else remains unchanged), $dz_m = z_m(\tau') - z_m(\tau) = 200$.[7]

Mobile response: keep track of how $h(z)$ itself has changed, that is, the change in $h(z;\tau)$ itself given $z \geq \bar{z}$.

Note that

$$h(z;\tau') = \begin{cases} z_3 = 100, \\ z_2 = z_4 = 300. \end{cases}$$

Comparing $h(z;\tau')$ with $h(z;\tau)$ yields $dz_m = 200$, since z_2 has moved upward in income from 100 to 300.

The top tax base T at τ' is equal to the same amount $(100 - 80) + 2(300 - 80)$, regardless of the behavioral or the mobile response.

When the members of the rich club remain unchanged as the top tax rate varies, the two approaches give the same answer with respect to dz_m as the above example demonstrates. However, the answers may be different when the members of the rich club change in response to variations in the top tax rate. We illustrate this possibility by the following two examples.

Example 2.

Suppose that the government raises the top tax rate from τ to τ'', and that the resulting outcome is: z_2 decreases from 100 to 70 and all else remains unchanged.

Behavioral response: keep track of how the taxable incomes of those in the rich club at the pre-change τ have changed, that is, the change in z given $h(z;\tau)$.

Since z_2 is in the rich club at τ and his taxable income has decreased from 100 to 70 as

[7] In our examples we abuse the notation z_m a little and use it to represent the sum of incomes instead of the mean of incomes above $\bar{z}$.

the top tax rate is raised to τ'' (all else remains unchanged), $dz_m = z_m(\tau'') - z_m(\tau) = -30$.

Mobile response: keep track of how $h(z)$ itself has changed, that is, the change in $h(z;\tau)$ itself given $z \geq \bar{z}$.

Note that

$$h(z;\tau'') = \begin{cases} z_3 = 100, \\ \\ z_4 = 300. \end{cases}$$

Comparing $h(z;\tau'')$ with $h(z;\tau)$ yields $dz_m = -100$, since z_2 has moved downward in income from 100 to 70 and so agent z_2 is no longer a member of the rich club at τ''.

According to the mobile response, the top tax base at τ'' is equal to $(100-80)+(300-80)$, which implies the loss of the top tax base $100-80$ if compared with that at τ. According to the behavioral response, however, the top tax base T at τ'' is not well defined since z_2 at τ'' is below the threshold $\bar{z} = 80$.

As far as measuring dz_m is concerned, the mobile response gives the correct answer in this example while the behavioral response does not.

Example 3.

Suppose that the government cuts the top tax rate from τ to τ''', and that the resulting outcome is: z_1 increases from 50 to 100 and all else remains unchanged. This scenario is possible because it provides incentives for agent z_1 to work harder to become a member of the rich club as the top tax rate is lower.

Behavioral response: keep track of how the taxable incomes of those in the rich club at the pre-change τ have changed, that is, the change in z given $h(z;\tau)$.

Since the taxable incomes for the members of the rich club at τ remain the same as those at τ''', $dz_m = z_m(\tau''') - z_m(\tau) = 0$.

Mobile response: keep track of how $h(z)$ itself has changed, that is, the change in $h(z;\tau)$ itself given $z \geq \bar{z}$.

Note that

$$h(z;\tau''') = \begin{cases} z_1 = z_2 = z_3 = 100, \\ \\ z_4 = 300. \end{cases}$$

Comparing $h(z;\tau''')$ with $h(z;\tau)$ yields $dz_m = 100 - 80$. Agent z_1 is not a member of the rich club at τ, but has become a member of the rich club at τ'''. As a result, the gain in the top tax base from τ to τ''' is $100 - 80$ rather than none as measured by the behavioral response. Again, the mobile response gives the correct answer in this example while the behavioral response does not.

Agents respond differently to changes in the top tax rate τ and so alter their taxable incomes accordingly. These changes in taxable incomes may make some agents move above or below the threshold $\bar{z}$. The main lesson from the above examples is: if there is little mobility into and out of the rich club (Example 1), the two approaches are equivalent; however, once there is a non-negligible fraction of agents who experience mobility into and out of the rich club (Examples 2 and 3), it is better to measure dz_m by changes in $h(z)$ itself given $z \geq \bar{z}$ rather than by changes in z given $h(z)$. At any rate, it is clear that our mobile response generalizes Saez's behavioral response by accounting for the "extensive" as well as the "intensive" income mobility in the rich club.

In deriving his Eq. (7) for measuring the behavioral response, Saez (2001) implicitly assumed that the set of agents who might jump discontinuously because of a small tax reform is negligible. However, Saez (2001, Footnote 7) admitted that this assumption may fail to hold.[8] Thus, it seems that Saez (2001) has recognized that his derived formula is not applicable once there is a high extensive mobility into or out of the rich club in response to tax changes. We provide evidences on the significance of extensive mobility in the next section.

[8] See also Saez et al. (2012, footnote 7), who made the same assumption.

Remark 1 *To empirically estimate the behavioral response as suggested by Saez (2001), panel data are required since the response tracks how the taxable incomes of those in the rich club at a point of time have changed. Following the influential work of Feldstein (1995), a great majority of empirical studies did employ panel data to estimate the behavioral response. However, our analysis suggests that, as far as estimating changes in the top tax base is concerned, panel data may not be so useful once there exists significant extensive mobility in the rich club over time. We propose an alternative method of utilizing data below.*

2.1.3 The elasticity of the Pareto parameter vs. the aggregate elasticity of taxable income

Let $T = (z_m - \bar{z})$, which is the top income tax base associated with the top tax rate τ. There are two possible approaches to measuring the mobile response of the top tax base. One involves the "aggregate elasticity of taxable income" as defined in Saez et al. (2012), while the other is through our defined "elasticity of the Pareto parameter." We explain their meanings and differences below.

The aggregate elasticity of taxable income in the top bracket with respect to the net-of-top-tax rate $(1 - \tau)$ is defined as

$$e \equiv \frac{1-\tau}{z_m} \cdot \frac{\partial z_m}{\partial (1-\tau)}. \tag{4}$$

Saez et al. (2012, p. 7) commented: "This aggregate elasticity is equal to the average of the individual elasticities weighted by individual income, so that individuals contribute to the aggregate elasticity in proportion to their incomes." Piketty and Saez (2013, Footnote

54) explicitly noted that[9]

$$e = e^u - \eta \cdot \frac{\bar{z}}{z_m}, \tag{5}$$

where $e^u \equiv \int_{\bar{z}}^{\infty} e_{(z)} z h(z) dz / z_m$ is a weighted average of the uncompensated individual elasticity $e_{(z)}$;[10] $\eta \equiv \int_{\bar{z}}^{\infty} \eta_{(z)} h(z) dz$ is the average of the individual income effect $\eta_{(z)}$;[11] and $\bar{z}/z_m = (\alpha - 1)/\alpha$ since $z \geq \bar{z}$ is Pareto-distributed. The decomposition of e in (5) was derived by Saez (2001) and this decomposition was used in his Eq. (8), which is the formula for top income optimal tax rates in Saez (2001). Saez (2001) emphasized that the aggregates e^u and η in (5) are microfounded, that is, they are built on individual $e_{(z)}$ and $\eta_{(z)}$.

As explained earlier, if we start from the right-hand side of (5) by employing panel data of individuals to first estimate $e_{(z)}$ and $\eta_{(z)}$ and then obtain the aggregate elasticity e via (5), the approach will miss extensive mobility. This is true because both e^u and η as defined above are evaluated with a given $h(z)$, and so they fail to take into account people's extensive mobility into and out of the top tax bracket.

How about estimating the aggregate elasticity e by the share analysis of Feenberg and

[9] Piketty and Saez (2013) specified it as

$$e = e^u + \eta \cdot \frac{\bar{z}}{z_m},$$

where the positive rather than the negative sign seems to be a typo.

[10] More precisely, the elasticity term $e_{(z)}$ is the average elasticity over individuals with income z. Define the uncompensated elasticity for an individual with income z as

$$e(z) \equiv \frac{1-\tau}{z} \cdot \frac{\partial z}{\partial(1-\tau)}.$$

See equation (1) of Saez (2001). $e_{(z)}$ is the average of $e(z)$ since Saez (2001) allowed for heterogeneity in preferences.

[11] Define the income effect for an individual with income z as

$$\eta(z) \equiv (1-\tau) \frac{\partial z}{\partial R}.$$

See equation (2) of Saez (2001). $\eta_{(z)}$ is the average of $\eta(z)$ since Saez (2001) allowed for heterogeneity in preferences.

Poterba (1993) and Slemrod (1996)[12] or by employing repeated cross-section analysis rather than panel analysis? Both share analysis and repeated cross-section analysis take into consideration people's extensive mobility in their estimation of e. However they suffer from a problem in measuring the mobile response of the top tax base, since the tax base of the top bracket is $(z_m - \bar{z})$ rather than z_m itself.

The aggregate elasticity e as defined by Eq. (4) reflects the response of z_m rather than $(z_m - \bar{z})$ with respect to top tax rates. This simple fact implies that when using the aggregate elasticity e to measure the response of the top tax base, some adjustment to e is necessary, even if the estimate of e incorporates extensive mobility.

Let us define

$$\zeta \equiv \frac{1-\tau}{z_m - \bar{z}} \cdot \frac{\partial(z_m - \bar{z})}{\partial(1-\tau)}, \tag{6}$$

which is the elasticity of the top tax base with respect to the net-of-top-tax rate $(1-\tau)$. From (4) and (6), we obtain the following relationship between ζ and e,

$$\zeta = e \cdot \frac{z_m}{z_m - \bar{z}}. \tag{7}$$

This relationship is intuitive. The elasticity ζ defined by (6) dictates that a 1% change in $(1-\tau)$ causes a ζ% change in the *top tax base* $(z_m - \bar{z})$, while the elasticity e defined by (4) dictates that a 1% change in $(1-\tau)$ causes a e% change in the *top income* z_m. To equate e with ζ, the term $\frac{z_m}{z_m - \bar{z}}$ in equality (7) plays a role of the necessary adjustment. This is also the key difference between the elasticity ζ and the elasticity e.

[12] Saez et al. (2012, p. 19) noted: "a natural measure of the evolution of top incomes relative to the average is the change in the share of total income reported by the top percentile." This is the basic idea of share analysis.

Since $z \geq \bar{z}$ is Pareto-distributed, using $z_m/\bar{z} = \alpha/(\alpha - 1)$ we then derive from (7)

$$\zeta = e \cdot \alpha. \tag{8}$$

That is, when using the aggregate elasticity e to measure the response of the top tax base, the term e needs to be adjusted by multiplying with the Pareto parameter α. Saez (2001) claimed that the formula for top income optimal tax rates "can be applied using directly the empirical value of α" (p. 212). If e is estimated through one tax reform, the choice for α in Eq. (8) is clear. In this case α should take the pre-reform value. However, if the elasticity e is identified through pooling several or even many tax reforms (in, for example, the repeated cross-section analysis of Saez et al. (2012)), then a difficult problem with Saez's claim in empirically implementing the right-hand side of (8) is that there are too many empirical values of α to choose from as Figure 1 has demonstrated.

Alternatively we propose empirically implementing the left-hand side of (8) directly to measure the response of the top tax base. Note that

$$\zeta = \frac{1-\tau}{z_m - \bar{z}} \cdot \frac{\partial(z_m - \bar{z})}{\partial(1-\tau)} = \frac{1-\tau}{z_m/\bar{z} - 1} \cdot \frac{\partial(z_m/\bar{z} - 1)}{\partial(1-\tau)} = \frac{1-\tau}{\frac{1}{\alpha-1}} \cdot \frac{\partial(\frac{1}{\alpha-1})}{\partial(1-\tau)},$$

where the last equality has utilized the property $z_m/\bar{z} = \alpha/(\alpha - 1)$ because $z \geq \bar{z}$ is Pareto-distributed. Thus, we obtain

$$\zeta = \frac{1-\tau}{\frac{1}{\alpha-1}} \cdot \frac{\partial(\frac{1}{\alpha-1})}{\partial(1-\tau)}, \tag{9}$$

which is our defined elasticity of the Pareto parameter. By (6) ζ incorporates extensive mobility.

The endogenous dependence of α upon τ as embodied in (9) is consistent with the evidence provided by Alvaredo et al. (2013) and Mertens (2015). The former paper plots the changes in top marginal income tax rates since the early 1960s against the changes over that period in the top 1% income shares for 18 countries in the World Top Incomes Database (Alvaredo et al., 2011). It is found that there is a strong negative correlation between the changes in top tax rates and the evolution of top income shares. The latter paper applies a structural vector autoregression approach to the U.S. tax return data over the 1946-2012 period and reaches a consistent finding that a hypothetical cut in the top tax rate for the top 1% leads to sizeable increases in top 1% incomes and greater inequality in pre-tax incomes.

In Section 3 of empirical applications we use Eq. (9) to estimate ζ.

2.1.4 Revenue-maximizing top tax rates

If the goal of the government is simply to maximize the amount of tax revenue collected from the top tax base, then the revenue-maximizing top tax rate τ^* is implicitly defined by adding up the mechanical effect and the mobile response and setting the sum to zero,

$$M + D = 0.$$

From (6), we have $dz_m = -\zeta \frac{z_m - \bar{z}}{1-\tau} d\tau$. Substituting it into D, the above first-order condition then gives rise to

$$\frac{\tau^*}{1-\tau^*} = \frac{1}{\zeta}, \tag{10a}$$

which, with ζ given by (9), is our proposed formula for calculating the revenue-maximizing top tax rate τ^*.

Compared with the one proposed by Saez (2001), our formula (10a) has two advantages. First, the elasticity of the Pareto parameter ζ as defined by Eq. (6) takes into consideration the extensive as well as the intensive mobility of the top income group. Second, our formula only requires one parameter ζ. To see this second advantage more clearly, using Eq. (8) to replace ζ in Eq. (10a) yields

$$\frac{\tau^*}{1-\tau^*} = \frac{1}{e \cdot \alpha}, \tag{10b}$$

which is the first equation derived in Diamond and Saez (2011) and Eq. (7) derived by Saez et al. (2012). Formula (10b) has been the focal point of recent studies on revenue-maximizing top tax rates, including Guner et al. (2014), Kindermann and Krueger (2014), and Badel and Huggett (2015a, 2015b). The replacement of $e \cdot \alpha$ in (10b) by ζ in (10a) is significant in methodology, in that variations in α as observed in the real-world data are fully incorporated by ζ; by contrast, in applying formula (10b) one faces a difficult problem in practice: which value of α to plug in, since there are so many of them that can be chosen according to the data shown in Figure 1.

If $f(z)$ is Pareto-distributed asymptotically, then $f(z) \approx C/z^{\alpha+1}$for $z \geq \bar{z}$, where C is a positive constant. Thus we have

$$z_m - \bar{z} = \frac{\int_{\bar{z}}^{\infty} z f(z) dz}{\int_{\bar{z}}^{\infty} f(z) dz} - \bar{z} = \frac{\int_{\bar{z}}^{\infty} z^{-\alpha} dz}{\int_{\bar{z}}^{\infty} z^{-(\alpha+1)} dz} - \bar{z} = \bar{z} \cdot \frac{1}{\alpha - 1},$$

which indicates that once the threshold $\bar{z}$ is given, the top tax base $(z_m - \bar{z})$ is completely characterized by the Pareto parameter α. This implies that the elasticity ζ defined in (6) and estimated by (9) is completely measured by changes in α, which is true regardless of whether the mobility involved is intensive or extensive, and regardless of who are in the rich club. Intuitively, given the threshold $\bar{z}$, all the income mobilities involving $z \geq \bar{z}$ will make the income distribution at the top become either thinner or fatter and, as a consequence, alter its measure – the Pareto parameter α. As such, the elasticity of the Pareto parameter ζ is a sufficient statistic for summarizing changes in the top tax base with respect to top tax rates. By contrast, as is clear from (7) or (8), the aggregate elasticity of taxable income e by itself is not a sufficient statistic for summarizing changes in the top tax base with respect to top tax rates.

2.1.5 Compared with Badel and Huggett (2015b)

Badel and Huggett (2015b) studied the issue of revenue-maximizing top tax rates but extended formula (10b) to mainly count in the following possibility:

Example 4.

Suppose that the government cuts the top tax rate from τ to τ'''' so that the resulting outcome is: z_1 increases from 50 to 70 and all else remains unchanged. According to Badel and Huggett (2015b), this result may arise in dynamic models because the anticipation of the possibility of becoming a member of the rich club in the future may change the behavior of agents below the top bracket, even though the marginal tax rates these agents face do not change at all.

Since $h(z;\tau'''') = h(z;\tau)$ and (z_2, z_3, z_4) in the rich club remains exactly the same as before the tax cut, both Saez's (2001) behavioral response and our mobile response will assign no change in the top tax base and so there is no change in the tax revenue collected by the government. However, this lack of a change in the tax revenue may not be correct according to Badel and Huggett's (2015b) extended formula. The reason is that the revenue collected by the government may increase as a result of the increase in

z_1 from 50 to 70.

One may well refer to the effect such as that on z_1 in Example 4 as the "indirect effect," in the sense that they capture the impact of varying the top tax rate τ on the tax base other than the top one. With a calibrated structural model, it is relatively straightforward to take into consideration the indirect effect. However, to empirically estimate the indirect effect in the real world, it may not be easy. Our approach proposes a compromise: count in the indirect effect if and only if the change in z_1 is large enough so that agent z_1 becomes a member of the rich club. Note that we count in this indirect effect by a change in $h(z)$ itself or, more precisely, a change in the Pareto parameter α.

2.2 Welfare effect

Denote the utility of an agent with taxable income z by u_z. The welfare of all agents in the top tax bracket is defined by

$$S = \int_{\bar{z}}^{\infty} G(u_z)h(z)dz, \tag{11}$$

where the social welfare function $G(.)$ is increasing and concave in u_z, and recall that $h(z) = f(z)/\int_{\bar{z}}^{\infty} f(z)dz$. This welfare definition is in line with the literature following Mirrlees (1971); see Saez (2001, Eq. (10)).

The welfare criterion (11) allows for the possibility that $G(u_{z_1}) \neq G(u_{z_2})$ if $z_1 \neq z_2$. This is so despite both z_1 and z_2 belong to the rich club with $z_1 \geq \bar{z}$ and $z_2 \geq \bar{z}$. An alternative welfare criterion is to assume that the government ignores inequality within the rich and sees the rich only on average (i.e., only through their mean income z_m). It seems true that most people care about the income inequality of their economy, but not particularly about the income inequality among the rich. Under this alternative criterion,

S in (11) is replaced by

$$S(z_m) = \int_{\bar{z}}^{\infty} G(u_{z_m}) h(z) dz = G(u_{z_m}) \int_{\bar{z}}^{\infty} h(z) dz = G(u_{z_m}). \tag{12}$$

A neat feature of the criterion $S(z_m)$ is that variations in $h(z)$ due to $d\tau$ no longer matter for the welfare effect (though they still matter for the revenue effect).

Let p be the multiplier on the government budget constraint, which represents the social marginal value of government public funds; see Saez (2001) and Piketty and Saez (2013). Using the envelope theorem as shown by Saez (2001) and Piketty and Saez (2013), the social welfare variation for the rich due to $d\tau$ under the welfare criterion $S(z_m)$ is given by

$$W \equiv \frac{dS(z_m)}{p} = \frac{\partial G(u_{z_m})}{\partial \tau} \frac{1}{p} d\tau = -g(z_m)(z_m - \bar{z}) d\tau = -g(z_m) M, \tag{13}$$

which yields the same formula for the welfare effect as that derived by Saez (2001). Saez (2001) did not specify explicitly which welfare criterion he employed in his derivation of top income optimal tax rates. However, Saez (2001, p. 210) did note: "$\bar{g}$ [$g(z_m)$ in our notation] is defined such that the government is indifferent between $\bar{g}$ more dollars of public funds and one more dollar consumed by the taxpayers with income above $\bar{z}$."

To facilitate comparisons with Saez (2001), our welfare effect will be set equal to (13).

2.3 Top income optimal tax rates

The sum of the revenue and the welfare effect constitutes the first-order effect, which is equal to zero at the optimum. Thus, $M + D + W = 0$ gives rise to the optimal top tax rate implicitly determined by

$$[1 - g(z_m)]M = -\tau dz_m, \tag{14}$$

where the term $[1 - g(z_m)]M$ represents the marginal benefit of $d\tau$ (MB), while the term $-\tau dz_m$ represents the corresponding marginal cost (MC). As is typical, we have $MB = MC$ at the optimum. On the basis of (14) and using the definition of M and ζ, one can derive the optimal top tax rate τ^{**} as

$$\frac{\tau^{**}}{1-\tau^{**}} = \frac{1-g(z_m)}{\zeta}. \tag{15a}$$

Using (8), we also have

$$\frac{\tau^{**}}{1-\tau^{**}} = \frac{1-g(z_m)}{e \cdot \alpha}. \tag{15b}$$

The optimal tax rate τ^{**} of (15a)-(15b) differs from the revenue-maximizing tax rate τ^* of (10a)-(10b) simply because the numerator is $1 - g(z_m)$ rather than 1 or, put it differently, $g(z_m) \neq 0$. Note that formula (15b) is basically the same as that derived by Piketty and Saez (2013, Eq. (7)).[13]

Two extremes If $\bar{z} \to z_m$, then $\zeta \to \infty$ according to (7) and so $\tau^{**} \to 0$ from (15a). This is the celebrated "zero top" result of Sadka (1976) and Seade (1977). On the other hand, if the threshold $\bar{z} = 0$, then $\alpha = 1$ and $\zeta = e$ according to (7)-(8); as a result, implementing (15a) or (15b) to empirically estimate τ^{**} no longer makes any real difference.

[13] A deviation is that their g term is defined more generally, based on S in (11) rather than $S(z_m)$ in (12). They can have a more general g mainly because the derivation of their formula builds on the behavioral response and so $h(z)$, i.e., those in the rich club, remain unchanged throughout their derivation.

Remark 2 *Saez (2001) is a seminal paper that finds an inverse relationship between the optimal top tax rate* τ^{**} *and* $e \cdot \alpha$*, where* e *is specified according to (5). After the derivation, he noted: "Interestingly, the optimal rate is an increasing function of* $z_m/\bar{z}$*. The ratio* $z_m/\bar{z}$ *is a key parameter for the high income optimal tax problem. ... Distributions with constant ratio* $z_m/\bar{z}$ *are exactly Pareto distributions." (pp. 210-211) However, Saez did not explicitly explain why the Pareto parameter* α *itself is present in the formula for top income optimal tax rates as shown in formula (15b). The equality* $\zeta = e \cdot \alpha$ *as given by (8) explains why: the top tax base is* $z_m - \bar{z}$ *rather than* z_m*; as a result, if the elasticity* e *rather than the elasticity* ζ *is employed to summarize changes in the top tax base, it needs to be adjusted by the term* $\frac{z_m}{z_m - \bar{z}}$ *as shown in (7) and this term happens to equal the Pareto parameter* α *since* $z \geq \bar{z}$ *is Pareto-distributed. Although a simple feature, it seems that the extant literature fails to recognize it. This may largely explain why elasticity* e *of (10b) and (15b) rather than the sufficient-statistic elasticity* ζ *of (10a) and (15a) has been the focal point of the literature.*

3 Empirical application

This section applies our derived formula to quantitatively characterize top income optimal tax rates for the U.S. economy. To empirically implement formula (15a), we need to know both $g(z_m)$ and ζ. We shall consider two values for $g(z_m)$ as in Saez (2001): $g(z_m) = 0$ and $g(z_m) = 0.25$. This then leaves us with the task of estimating ζ only.

Before reporting our estimation on the elasticity of the Pareto parameter ζ, however, we provide preliminary evidences on the significance of extensive mobility in the rich club. As argued earlier, a critical difference between our mobile response and Saez's (2001) behavioral response lies in that while we explicitly consider agents' mobility into and out of the rich club, Saez (2001) did not.

3.1 Evidence on extensive mobility

We focus on the case of the 1986 Tax Reform Act (TRA 1986) and utilize the University of Michigan Tax Panel to estimate tax effects on agents' mobility into and out of the top income groups. TRA 1986 might have been anticipated earlier than the year 1986 and so agents might have began to adjust their behavior before it was actually enacted. Following the practice of previous studies, we use the 1985 and 1988 panel data for the investigation.[14]

We conduct two types of regressions to estimate the tax effects on agents' mobility into and out of the top income groups based on transition analyses (e.g., Bruce, 2000). The basic idea of transition analyses is to study how some factor may affect an individual's transition between different states. We follow the approach to carry out our estimation. The first type of regression is mainly to estimate the effect of changing taxes on agents' mobility out of the top income groups, which is illustrated by Example 2 in Section 2.1.2. We keep the top 1% of taxpayers in 1985 and categorize them into two groups according to their income in 1988, one for those remaining in the top 1% in 1988 while the other for those moving out of the top 1% in 1988.[15] Our focus is on the top 1% of taxpayers, but we also consider the top 5% of taxpayers as a comparison.

The second type of regression is mainly to estimate the effect of changing taxes on agents' mobility into the top income groups, which is illustrated by Example 3 in Section 2.1.2. We exclude the top 1% group in 1985 and keep the taxpayers who were not in

[14]For example, Auten and Carroll (1999) estimated the effect of taxes on the change in taxable income between 1985 and 1989 while Feldstein (1995) estimated the effect between 1985 and 1988.

[15]Only the returns available for both 1985 and 1988 are kept for transition analyses to avoid the influence of unbalanced panels on the results.

the top 1% group in 1985. We then categorize them into two groups according to their income in 1988, one for those moving up to the top 1% in 1988 while the other for those remaining in the lower 99% in 1988. Again, we also consider the top 5% of taxpayers as a comparison.

We run probit estimations for the two above-mentioned types of regressions with respect to the main explanatory variable, namely, the change in the logarithm of the net-of-tax rates. Because the tax rates are functions of taxable income in a graduated income tax system, they are endogenous to income. We employ an instrumental variable approach to deal with the endogeneity problem. The instruments for the change in the logarithm of the net-of-tax rates are calculated based on the synthetic tax price as in Auten and Carroll (1999). We also include various control variables based on Auten and Carroll (1999) and Bruce (2000). These variables include the income in the initial year, marital status, dependents, residence regions, whether having sole proprietorship, and whether having capital gains or losses. Taxpayers' mobility can result from mean reversion (Auten and Carroll, 1999), so including the income in the initial year as a control variable is particularly important for estimating the effect of taxes on mobility. Since the tax return data do not provide demographic variables such as age, education, occupation, etc., they are not accounted for in the estimations.

(Insert Table 1 about here)

Table 1 reports the results, including the two types of probit estimations for the top 1% and 5% income groups. Following Gruber and Saez's (2002) definition of high-income in terms of broad income (explained in Footnote 4), we also conduct these two types of probit estimations based on the income threshold of $100,000 in 1992 dollars. As shown in Table 1, the tax effects are only significant for agents' mobility up to the very top income groups but are statistically insignificant for agents' mobility out of the top income groups.[16]

The coefficient estimates for agents' mobility up to the very top income groups (i.e., top 1% or above the threshold of $100,000) are positive and statistically significant. For example, the coefficient estimate of 3.11 in Table 1 implies that a reduction in marginal tax rates on top-income taxpayers from 50% to 28% would on average raise the probability of taxpayers' mobility to the rich group by 6.54%. Since the influence of mean reversion has been controlled for, the results indicate that a tax cut makes it more likely for agents to move up to the very top income groups. The results also suggest that tax reforms would change the top tax base as tax reductions induce taxpayers into the top income group. By contrast, although the coefficient estimates are negative as expected in theory, the tax effects on mobility out of the top income group are not statistically significant for a variety of definitions of the top income group. This asymmetry of the results on income mobility seems not unreasonable, in view of the fact that TRA 1986 is generally associated with a reduction rather than an increase in the tax rates for high-income taxpayers.

3.2 Estimating the elasticity of the Pareto parameter

On the basis of Eq. (9), let us consider a basic time-series regression of the following form for estimating the elasticity ζ,

$$\log(\frac{1}{\alpha_t - 1}) = b + \zeta \log(1 - \tau_t) + \varepsilon_t, \tag{16}$$

[16]Since transition analyses are based on only a sample who are in a state in the initial year, a potential bias may arise with the sample selection. Bruce (2000) suggested calculating an inverse Mills ratio using the estimates of the first-stage probit of agents' initial states and then including the ratio as a regressor in the transition probit regressions to correct for the bias. We try including the inverse Mills ratio in the transition probit regressions, but do not find evidence of a significant influence on the estimates.

where b is some constant. It is interesting to observe from (16) that the elasticity ζ is not simply of $\log \alpha_t$ with respect to $\log(1-\tau_t)$, but of $\log(\frac{1}{\alpha_t - 1})$ with respect to $\log(1-\tau_t)$. Since $(1-\tau_t)$ is likely to be correlated with the error term ε_t in the regression (16), it is necessary to find exogenous $(1-\tau_t)$, or instrumental variables correlated with $(1-\tau_t)$ but uncorrelated with ε_t, to identify the elasticity ζ.

There is a large literature on estimating the elasticity of taxable income (ETI) with respect to marginal tax rates using tax return data. Although ETI is different from our ζ, its estimation faces a similar problem as in our regression (16): the explanatory variable $(1-\tau_t)$ is likely to be correlated with the error term ε_t. Saez et al. (2012) critically reviewed this literature. The review by Saez et al. (2012) overall conveys the message that finding satisfactory instruments by the difference-in-differences method is challenging, if not formidable.[17] Note in particular that when applied to panel data, the method fails to count in the extensive mobility of the top income group as Examples 2 and 3 in Section 2.1.2 illustrate.

Romer and Romer (2010) and Mertens (2015) made use of narrative records such as presidential speeches and Congressional reports in relation to significant pieces of U.S. federal tax legislation to identify exogenous tax changes. We follow their approach in our estimation of the elasticity ζ.

To employ the narrative methodology of utilizing exogenous tax changes, we consider a regression specification in terms of changes rather than levels in the tax rate. From (16), we have

$$\log(\frac{1}{\alpha_{t-1} - 1}) = b + \zeta \log(1 - \tau_{t-1}) + \varepsilon_{t-1}. \tag{17}$$

[17] Weber (2014) and Burns and Ziliak (2016) recently proposed new methods to estimate ETI. The former paper particularly showed that most of the existing instruments used in difference-in-differences are not exogenous.

Eq. (16) minus Eq. (17) leads to

$$\log(1-\frac{\Delta\alpha_t}{\alpha_t-1}) = \zeta\log(1-\frac{\Delta\tau_t}{1-\tau_{t-1}}) + v_t, \tag{18}$$

.

where $\Delta\alpha_t = \alpha_t - \alpha_{t-1}$, $\Delta\tau_t = \tau_t - \tau_{t-1}$ and $v_t = \varepsilon_t - \varepsilon_{t-1}$. Analogous to Eq. (7) in Romer and Romer (2010), we estimate

$$\log(1-\frac{\Delta\alpha_t}{\alpha_t-1}) = \zeta\log(1-\frac{\Delta\tau_t}{1-\tau_{t-1}})+\sum_{j=1}^{M}\zeta_j\log(1-\frac{\Delta\tau_{t-j}}{1-\tau_{t-1-j}})+\sum_{j=1}^{N}\beta_j\log(1-\frac{\Delta\alpha_{t-j}}{\alpha_{t-j}-1})+v_t, \tag{19}$$

where the lagged terms of $\Delta\tau$ and $\Delta\alpha$ are added to (18) as controls. We use Akaike information criterion (AIC) and Bayesian information criterion (BIC) to choose the optimal time lags for $\Delta\tau$ and $\Delta\alpha$.

Romer and Romer (2010) argued that governments often implement tax changes in response to macroeconomic conditions and thus tax changes are usually correlated with aggregate economic activities. As such, instrumenting with statutory tax changes may not address the endogeneity of tax changes adequately if tax reforms are also results of aggregate economic conditions. They examined the narrative records of major U.S. legislated tax changes over the period 1945-2007, which described the history and motivation of tax policy changes, to determine whether tax changes are taken for reasons related to prospective economic conditions or taken for more exogenous reasons such as arising from ideological or inherited deficit concerns. Romer and Romer (2010) identified 54 quarterly exogenous tax changes over the period 1945-2007 in estimating the effect of taxes on aggregate output. Built on Romer and Romer's (2010) classification of exogenous tax changes, Mertens (2015) further restricted exogenous tax changes to those legislative changes that affected individual income tax liabilities and were implemented without delay in estimating the effect of marginal tax rate changes on reported income. In the end, he identified 15 yearly exogenous changes in statutory income tax rates over the post-World War II period 1946-2012. Since our focus is also on the effect of marginal tax rate changes as in Mertens (2015), we apply these 15 yearly exogenous tax changes to identify the elasticity ζ in the regression (19).[18]

To carry out the estimation of (19), we need to obtain values for the Pareto parameter, α_t, and the top tax rate, τ_t. The Pareto parameter can be calculated by utilizing the formula $\alpha = z_m/(z_m - \bar{z})$, where $\bar{z}$ denotes the threshold income for the top income group, and z_m denotes the mean income for the top income group. Piketty and Saez (2003) provided data for the threshold income and the mean income for different income groups. On the basis of the reasonable assumption that the income distribution at the top satisfies the shape of the Pareto distribution, we can obtain the Pareto parameter using the data provided by Piketty and Saez (2003).[19]

The top tax rates applicable to taxpayers in the top income group, say, the top 1%, could vary over a wide range. This is especially true before the passage of the Economic Recovery Tax Act of 1981 and the Tax Reform Act of 1986, both of which substantially cut the number of tax brackets.[20] Based on the methodology of Barro

[18] The estimated impact of a selected tax reform is the change between the counterfactual tax rate based on the current tax law and pre-reform income distribution, and the tax rate based on the pre-reform tax law and pre-reform income distribution (Mertens, 2015).

[19] Please see the extended data file attached to Piketty and Saez (2003) at http://eml.berkeley.edu/~saez/. Income is broad income as defined in Footnote 4. Piketty and Saez (2003) provided various categories of individual income. To be comparable to previous studies like Gruber and Saez (2002), the Pareto parameters in our estimations are based on individual income including all income items reported in tax returns before all deductions but excluding capital gains realization. Specifically, they are based on Table A4 of the data file of Piketty and Saez (2003).

[20] For example, the cutoff income for the top 1% income was about $\$76,300$ in 1980 in current dollars. For a married household of five people, the taxable income equals $\$67,900$ ($= 76,300 - 3,400 - 5 \times 1,000$) after exemptions and standard deductions are subtracted. However, the marginal tax rates for taxable income above $\$67,900$ could range from 54% to 70% in 1980 and, therefore, the marginal tax rates for the top 1% taxpayers could range from 54% to 70% in 1980.

and Sahasakul (1986) and the income percentiles of Piketty and Saez (2003), Mertens (2015) constructed from U.S. federal tax return statistics annual time series for a weighted average of individual marginal tax rates with weights given by income shares or the so-called "average marginal tax rate" (AMTR), namely, the combined tax rates of both average marginal individual income tax rates and average marginal social security tax rates, from 1946 to 2012 for different income percentile brackets. We directly use his calculated AMTRs as the marginal tax rates facing different income groups.

To sum up, we obtain Pareto parameters based on the extended income distribution data provided by Piketty and Saez (2003), and obtain the top tax rates by utilizing the AMTRs corresponding to different top income groups from Mertens (2015).

Applying the unit root test shows that the time series of both left- and right-hand-side variables of the regression (19) are stationary. The AIC and BIC generally suggest models with one lag or at most two lags. Therefore, besides the contemporaneous $(1 - \frac{\Delta\tau_t}{1-\tau_{t-1}})$ on the right-hand side of (19), we include two lags of it and two lags of the dependent variable to account for the possible lagged impact from tax changes and serial correlation. To account for possible macro fluctuations of GDP on income distribution, we also include the growth rate of per capita GDP. Government expenditures affect output and possibly income distribution, and so the growth rate of per capita government expenditure is also included as a part of the controls in our regressions.[21] Similar to the arguments put forth by Romer and Romer (2010) and Mertens (2015), if the underlying reasons for tax changes are exogenous, a regression on these exogenous tax changes should yield unbiased estimates of the elasticity ζ. We thus utilize the exogenous tax changes over 1946-2012

[21] The government expenditure data are from Table 3.2 of the National Income and Product Accounts, which were accessed on July 24, 2015. We divide government expenditures by both the price index for GDP and population to convert them to per capita real government expenditures. We then take the first difference to obtain the growth of per capita government expenditures.

identified by Mertens (2015) to instrument actual changes in AMTR in the regressions. As Mertens (2015) noted, however, only eight out of these 15 exogenous tax changes have measurable changes in AMTRs; the other seven exogenous tax changes are set to zero.

Our empirical focus is on the 1% top income group, which corresponds approximately to the top federal income tax bracket in recent years (Saez et al., 2012). However, for reference purposes, we also estimate ζ for the income groups of the top 10% and 5%. Our data period is from 1946 to 2012. The estimations are based on 64 observations over 1949-2012 because the variables are in the first-difference and two periods of lags are included in the regression. Table 2 reports our estimation results.

(Insert Table 2 about here)

We are mainly interested in the estimate of the elasticity ζ, namely, the coefficient estimate for the variable $(1 - \frac{\Delta\tau_t}{1-\tau_{t-1}})$ in the regression (19). The coefficient estimates of ζ for the top 10%, 5%, and 1% are equal to 1.220, 1.574, and 0.899 in the baseline cases in columns (1)-(3) of Table 2. The estimates are all statistically significant. Although our focus is on the top 1% income group, it is interesting to note that the estimates of ζ do not monotonically decrease or increase with respect to the top income groups. This finding is somewhat similar to the non-monotonic estimates found by Gruber and Saez (2002) for the elasticity of taxable income across different income groups.

From the definition of the elasticity ζ in Eq. (6) we know that the estimation results of ζ in Table 2 have clear economic interpretations. For example, the estimation result of ζ in column (3) is 0.899. This estimate means that a 1% increase in the net-of-tax rate of the top 1% group raises the corresponding top tax base by 0.899%.

3.3 Robustness

Although we adopt the method of Romer and Romer (2010) and Mertens (2015) by utilizing exogenous tax changes to instrument actual tax changes in estimating the effect of tax changes on income distribution, other non-tax factors may also affect the income distribution and cause omitted variable bias if they are not controlled for. Besides the growth of per capita GDP and per capita real government expenditure in the baseline specification, we examine the effects of including other control variables.

Political ideology is a factor which is likely to affect government policy and income distribution. Presidents of different parties emphasize income growth for different income groups. For example, Bartels (2004) contended that Democratic presidents have advocated for policies which produce slightly more income growth for poor families than for rich families and consequently result in a modest decrease in overall inequality. Because Eq. (19) is in the first-difference form, we thus re-run Eq. (19) by including two dummies to represent changes in the party affiliation of presidents.[22] Columns (4)-(6) of Table 2 reports the results. Compared to the baseline cases, the estimates of ζ for the top 10%, 5%, and 1% become 1.234 (S.E.=0.655), 1.402 (S.E.=0.464), and 0.808 (S.E.=0.198), respectively, when we account for the party affiliation of presidents. There are no significant differences in general.

It is possible that the changes in income inequality over time are owing to reasons unrelated to tax changes (e.g., Saez et al., 2012 (p. 22)). The ETI literature suggests including time trends to control for these possible factors. Therefore, we further account for linear, square and cubic time trends in the estimations and find that the estimates of

[22] Specifically, one dummy represents the change from a Democratic president to a Republican president, while the other dummy represents the change from a Republican president to a Democratic president.

ζ for the top 10%, 5%, and 1% are equal to 1.085 (S.E.=0.656), 1.392 (S.E.=0.396), and 0.812 (S.E.=0.163), respectively. The estimates in columns (7)-(9) are similar to those in columns (4)-(6), except that the estimate for the top 10% becomes somewhat smaller.

The estimates provided in Table 2 are based on AMTR, namely, the combined tax rates of both average marginal individual income tax rates (AMIITR) and average marginal social security tax rates. Mertens (2015) provided separate data on both tax rates for different income groups over the period 1946-2012. Although social security taxes did not comprise a major part of AMTR until the 1980s, average marginal social security tax rates have been substantial for top income groups in recent decades. Whether individuals respond to individual income taxes only or to the sum of individual income taxes and payroll taxes is an empirical issue. We therefore re-estimate Eq. (19) by replacing AMTR with AMIITR. The estimates of ζ are similar in magnitude. For example, the estimates of ζ for the specifications with controls for party affiliation of presidents and time trends are 1.171 (S.E.=0.845), 1.523 (S.E.=0.493), and 0.818 (S.E.=0.173) for the top 10%, 5%, and 1% groups, respectively. The estimates of ζ are all statistically significant except for the estimate of the top 10%.

Romer and Romer (2010) and Mertens (2015), as noted above, contended that instrumenting actual tax changes with statutory tax changes alone does not overcome the endogeneity of tax policy because tax reforms might be driven by the concern for prospective economic conditions and statutory tax changes are not totally independent of economic conditions and income distribution. We re-estimate Eq. (19) to evaluate potential bias utilizing all statutory tax changes for instruments.[23] We find that the estimates utilizing the instruments of all statutory tax changes are substantially larger than those in Table 2 for the specifications without controlling for party affiliations and

[23] Mertens (2015) identifies twenty-eight yearly statutory tax changes over 1946-2012.

time trends. Besides, the estimates corresponding to columns (6) and (9) become larger and are equal to 0.975 (S.E.=0.203) and 0.937 (S.E.=0.203), respectively. The larger estimates suggest potential endogeneity of some statutory tax changes.

The identification of the tax effects hinges on the relatively low number of tax changes, namely, eight measurable exogenous tax changes, and so the estimates may thus be sensitive to the inclusion of a particular tax reform (Mertens, 2015). We follow Mertens (2015) to examine this issue by alternatively replacing one of the eight exogenous tax changes with zero and re-run the estimation. Taking the baseline specification as an example, we find that the estimates of ζ are basically robust to exclusions of any particular tax reform in the sense that the magnitudes of the estimates are similar to those in the baseline case and the point estimates are statistically significant except for the cases of the 1964 Kennedy tax reform and the 1986 tax reform. Excluding the 1964 Kennedy tax reform leads to larger estimates only for the top 10% and 5%, which become 1.712 (S.E.=0.180) and 1.774 (S.E.=0.280). By contrast, excluding the 1986 tax reform results in smaller estimates for all three top income groups, which are 0.419 (S.E.=0.797), 0.986 (S.E.=0.509) and 0.741 (S.E.=0.238) for the top 10%, 5%, and 1% groups, and the estimate for the top 10% becomes statistically insignificant.

To sum up, the empirical evidence overall suggests that our estimates of the elasticity ζ in Table 2 are robust to accounting for the party affiliation of presidents and time trends. Although one or two tax reforms exert a larger influence on the coefficient estimates, the influence is mainly on the top 10% and 5% groups. The estimates for the top 1%, which are our central focus, are robust to different specifications and the exclusion of a particular tax reform.

3.4 Quantitative top income optimal tax rates

After obtaining the estimates of the elasticity ζ, it becomes straightforward to quantitatively characterize top income optimal tax rates – by simply applying our derived formula (15a). Our focus is on the top 1% income group. Choosing $\zeta = 0.812$ in column (9) of Table 2, which is our preferred estimate, we report the results in Table 3.

(Insert Table 3 about here)

The optimal top tax rates for the top 1% income group range from 0.48 to 0.55, depending on whether $g(z_m) = 0.25$ or $g(z_m) = 0$. Note that when $g(z_m) = 0$, it corresponds to the revenue-maximizing top tax rate.

We next apply (15b), which is the formula widely used in the extant literature both theoretically and empirically. It is known that top income optimal tax rates are sensitive to the choice of e (Saez et al., 2012). In their quantitative illustration for revenue-maximizing top tax rates, Saez et al. (2012) chose an elasticity $e = 0.25$, which roughly corresponds to the mid-range of the estimates from the ETI literature. We use the same quantitative choice for the elasticity e. As in Saez (2001), we consider three values for the Pareto parameter α, 1.5, 2, and 2.5. Table 4 reports the results.

(Insert Table 4 about here)

Saez (2001) noted in reporting his results that the Pareto parameter exerts a big effect on top income optimal tax rates. This feature also stands out in Table 4. If the Pareto parameter $\alpha = 2.5$, then the optimal top tax rates in Table 4 (0.62 and 0.55) are not far away from those in Table 3 (0.55 and 0.48). However, if $\alpha = 2$ or $\alpha = 1.5$, which are the values close to more recent data on income concentration, top income optimal tax rates from our formula will be lower than those in Saez (2001) to a significant extent. Note that all of these three values $\alpha = 1.5$, 2, 2.5 are compatible with the data shown in Figure 1. In fact, all values between 1.5 and 2.5 are compatible with the data shown in Figure 1. As we emphasize earlier, a problem with applying formula (15b) in empirical studies is that there are so many values of α that can be chosen according to the real-world data.

Given $g(z_m)$, Table 4 shows a clear inverse pattern between the Pareto parameter and the optimal top tax rate: the lower the Pareto parameter, the higher the optimal top tax rate. This inverse relationship can be seen directly from formula (15b). By contrast, our derived formula (15a) shows that top income optimal tax rates are inversely related to the elasticity of the Pareto parameter rather than the Pareto parameter itself.

4 Conclusion

The academic recommendation for the top income tax rate has changed dramatically from the "zero top" result of Sadka (1976) and Seade (1977) to the "fairly high" result of Saez (2001). This paper revisits the work of Saez (2001) and refines his derived formula for top income optimal tax rates. We contribute to the literature by using the elasticity of the Pareto parameter rather than the elasticity of taxable income to summarize the response of the top tax base with respect to top tax rates. Our approach generalizes Saez's (2001) formula by incorporating the extensive as well as the intensive income mobility in the top income group. Importantly, while the elasticity of the Pareto parameter is a sufficient statistic for summarizing changes in the top tax base, the elasticity of taxable income is not. The results from our proposed formula are immune from a big impact on optimal top tax rates by arbitrarily choosing some value of the Pareto parameter as in Saez (2001). We empirically estimate the elasticity of the Pareto parameter for the top 1% income group based on the exogenous tax changes identified by the narrative method of Romer and Romer (2010) and Mertens (2015). The top income optimal tax rates resulting from our refined formula are to some extent lower than those in Saez (2001).

How a government should set the top income tax rate is not only of academic interest but also of practical relevance. Both the Obama administration of the U.S. government and the Hollande administration of the French government have made moves to substantially raise the top income tax rates applied to the rich. The Japan and the Taiwan governments recently raised the top income tax rate from 40% to 45%, and the Singapore government plans to raise the top income tax rate from 20% to 22% from 2017 onward. In her campaign for Presidency, Hillary Clinton recently proposed 4% income-tax surcharge for wealthy Americans whose incomes exceed $5 million a year. All of these moves have given rise to both supporting and opposing arguments. Our derived formula for top income optimal tax rates could provide a platform for exchanging different arguments both for and against.

References

Alvaredo, F., Atkinson, A.B., Piketty, T., Saez, E., 2011. The World Top Income Database, online at http://g-mond.parisschoolofeconomics.eu/topincomes.

Alvaredo, F., Atkinson, A.B., Piketty, T., Saez, E., 2013. The top 1 percent in international and historical perspective. Journal of Economic Perspectives 27(3), 3-20.

Auten, G., Carroll, R., 1999. The effect of income taxes on household behavior. Review of Economics and Statistics 81, 681-693.

Badel, A., Huggett, M., 2015a. Taxing top earners: a human capital perspective. Mimeo.

Badel, A., Huggett, M., 2015b. The sufficient statistic approach: predicting the top of the Laffer curve. Federal Reserve Bank of St. Louis Working Paper 2015-038A.

Bartels, Larry M., 2004. Partisan Politics and the U.S. Income Distribution. Mimeo.

Barro, R.J., Sahasakul, C., 1986. Average marginal tax rates from social security and the individual income tax. Journal of Business 59(4), 555-566.

Bruce, D., 2000. Tax effects of the United States tax system on transition into self-employment, Labour Economics 7, 545-574.

Burns, S.K., Ziliak, J.P., 2016. Identifying the elasticity of taxable income. Economic Journal, forthcoming.

Diamond, P., Saez, E., 2011. The case for a progressive tax: from basic research to policy recommendations. Journal of Economic Perspectives 25, 165-190.

Feenberg, D.R., Poterba, J.M., 1993. Income inequality and the incomes of very high-income taxpayers: evidence from tax returns. In: Poterba, J. (Ed.), Tax Policy and the Economy, Volume 7. MIT Press, Cambridge.

Feldstein, M., 1995. The effect of marginal tax rates on taxable income: a panel study of the 1986 Tax Reform Act. Journal of Political Economy 103, 551-572.

Feldstein, M., 1999. Tax avoidance and the deadweight loss of the income tax. Review of Economics and Statistics 81, 674-680.

Gruber, J., Saez, E., 2002. The elasticity of taxable income: evidence and implications. Journal of Public Economics 84, 1-32.

Guner, N., Lopez-Daneri, M., Ventura, G., 2014. Heterogeneity and government revenues: higher taxes at the top? IZA Discussion Paper No. 8335.

Kindermann, F., Krueger, D., 2014. High marginal tax rates on the top 1%? Lessens from a life cycle model with idiosyncratic income risk. NBER Working Paper No. 20601.

Mertens, K., 2015. Marginal tax rates and income: new time series evidence. Mimeo.

Mirrlees, J.A., 1971. An exploration in the theory of optimum income taxation. Review of Economic Studies 38, 175-208.

Piketty, T., Saez, E., 2003. Income inequality in the United States, 1913-1998. Quarterly Journal of Economics 118, 1-39.

Piketty, T., Saez, E., 2013. Optimal labor income taxation, in Handbook of Public Economics, Volume 5, edited by A.J. Auerbach, R. Chetty, M. Feldstein, E. Saez, Elsevier.

Piketty, T., Saez, E., Stantcheva, S., 2014. Optimal taxation of top labor incomes: a tale of three elasticities. American Economic Journal: Economic Policy 6, 230-271.

Romer, C.D., Romer, D.H. 2010. The macroeconomic effects of tax changes: estimates based on a new measure of fiscal shocks. American Economic Review 100, 763-801.

Sadka, E., 1976. On income distribution, incentive effects, and optimal income taxation. Review of Economic Studies 43, 261-267.

Saez, E., 2001. Using elasticities to derive optimal income tax rates. Review of Economic Studies 68, 205-229.

Saez, E., Slemrod, J.B., Giertz, S.H., 2012. The elasticity of taxable income with respect to marginal tax rates: a critical review. Journal of Economic Literature 50, 3-50.

Scheuer, F., Werning, I., 2015. The taxation of superstars. CESIFO Working Paper No. 5479.

Seade, J., 1977. On the shape of optimal income tax schedules. Journal of Public Economics 7, 203-235.

Slemrod, J., 1996. High-income families and the tax changes of the 1980s: the anatomy of behavioral response. In: Empirical Foundations of Household Taxation. Feldstein, M., Poterba, J.M., (Eds.), University of Chicago Press, Chicago.

Weber, C.E., 2014, Toward obtaining a consistent estimate of the elasticity of taxable income using difference-in-differences. Journal of Public Economics 117, 90-103.

Table 1. Probit estimates of tax effects on agents' extensive mobility

	Mobility into			Mobility out of		
	Top 1%	Top 5%	\$100,000	Top 1%	Top 5%	\$100,000
$\Delta \log(1-\tau)$	5.17***	-0.79	3.11***	-1.37	-1.26	-0.52
	(0.97)	(1.78)	(0.93)	(2.38)	(0.93)	(1.53)
log(initial income)	0.83***	1.47***	1.35***	-0.48*	-1.13***	-0.70***
	(0.22)	(0.22)	(0.19)	(0.29)	(0.23)	(0.29)
Other control variables						
Marital status	✓	✓	✓	✓	✓	✓
Dependents	✓	✓	✓	✓	✓	✓
Sole proprietorship/farming	✓	✓	✓	✓	✓	✓
Capital gains/losses	✓	✓	✓	✓	✓	✓
Regions	✓	✓	✓	✓	✓	✓
Log likelihood	17233.0	12025.8	16561.7	108.7	600.0	304.5
Observations	16712	16115	16561	171	856	403

Notes: Standard errors are in parentheses; ***, **, and * denote statistical significance at the level of 1%, 5%, and 10%, respectively. The dummies for sole proprietorship and/or farming are based on Schedule C or F while the dummies for capital gains or losses are based on the capital gains indicator in tax returns.

Table 2. Estimates of ζ for different top income groups

Income group	Top 10%	Top 5%	Top 1%	Top 10%	Top 5%	Top 1%	Top 10%	Top 5%	Top 1%
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$\log(1-\frac{\Delta\tau_t}{1-\tau_{t-1}})$	1.220**	1.574***	0.899***	1.234*	1.402***	0.808***	1.085*	1.392***	0.812***
	(0.519)	(0.367)	(0.148)	(0.655)	(0.464)	(0.198)	(0.656)	(0.396)	(0.163)
1-year lag	0.305	-0.252	0.527	0.471	0.033	0.804*	0.689	0.235	1.138***
	(0.615)	(0.603)	(0.447)	(0.652)	(0.630)	(0.445)	(0.650)	(0.583)	(0.422)
2-year lag	-0.550	0.155	-0.804	-0.440	0.032	-0.699	-0.051	0.347	-0.137
	(0.675)	(0.738)	(0.541)	(0.623)	(0.764)	(0.536)	(0.548)	(0.622)	(0.517)
GDP growth	0.008**	0.011**	0.009*	0.007*	0.009*	0.007	0.005	0.008	0.006
	(0.004)	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
1-year lag	0.001	0.001	0.002	0.001	0.003	0.003	0.002	0.004	0.005
	(0.003)	(0.004)	(0.005)	(0.004)	(0.005)	(0.005)	(0.003)	(0.004)	(0.005)
2-year lag	-0.004	-0.006	-0.006	-0.001	-0.003	-0.002	0.002	-0.000	0.002
	(0.004)	(0.005)	(0.005)	(0.003)	(0.004)	(0.004)	(0.003)	(0.005)	(0.005)
Gov. exp. growth	-0.137	-0.259*	-0.239*	-0.161	-0.272*	-0.249**	-0.167	-0.206	-0.148
	(0.127)	(0.154)	(0.136)	(0.138)	(0.140)	(0.126)	(0.143)	(0.164)	(0.151)
1-year lag	-0.110	-0.007	0.060	-0.165	-0.055	0.034	-0.220*	-0.118	-0.005
	(0.150)	(0.179)	(0.161)	(0.125)	(0.167)	(0.141)	(0.122)	(0.128)	(0.151)
2-year lag	0.058	-0.005	-0.076	0.163	0.186	0.106	0.184	0.230*	0.143
	(0.123)	(0.142)	(0.158)	(0.114)	(0.137)	(0.131)	(0.117)	(0.136)	(0.141)
$\log(1-\frac{\Delta\alpha_t}{\alpha_t-1})$									
1-year lag	0.052	-0.004	-0.208	-0.085	-0.164	-0.393*	-0.310*	-0.376**	-0.728***
	(0.162)	(0.168)	(0.230)	(0.167)	(0.171)	(0.218)	(0.172)	(0.162)	(0.263)
2-year lag	0.301*	0.067	0.483**	0.168	-0.012	0.363	-0.116	-0.255	-0.067
	(0.164)	(0.194)	(0.227)	(0.153)	(0.198)	(0.230)	(0.177)	(0.188)	(0.302)
Political affiliation				✓	✓	✓	✓	✓	✓
Time polynomial							✓	✓	✓
Sample period	1946-2012								

Notes: Party affiliation denotes the party of the President. Time polynomials include linear, square, and cubic time trends from 1946-2012. Newey-West standard errors are in parentheses; ***, **, and * denote statistical significance at the level of 1%, 5%, and 10%, respectively.

Table 3. Optimal top tax rates

Income group	Top 1%
$g(z_m) = 0$	55
$g(z_m) = 0.25$	48

Table 4. Saez's (2001) optimal top tax rates with $e = 0.25$

Pareto parameter	1.5	2	2.5
$g(z_m) = 0$	73	67	62
$g(z_m) = 0.25$	67	60	55

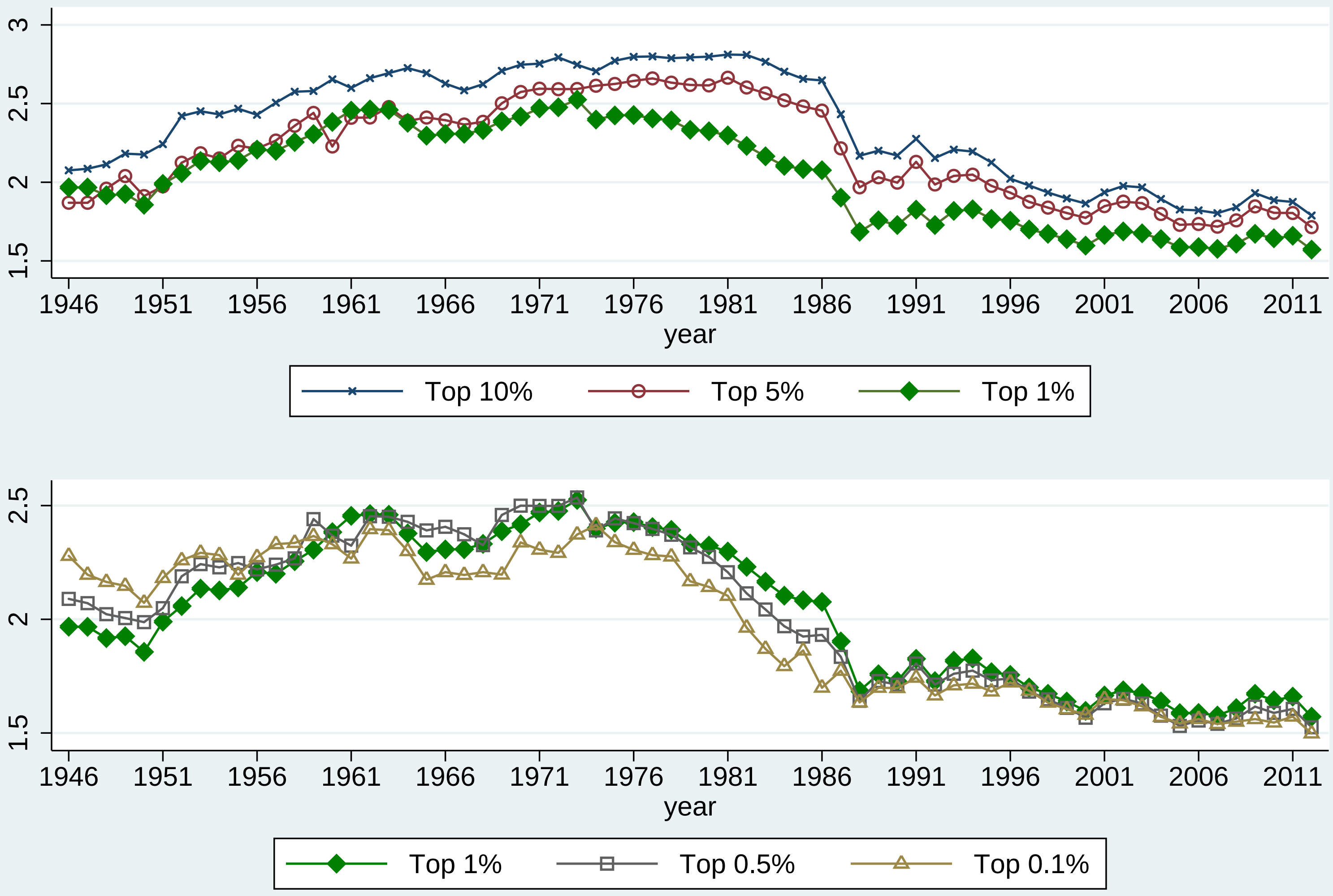

Figure 1. Time−series patterns of the Pareto parameter in the U.S. during 1946−2012