Revenue-maximizing top tax rates revisited: mobile response and the elasticity of the Pareto parameter*

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

June 15, 2017

Abstract

The Pareto parameter is a measure of the thickness of the income distribution at the top. We show that the elasticity of the Pareto parameter with respect to the top tax rate *alone* provides a sufficient statistic for the characterization of revenue-maximizing top tax rates. The derivation of this result builds on our proposed "mobile response," which extends Saez's (2001) behavioral response to incorporating the extensive mobility of the top income group into and out of the top tax bracket.

Keywords: Mobile response, the elasticity of the Pareto parameter, revenue-maximizing top tax rates
JEL Classification: H21, H31

*We thank Fabian Kindermann, Emmanuel Saez, Thomas Sargent, Pengfei Wang, Ye Yuan, and seminar participants of the 2015 SAET conference held in Cambridge.

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, Saez (2001) derived a practical formula for the calculation of 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, α is the Pareto parameter of the income distribution at the top, and e is the elasticity of taxable income (ETI) in the top tax bracket with respect to the net-of-top-tax rate $(1-\tau)$.[1] Recently, formula (1) has become the focal point of studies on revenue-maximizing top tax rates, including Guner et al. (2014), Kindermann and Krueger (2014), and Badel and Huggett (2015, 2017).

Saez (2001) proposed the "behavioral response" to measure how the taxable incomes of those in the top income group have varied in response to the top tax rate. He emphasized that the ETI is microfounded with the behavioral response, building on the aggregation of the individual elasticity of taxable income in the top tax bracket. In their review of the literature on the ETI, Saez et al. (2012, p. 7) explicitly explained: "This aggregate elasticity is equal to the average of the individual elasticities weighted by individual

[1] Saez (2001) also addressed welfare-maximizing top tax rates. The formula is given by $\frac{1-\bar{g}}{1-\bar{g}+e\cdot\alpha}$ where $\bar{g}$ is defined such that (p. 210) "the government is indifferent between $\bar{g}$ more dollars of public funds and one more dollar consumed by the taxpayers with income above $\bar{z}$ [the threshold of the top income]." Note that welfare-maximizing top tax rates reduce to revenue-maximizing top tax rates if $\bar{g} = 0$.

income, so that individuals contribute to the aggregate elasticity in proportion to their incomes." Saez (2001) and Piketty and Saez (2013, Footnote 54) spelled out the exact relationship between the aggregate elasticity e and the individual elasticities.

Following the influential work of Feldstein (1995), a great majority of empirical studies employ panel data analysis to estimate the behavioral response; see Saez et al. (2012) for a review. However, income variations in response to tax changes may cause some individuals to move above or below the income threshold of the top tax bracket so as to alter the tax base at the top. The behavioral response tracks the change of the taxable incomes of those who are in the top income group. A major problem of the behavioral response is that it fails to account for the extensive mobility into and out of the top tax bracket. Indeed, in their derivations of the ETI, both Saez (2001, Footnote 7) and Saez et al. (2012, Footnote 7) explicitly assumed that the top earners' extensive mobility is negligible. We make use of the 1986 Tax Reform Act (TRA 1986) to conduct empirical tests and find that the impact of the tax reform on the top earners' extensive mobility is statistically significant.[2]

When applying formula (1), researchers also face the problem of choosing the "right" value for the Pareto parameter α. There are many values of α that can be chosen in time-series 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 vary widely during 1946-2012 in the United States, from well above 2 in the 1970s to close to 1.5 in the most recent years.[3]

[2] Slemrod (1992) reported substantial mobility into and out of the top 1 percent households based on data from the 1970s and 1980s. For example, between 28 and 40 percent of the top 1% households in the income distribution were new from one year to the next year over the 1980–1986 period. Slemrod (1992) did not explicitly identify the underlying reasons for the mobility. In this paper we provide evidences on the significance of the extensive mobility of the top income group in response to the change of top tax rates.

[3] 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.

In fact, in his numerical 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)

Starting from the work of Saez (2001), papers including Simula and Trannoy (2010), Diamond and Saez (2011), Piketty and Saez (2013), Piketty et al. (2014), Guner et al. (2014), Kindermann and Krueger (2014), Scheuer and Werning (2015), and Badel and Huggett (2015, 2017), have addressed optimal taxation on top earners from a variety of perspectives. In this paper we proposes the "mobile response" to replace the behavioral response. We derive a new formula for revenue-maximizing tax rates in the Mirrlees (1971) framework,

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

where

$$\zeta = \frac{1-\tau}{\frac{1}{\alpha-1}} \cdot \frac{\partial(\frac{1}{\alpha-1})}{\partial(1-\tau)}. \tag{3}$$

We call ζ the "elasticity of the Pareto parameter." The mobile response tracks how the income distribution at the top itself has altered in response to the change of top tax rates. The elasticity ζ reflects the mobile response. We demonstrate through examples that the mobile response will be identical to the behavioral response if variations in taxable incomes involve only "intensive mobility" in that no agents move into or out of the top tax bracket. However, once variations in taxable incomes involve "extensive mobility" in that there are agents moving into or out of the top tax bracket, the mobile response will generalize the behavioral response to account for the extensive mobility.

Measuring the mobile response could be a difficult task, since it requires tracking how the income distribution at the top itself has changed with respect to the top tax rate. Nevertheless, we show that if the income distribution at the top is asymptotically Pareto-distributed as practically observed, then, given the income threshold of the top tax bracket, the tax base at the top is completely characterized by the Pareto parameter α. This complete characterization paves a convenient way to measure our proposed mobile response. We demonstrate that the elasticity ζ accounts for the extensive as well as intensive mobility of the top income group so as to fully summarize the change of the top tax base in response to the change of τ. As such, unlike the well-known formula (1) that is characterized by two parameters, the single parameter ζ alone is a *sufficient statistic* for the characterization of the revenue-maximizing top tax rate τ^*.

We also develop a method to estimate the elasticity ζ. Our estimation is based on a relationship between ζ and the Pareto parameter α. In stead of using panel data, we use time-series data of α to estimate ζ. The main challenge of the estimation is the endogeneity problem. We empirically estimate ζ 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). 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.

The rest of the paper is organized as follows. In Section 2 we revisit the derivation of revenue-maximizing tax rates by using our proposed mobile response. Section 3 comments on the ETI at the top. Section 4 provides evidences on the significance of the extensive mobility at the top in response to the change of top tax rates. In Section 5 we estimate the elasticity of the Pareto parameter and numerically implements our new formula. Section 6 concludes the paper.

2 Derivation of revenue-maximizing top tax rates

We follow the approach 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 the intensity of work, form and timing of compensation, and tax avoidance/evasion.[4]

In deriving the formula for calculating top income optimal tax rates, 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 involved (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 a small perturbation at the optimum. We follow the same methodology but make some important refinements.

[4] 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 3. 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).

The key difference between our derived formula and Saez's (2001) stems from the revenue effect, not from the welfare effect. Thus, we shall confine our analysis to the revenue effect and focus on the derivation of revenue-maximizing rather than optimal top tax rates.

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 τT. 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 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." The mobile response generalizes Saez's behavioral response, in that it accounts for the "extensive" as well as the "intensive" mobility in the top income tax base.

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 is 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,

$$M \equiv (z_m - \bar{z})\, d\tau, \tag{4}$$

where $(z_m - \bar{z})$ is the tax base of the top tax bracket, and z_m denotes the mean of taxable incomes above the threshold $\bar{z}$ of the top tax bracket. Note that $T = z_m - \bar{z}$, due to the graduated-rate nature of the modern income tax system. By the definition of z_m we have

$$z_m = \frac{\int_{\bar{z}}^{\infty} z f(z) dz}{\int_{\bar{z}}^{\infty} f(z) dz} = \int_{\bar{z}}^{\infty} z h(z) dz, \tag{5}$$

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, equation (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

$$D \equiv \tau dz_m, \tag{6}$$

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 (5) serves as the weight; see his equation (7). He called this change the "behavioral response," which calculates changes in z above $\bar{z}$, 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[5]

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

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.

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.

Behavioral response: track how the taxable incomes of those in the rich club at the pre-change τ have changed, that is, the change in z above $\bar{z}$ 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), $dz_m = z_m(\tau') - z_m(\tau) = 200$.[6]

Mobile response: track 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.

When the members of the rich club remain unchanged as the top tax rate varies,

[5] For ease of exposition, we do not express $h(z)$ in terms of frequencies. Our examples are illustrative.

[6] 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 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: since z_2 is in the rich club at τ and his taxable income has decreased from 100 to 70 as the top tax rate is raised to τ'' (all else remains unchanged), $dz_m = z_m(\tau'') - z_m(\tau) = -30$.

Mobile response:

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''.

Since agent z_2 is no longer a member of the rich club at τ'', dz_m is equal to -100 rather than -30. As far as measuring changes in the top tax base is concerned, the mobile response for calculating dz_m 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: 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:

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 $dz_m = 100$. Agent z_1 is not a member of the rich club at τ, but has become a member of the rich club at τ'''. 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 their taxable incomes change accordingly. These variations 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 nonnegligible fraction of agents who experience mobility into or out of the rich club (Examples 2 and 3), it is better to measure dz_m by the change of $h(z)$ itself given $z \geq \bar{z}$ (the mobile response) rather than by the change of z above $\bar{z}$ given $h(z)$ (the behavioral response). At any rate, it is clear that our mobile response of measuring dz_m generalizes Saez's behavioral response by accounting for the "extensive" as well as the "intensive" income mobility in the rich club.

Measuring dz_m through the change of $h(z)$ given $z \geq \bar{z}$ could be a difficult task. We later come up with a way to overcome the difficulty.

2.1.3 Compared with Badel and Huggett (2017)

Badel and Huggett (2017) also studied the revenue-maximizing top tax rates. But they extended the formula (1) 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 (2017), 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 (2017) 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, it is not easy to empirically estimate the indirect effect. 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. We count in this indirect effect by the change in $h(z)$ itself or, more precisely, the change in the Pareto parameter α.

2.2 Revenue-maximizing top tax rates

The goal of the government is to maximize tax revenues collected from the top tax base. The revenue-maximizing top tax rate τ^* is determined by setting the sum of the mechanical effect (defined by (4)) and the mobile response (defined by (6)), to zero,

$$M + D = 0. \tag{7}$$

A key term in the above condition is dz_m in D. Our mobile response measures dz_m by tracking how $h(z)$ itself has changed, given $z \geq \bar{z}$.

We measure dz_m by using the relationship between z_m and the Pareto parameter α. If $f(z)$ is asymptotically Pareto-distributed as practically observed, then $f(z) \approx C/z^{\alpha+1}$ for $z \geq \bar{z}$, where C is a positive constant.[7] Thus we have

$$z_m - \bar{z} = \int_{\bar{z}}^{\infty} zh(z)dz - \bar{z} = \frac{\int_{\bar{z}}^{\infty} zf(z)dz}{\int_{\bar{z}}^{\infty} f(z)dz} - \bar{z} = \frac{1}{\alpha - 1}\bar{z}. \tag{8}$$

Once the threshold $\bar{z}$ is given, the top tax base $(z_m - \bar{z})$ is completely characterized by the Pareto parameter α. This result provides us a convenient way of measuring dz_m by changes in $h(z)$, given $z \geq \bar{z}$. Specifically, combining equations (8) and (6) we have

$$D = -\frac{\tau}{1-\tau}\zeta\frac{1}{\alpha - 1}\bar{z}d\tau, \tag{9}$$

where $\zeta = \frac{1-\tau}{\frac{1}{\alpha-1}} \cdot \frac{\partial\left(\frac{1}{\alpha-1}\right)}{\partial(1-\tau)}$.[8] Plugging equations (4) and (9) into equation (7) and using the

[7] Saez (2001) provided empirical evidences showing that (p. 211): "the tails of empirical earnings distributions can be remarkably well approximated by Pareto distributions."

[8] We can find an alternative interpretaton of ζ in equation (13).

relationship (8), we have the revenue-maximizing top tax rate formula,

$$\tau^* = \frac{1}{1+\zeta}. \tag{10}$$

Since ζ is completely characterized by changes in α, we call it the elasticity of the Pareto parameter.

To derive our formula (10) we do not need the assumption that the top earners' extensive mobility is negligible. From the expression (8) we know that the change of $\frac{1}{\alpha-1}$ exactly reflects the change of the top tax base. The mobile response D represents the change of the top tax base, irrespective of whether the mobility involved is intensive or extensive and regardless of who are in the top income group. In other words, the elasticity ζ itself is a sufficient statistic for summarizing the change of the top tax base with respect to that of the top tax rate.

The dependence of α upon τ as embodied in (3) is consistent with the evidence provided by Alvaredo et al. (2013). They plot 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). There is a strong negative correlation between the changes in top tax rates and the evolution of top income shares.

3 The ETI at the top

As noted in the Introduction, the literature has focused on the ETI at the top, the elasticity e in formula (1). We address the relationship between e and our proposed elasticity of the Pareto parameter ζ in this section.

3.1 The microfounded ETI

Saez (2001) emphasized that the ETI at the top is microfounded with the behavioral response, building on the aggregation of different individuals' elasticities of taxable income in the top tax bracket. Saez et al. (2012) critically reviewed the literature on the ETI. They defined the aggregate elasticity of taxable income in the top bracket with respect to the net-of-top-tax rate $(1-\tau)$ as

$$e \equiv \frac{1-\tau}{z_m} \cdot \frac{\partial z_m}{\partial(1-\tau)}. \tag{11}$$

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]

$$e = e^u - \eta \cdot \frac{\bar{z}}{z_m}, \tag{12}$$

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] and $\eta \equiv \int_{\bar{z}}^{\infty} \eta_{(z)} h(z) dz$ is the average of the individual income effect $\eta_{(z)}$.[11] Saez (2001) emphasized that the aggregates e^u and η in (12) are microfounded,

[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 for 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

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

that is, they are aggregations of individual $e_{(z)}$ and $\eta_{(z)}$ respectively, with the suitable weights. Both Saez (2001, Footnote 7) and Saez et al. (2012, Footnote 7) explicitly assumed that the top earners' extensive mobility is negligible. Note in particular that the elasticity e in (12) is measured by the change of z above $\bar{z}$ given $h(z)$ (Saez's behavioral response) rather than the change of $h(z)$ given $z \geq \bar{z}$ (our mobile response). Thus, a key difference between the ETI e and the elasticity of the Pareto parameter ζ lies in whether people's extensive mobility into and out of the top tax bracket is accounted for. We provide evidences on the significance of such extensive mobility in response to the change of top tax rates in Section 4.

To empirically estimate the behavioral response (changes of z above $\bar{z}$ given $h(z)$) as suggested by Saez (2001), panel data are required since the response tracks how the taxable incomes of those in the rich club have changed with respect to the top tax rate. Following the influential work of Feldstein (1995), a great majority of empirical studies did employ panel analysis to estimate the behavioral response. However, our examples in Section 2 suggest that, as far as estimating the change of the top tax base is concerned, panel data may not be so useful if there exists significant extensive mobility in the top tax bracket.

3.2 The Pareto parameter and the ETI at the top

We now establish an exact relationship between the ETI at the top e and the elasticity of the Pareto parameter ζ when the top earners' extensive mobility is negligible.

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

From equations (3) and (8) we have

$$\zeta = \frac{1-\tau}{z_m - \bar{z}} \cdot \frac{\partial(z_m - \bar{z})}{\partial(1-\tau)}. \tag{13}$$

Thus ζ represents the elasticity of the top tax base $(z_m - \bar{z})$ with respect to the net-of-top-tax rate $(1-\tau)$. We assume that the top earners' extensive mobility is negligible. Combining equations (8), (11), and (13), we have

$$\zeta = e \cdot \frac{z_m}{z_m - \bar{z}} = e \cdot \alpha. \tag{14}$$

It seems that the extant literature fails to recognize this relationship, even though it is intuitive. Equation (13) implies 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 (11) implies that a 1% change in $(1-\tau)$ causes a e% change in the *top income* z_m. The term $\frac{z_m}{z_m-\bar{z}}$ in equality (14) adjusts the percentage-change base.

When the top earners' extensive mobility is negligible, we have the relationship (14). Then our formula (10) reduces to the well-known formula (1),

$$\tau^* = \frac{1}{1+\zeta} = \frac{1}{1+e\cdot\alpha}.$$

4 Evidences on extensive mobility

We utilize the 1986 Tax Reform Act (TRA 1986) in the United States as an illustration to empirically show the impact of a tax change on the extensive mobility of top earners. It should be noted that the extensive mobility that we emphasize is different from the mean reversion bias that has been emphasized in the literature such as in Auten and Carroll (1999), Gruber and Saez (2002), and Saez et al. (2012). The main difference is that extensive mobility is caused by tax reforms while mean reversion is caused by large transitory income instead of tax reforms. We focus on the case of the 1986 Tax Reform Act (TRA 1986) and utilize the University of Michigan Tax Panel to estimate the 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 1985 and 1988 panel data for the investigation.[12]

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 and the other for those moving out of the top 1% in 1988.[13] 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

[12] 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.

[13] 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 and 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 calculate the synthetic net-of-tax price for 1989 by replacing the actual tax rate in 1989 with its instrument. The instrument for the actual tax rate in 1989 is constructed by computing a taxpayer's tax rate under 1989 tax law using 1985 income inflated to 1989 levels. 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 identifying the effect of taxes on mobility. We therefore also include the income in the initial year as a control variable to isolate mean reversion (Auten and Carroll, 1999). 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 2), 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.[14]

The coefficient estimates for agents' mobility up to the very top income groups (i.e., the 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 that TRA 1986 is generally associated with a reduction rather than an increase in the tax rates for high-income taxpayers.

[14] Since transition analyses are based on only a sample of agents that 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 have tried including the inverse Mills ratio in the transition probit regressions, but do not find evidence of a significant influence on the estimates.

5 Estimating the elasticity of the Pareto parameter

On the basis of equation (3), let us consider a basic time-series regression of the following form for estimating the elasticity ζ,

$$\log(\frac{1}{\alpha_t - 1}) = b + \zeta \log(1 - \tau_t) + \varepsilon_t, \tag{15}$$

where b is some constant. Since $(1 - \tau_t)$ is likely to be correlated with the error term ε_t in the regression (15), 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 the ETI is different from our ζ, its estimation faces a similar problem as in our regression (15): 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.[15]

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

[15] Weber (2014) and Burns and Ziliak (2017) recently proposed new methods to estimate the ETI. The former paper particularly showed that most of the existing instruments used in difference-in-differences are not exogenous.

(15), we have

$$\log(\frac{1}{\alpha_{t-1}-1}) = b + \zeta \log(1-\tau_{t-1}) + \varepsilon_{t-1}. \tag{16}$$

equation (15) minus equation (16) leads to

$$\log(1-\frac{\Delta\alpha_t}{\alpha_t-1}) = \zeta \log(1-\frac{\Delta\tau_t}{1-\tau_{t-1}}) + v_t, \tag{17}$$

.

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 equation (7) in Romer and Romer (2010), we estimate

$$\log(1-\frac{\Delta\alpha_t}{\alpha_t-1}) = \zeta \log(1-\frac{\Delta\tau_t}{1-\tau_{t-1}}) + \sum_{j=1}^{M} \zeta_j \log(1-\frac{\Delta\tau_{t-j}}{1-\tau_{t-1-j}}) + \sum_{j=1}^{N} \beta_j \log(1-\frac{\Delta\alpha_{t-j}}{\alpha_{t-j}-1}) + v_t, \tag{18}$$

where the lagged terms of $\Delta\tau$ and $\Delta\alpha$ are added to (17) 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 (18).[16]

To carry out the estimation of (18), 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).[17]

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

[16] 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).

[17] 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 5. 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).

substantially cut the number of tax brackets.[18] Based on the methodology of Barro 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 (18) 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 (18), 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.[19] Similar to the arguments put

[18] 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.

[19] 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.

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 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 (18). 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 equation (13) 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%.

5.1 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 equation (18) is in the first-difference form, we thus re-run equation (18) by including two dummies to represent changes in the party affiliation of presidents.[20] 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

[20] 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.

for linear, square and cubic time trends in the estimations and find that the estimates of 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 equation (18) 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 equation (18) to evaluate potential bias utilizing all statutory tax changes for instruments.[21] We find that the estimates utilizing the instruments of all statutory tax changes are substantially larger

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

than those in Table 2 for the specifications without controlling for party affiliations and 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.

5.2 Quantitative revenue-maximizing top tax rates

After obtaining the estimates of the elasticity ζ, it becomes straightforward to quantitatively characterize top income optimal tax rates. 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 revenue-maximizing top tax rate in Table 3.

(Insert Table 3 about here)

We next apply formula (1), which is widely used in the extant literature. 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 rate in Table 4 (0.62) are not far away from that in Table 3 (0.55). 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 (1) in empirical studies is that there are so many values of α that can be chosen according to the real-world data.

6 Conclusion

We contribute to the optimal taxation literature by proposing the mobile response to replace Saez's (2001) behavioral response. While the behavioral response tracks how the taxable incomes of those in the top income group have varied in response to changes in top tax rates, the mobile response tracks how the income distribution at the top itself has altered in response. It is shown that the mobile response will be identical to the behavioral response if income variations involve only the intensive mobility of the top income group; however, once income variations involve the extensive mobility of the top income group as well, the mobile response will generalize the behavioral response to account for the extensive mobility.

Measuring the mobile response could be a difficult task, since it requires tracking how the income distribution at the top itself has changed with respect to the top tax rate. Nevertheless, we show that if the income distribution at the top is asymptotically Pareto-distributed as practically observed, then the elasticity of the Pareto parameter can be used to measure the mobile response and, importantly, it is a sufficient statistic for summarizing changes in the top tax base.

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., 2015. Taxing top earners: a human capital perspective. Mimeo.

Badel, A., Huggett, M., 2017. The sufficient statistic approach: predicting the top of the Laffer curve. Journal of Monetary Economics 87, 1-12.

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

Burns, S., Ziliak, J., 2017. Identifying the elasticity of taxable income. Economic Journal 127, 297-329.

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

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%? Lessons 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.

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.

Simula, L., Trannoy, A., 2010. Optimal income tax under the threat of migration by top-income earners. Journal of Public Economics 94, 163-173.

Slemrod, J., 1992. Taxation and Inequality: A Time-Exposure Perspective. Tax Policy and the Economy 6, 105-127.

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 rate

Income group	Top 1%
$\tau^* = \frac{1}{1+\zeta}$	55

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

Pareto parameter	1.5	2	2.5
$\tau^* = \frac{1}{1+e\cdot\alpha}$	73	67	62

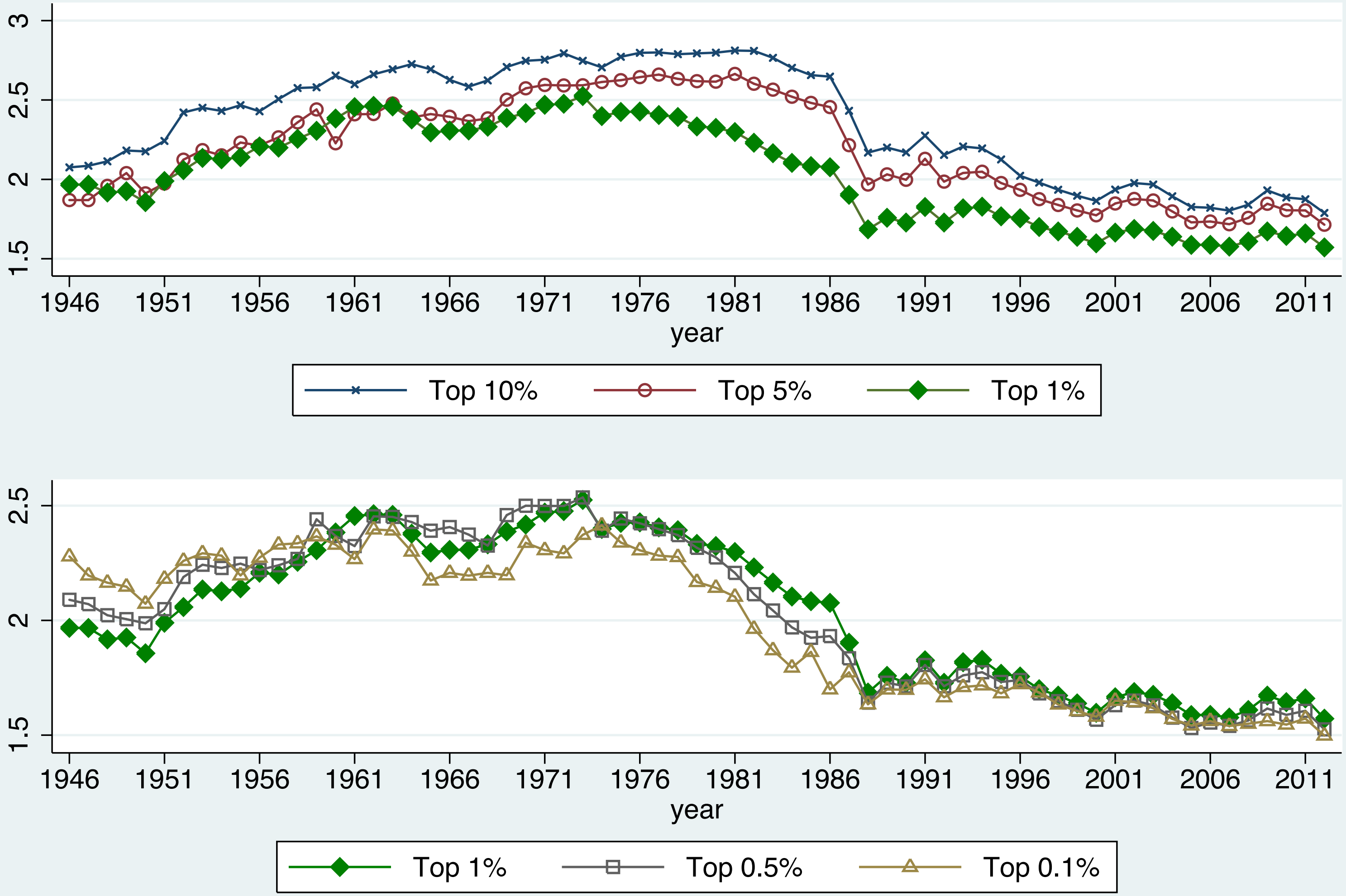

Figure 1. Time–series patterns of the Pareto parameter in the U.S. during 1946–2012