An Estimable Model of Income Redistribution in a Federation

Author(s): Kevin Milligan and Michael Smart

Source: *American Economic Journal: Economic Policy* , February 2019, Vol. 11, No. 1 (February 2019), pp. 406-434

Published by: American Economic Association

Stable URL: https://www.jstor.org/stable/10.2307/26641360

An Estimable Model of Income Redistribution in a Federation: Musgrave Meets Oates[†]

By Kevin Milligan and Michael Smart*

We develop a theory of cross-border income shifting in response to subnational personal taxation in a federation and examine its implications for the excess burden of personal taxes. We show how a properly chosen federal tax rate can offset the fiscal externality between states and facilitate decentralization, even in a heterogeneous federation where unitary taxation is suboptimal. Optimal taxes depend on the elasticities of national tax avoidance and of cross-state tax base shifting. We estimate these elasticities around a tax decentralization reform in Canada, finding both to be empirically relevant. We discuss the implications for optimal federalism. (JEL D31, H21, H23, H24, H26, H71, H77)

According to the received wisdom in public finance since Musgrave (1959), national governments should be assigned responsibility for redistribution of income. Decentralized income taxation can lead to inefficient tax base mobility, and this in turn can lead to tax competition among governments, which induces lower levels of redistribution than is optimal from a national perspective. In contrast, Oates (1972) emphasized the benefits of decentralization on the spending side of the budget. In the various states of a federation,[1] citizens have different tastes and capacities for public goods provision. Because a national government is generally constrained for political reasons to offer uniform policies for all states, decentralization of spending is preferred as long as interstate spillovers are small.

In this paper, we apply an Oatesian perspective to the (Musgravian) issue of redistribution through taxation. Just as on the spending side of the budget, regions have different tastes and capacities for redistribution—which favors decentralized redistribution. However, taxpayers may shift income between states of the federation in response to internal tax differentials, which favors national taxation. To analyze these competing effects, we develop a simple model of top-bracket income taxation where the tax base is mobile among states of a federation. We use the model to derive

*Milligan: Vancouver School of Economics, University of British Columbia, 6000 Iona Drive, Vancouver BC V6T 1L4 Canada (email: kevin.milligan@ubc.ca); Smart: Department of Economics, University of Toronto, 150 St. George Street, Toronto ON M5S 2G7 Canada (email: msmart@chass.utoronto.ca). Thanks for comments from Kory Kroft, Michael Devereux, members of the Oxford Centre for Business Taxation, and participants at the 2016 PEUK conference and the 2017 Canadian Economic Association meetings.
[1] In this paper, we mostly refer to subnational jurisdictions as "states" to keep the case general, but use the term "provinces" when referring to our specific empirical exercise involving Canada.

expressions for optimal tax rates as functions of estimable "sufficient statistics," we analyze alternative systems of tax assignment in the federation, and we estimate the model using data observed around a decentralizing reform in Canada.

We begin in Section I with a formal model of tax avoidance and tax shifting in a federation. Economists since Feldstein (1999) have understood that the elasticity of taxable income (ETI) is informative about the marginal excess burden of taxation. However, Gordon and Slemrod (1998) and Chetty (2009) have emphasized that when the ETI reflects shifting of income between two revenue sources, rather than simple tax avoidance behavior, then its implications for social welfare are more complicated. Such considerations obviously apply in our model, where some portion of the total ETI from a single state's perspective reflects shifting of income to other states. We show that the ETI may be decomposed into elasticities of pure avoidance and interstate income shifting, which are joint sufficient statistics in our model. We subject this framework of avoidance and shifting to an empirical test and explore its implications for an optimal tax setting within a federation.

Our framework at first suggests a Musgravian perspective on fiscal federalism. With tax base shifting, a tax increase in one state increases revenue in other states. This is a positive fiscal externality among state governments that favors national redistribution. However, consistent with the Oatesian perspective, differences in tax yield (or revenue-raising potential) of top-bracket taxes require differentiated tax rates. Tax yield differences are large in our data, and so this favors decentralization over uniform national taxation.

We analyze these trade-offs in a formal model. As benchmarks, we consider a fully decentralized model in which states set their own tax rates and a unitary national system in which a federal government sets one tax rate for the nation. We show that neither a fully decentralized nor a unitary national tax system can achieve the national welfare optimum. However, we find that a federal system can achieve the national welfare optimum, with shared taxation of the base by federal and state governments. Shared taxation creates a negative vertical externality between state and federal revenues, as a rate increase by one level of government shrinks the tax base available to the other level of government. This negative externality can offset the positive externality arising from horizontal tax shifting. As we show in our model, federal taxes may therefore be set to balance these negative and positive externalities in order to decentralize optimal tax-setting behavior to the state level, even in the presence of asymmetries among states.

Our results on the optimality of federalism are new, but our paper is not the first to analyze some of these issues. Keen and Kotsogiannis (2002) provided an early theoretical study of shared federal-state taxation of a mobile capital base. Like us, they studied the competing influences of horizontal and vertical externalities among state and federal governments to ask whether federal taxation tends to lead to tax rates that are too low or too high from a national perspective. Gordon and Cullen (2011) examine similar questions to ours in a different model of subnational taxation. However, they do not estimate tax elasticities, and they focus their analysis on the case of symmetric jurisdictions in which the "Oatesian" perspective that we emphasize does not arise. A distinct literature on the optimal size of nations following Alesina and Spolaore (1997) emphasizes the trade-off between horizontal externalities and

diverse populations. This trade-off is also present in our model, although we explore how federalism provides solutions rather than country formation and breakup.

Much of the previous literature deals with a symmetric model of states in which each state faces the same per capita tax base and sets the same tax rate in equilibrium. In contrast, we focus on tax base asymmetries among states in which the "Oatesian" trade-off between uniform centralized and heterogeneous decentralized policies comes to the fore. As well, our model is a normative one, establishing conditions under which federal policies are efficient, whereas the previous literature takes a more positive focus. Finally, our work takes a sufficient statistics approach in which we are able to estimate the relevant policy parameters in a model that is internally consistent and appropriate for the welfare analysis that we conduct.

Sections III and IV present our empirical application in which we estimate avoidance and shifting elasticities using data on top income shares and tax rates in Canadian provinces for the 1988–2013 period. Our data straddle the date of a federal reform that decentralized tax powers to subnational governments, and which led to substantial reductions in tax progressivity in some provinces—especially those with substantial nontax revenues—but not in others. We exploit this variation in order to estimate the shifting and avoidance components of the aggregate ETI. Our estimates suggest that interstate income shifting is large, with our point estimates suggesting about three-quarters of the elasticity for top taxpayers is accomplished through shifting to low-tax provinces, rather than other forms of avoidance.

The relatively small previous literature on cross-border effects of personal income taxation has generally examined specific mechanisms for tax shifting, rather than its aggregate effects on the ETI. Much of the literature has looked at tax-induced migration, building on the theoretical foundation set by Epple and Romer (1991).[2] In this vein, Kleven, Landais, and Saez (2013); Kleven et al. (2014); and Akcigit, Baslandze, and Stantcheva (2016) find substantial effects of high-income tax rates on migration of "superstars" in Europe. In the United States, Moretti and Wilson (2017) also find migration effects for star scientists, although earlier work by Young and Varner (2011) finds little response from the general high-income population to a particular state tax change. More recent work by Young et al. (2016) reinforces the evidence that tax-induced migration among high earners within the United States is small. Taken together, these results complement and reinforce the earlier conclusions of Feldstein and Wrobel (1998) that state income tax differences lead to offsetting differences in equilibrium pretax wages, leaving little scope for redistribution at the subnational level. A parallel literature looks for evidence of mobility of financial assets in response to personal tax differences. The recent European Savings Directive in particular appears to have substantially reduced the use of offshore bank accounts at least in some EU member countries (Johannesen 2014). A few papers have received relatively little attention in the Canadian empirical literature. Mintz and Smart (2004) document the potential for shifting corporate income between provinces and estimate tax base elasticities, but they do not consider personal income shifting. Saez and Veall (2005) estimate the ETI at the

---

[2] See also the optimal tax analysis of tax-induced migration in Lehmann, Simula, and Trannoy (2014).

national level using data similar to ours, but ignore provincial variation in tax rates. Veall (2012) documents the recent changes in top income shares at the provincial but does not estimate the effects of provincial taxes. Milligan and Smart (2015) estimate the ETI for Canada, but they do not look at subnational income shifting.

I. A Theory of Tax Avoidance and Tax Shifting in a Federal System

Consider a federation consisting of J states. Each state levels its own tax rate t_i on incomes above a top-bracket threshold k_i, while the federal government levies a common national tax rate T on the same base; the combined effective tax rate is $\tau_i = t_i + T$. Taxable income in the top income tax bracket in each state i is a function[3]

$$
(1) \qquad y_i(\tau_1, \ldots, y_J) = y_i(\tau_i, \tau_i - \bar{\tau}),
$$

where

$$
(2) \qquad \bar{\tau}(\tau_1, \ldots, \tau_J) = \sum_j \omega_j(\tau_1, \ldots, \tau_J) \tau_j
$$

is the average tax rate for all states in the nation, weighted by income shares ω_i, i.e.,

$$
\omega_i(\tau_1, \ldots, \tau_J) = \frac{y_i(\tau_1, \ldots, \tau_J)}{\sum_j y_j(\tau_1, \ldots, \tau_J)}.
$$

Together, equations (1)–(2) implicitly define the response of incomes to tax rates in the federation. In equation (1), the first argument measures how a state's tax base responds to a change in its own tax rate, holding interstate rate differentials constant. We call this the effect of taxes on *national avoidance behavior* and define the corresponding avoidance semi-elasticity:

$$
e_a = -\left.\frac{\partial \log y_i}{\partial \tau_i}\right|_{\tau_i - \bar{\tau} \text{ fixed}} = -\frac{y_{i1}}{y_i}.
$$

The second argument measures how the tax base in i changes in response to an increase in the tax rate differential between the home state and other states, holding the home rate fixed. We call this the effect of taxes on *cross-state shifting behavior* and define the corresponding shifting semi-elasticity:

$$
e_s = \left.\frac{\partial \log y_i}{\partial \bar{\tau}}\right|_{\tau_i \text{ fixed}} = -\frac{y_{i2}}{y_i},
$$

[3] This reduced-form representation of the tax base as a function of tax rates and differentials is similar to that used by Buettner (2003).

which we assume to be common for all states i.

The assumption that tax bases depend only on own marginal tax rates and national average tax differentials is of course restrictive, but it is a common assumption in the tax-competition literature.[4] Appendix A to this paper shows how such tax base functions can be derived from an optimizing model of individual taxpayer behavior when taxpayers may use costly avoidance devices to hide income from state and federal tax authorities or to shift income between states of the federation.[5] Our model could in principle be generalized, e.g., to one featuring asymmetric state tax-competition interactions within the federation.

In analyzing optimal tax policies later in the paper, we assume that governments maximize their respective revenues from income declared in the top bracket.[6] Recall that the federal government levies a uniform tax rate T on the top bracket in all states, whereas state governments (in the case of decentralization) levy heterogeneous tax rates $t_i = \tau_i - T$. So state government tax revenues are

$$(3) \qquad R_i(\tau_1, \ldots, \tau_J, T) = (\tau_i - T)[y_i(\tau_i, \tau_i - \bar{\tau}) - k_i], \quad i = 1, \ldots, J,$$

and federal tax revenues are

$$(4) \qquad R^F(\tau_1, \ldots, \tau_J, T) = T \sum_j [y_j(\tau_j, \tau_j - \bar{\tau}) - k_j].$$

Summing these expressions, national tax revenues are

$$(5) \qquad R^N(\tau_1, \ldots, \tau_J) = \sum_j \tau_j [y_j(\tau_j, \tau_j - \bar{\tau}) - k_j].$$

A. *Elasticities and Optimal Tax Policies*

Understanding the magnitudes of the pure avoidance and shifting responses is the key to evaluating the impact of tax rate changes on the revenues of federal and state governments and on the excess burden of the tax system. Our theory allows us to characterize optimal tax rates from a state and national perspective in terms of the estimable semi-elasticities e_a and e_s and a measure of income inequality that affects tax yields and which may vary among states. That is, we show that (e_a, e_s) are "sufficient statistics" for welfare analysis (Chetty 2009) in this model. Using this fact, we will proceed to contrast what can be achieved in a (Musgravian) unitary tax system, compared to an (Oatesian) equilibrium decentralized system.

---

[4] This approach was, for example, used in a two-country model by Keen (2001). Bucovetsky and Haufler (2007) uses a similar model to study taxation by two countries that differ in size. We extend the model to more general forms of heterogeneity and more than two jurisdictions.

[5] Because tax base functions depend on marginal tax rates, our focus is on the shifting of marginal financial income, rather than discrete changes in taxpayer residence.

[6] The assumption that governments maximize revenue rather than a broader notion of welfare is restrictive, but it may be most appropriate for the case of top-bracket taxation if the marginal utility of consumption for high-income taxpayers is sufficiently small. See, e.g., the discussion in Diamond and Saez (2011).

To that end, we study a simple extensive form game in which a federal government first chooses a common federal tax rate T applying to the top bracket in all states. We consider two possibilities:

(i) a unitary system in which there is no state taxation, so that $\tau_i = T$ in all states i; and

(ii) a federal system in which state governments observe T and simultaneously choose tax rates t_i, so that the combined top-bracket tax rates are $\tau_i = t_i + T$.

To understand the source of potential inefficiency in the state tax setting, consider the marginal revenue of a tax increase in state i from the state and national perspectives. As usual (Saez 2004), this will have a "mechanical" effect that is proportional to the inverse Pareto parameter measuring the inequality in the distribution of top incomes in state i, i.e.,

$$\theta_i = \frac{y_i - k_i}{y_i},$$

and to behavioral effects of taxes on reported incomes. Differentiating equations (1), (2), and (5), we can establish the following result.

PROPOSITION 1: *The marginal impacts of a unilateral tax increase in state i on state and national revenues are*

$$(6) \qquad \frac{\partial R_i / \partial \tau_i}{y_i} = \theta_i - \left[e_a + (1 - \omega_i) e_s \right] (\tau_i - T),$$

$$(7) \qquad \frac{\partial R^N / \partial \tau_i}{y_i} = \theta_i - e_a \tau_i - e_s (\tau_i - \bar{\tau}).$$

PROOF:
See Appendix B.

In both expressions, we have the familiar decomposition of marginal revenue into the mechanical and behavioral effects of a tax increase, with the mechanical affect proportional to the inverse Pareto parameter and the behavioral effect depending on estimable elasticities. But behavioral effects differ from the state and national perspectives, so that policies that are optimal from the state perspective are suboptimal nationally and conversely. Subtracting equation (6) from equation (7), state and national effects of tax increases differ by

$$(8) \qquad \frac{\partial R^N / \partial \tau_i - \partial R / \partial \tau_i}{y_i} = \underbrace{e_s \sum_{j \neq i} \omega_j (\tau_j - T)}_{\text{horizontal externality}} + \underbrace{(-e_a T)}_{\text{vertical externality}}.$$

In this expression, the first term is the (positive) horizontal fiscal externality—the tendency for state-level tax increases to increase average revenues of other states. The second term is the (negative) vertical fiscal externality—the tendency for state-level tax increases to decrease federal revenues. While the horizontal externality tends to result in states choosing tax rates that are too low from a national perspective, the vertical externality serves as a corrective, raising equilibrium tax rates.[7] It is this interplay between horizontal and vertical externalities that is the key to our results on optimal tax assignment.

We can characterize optimal tax rates as follows.

PROPOSITION 2: *The tax rates* τ_i^* *that maximize national revenues are*

$$(9) \qquad \tau_i^* \; = \; \frac{\theta_i}{e_a} - \frac{\theta_i - \bar{\theta}}{e_a + e_s}\frac{e_s}{e_a}, \quad i \; = \; 1,\ldots,J,$$

where

$$\bar{\theta} \; = \; \sum_j \omega_j\big(\tau_1^*,\ldots,\tau_J^*\big)\theta_j$$

is the weighted average of the individual state-yield parameters.

PROOF:
This is immediate from setting $\partial R^N / \partial \tau_i \; = \; 0$ and using equation (7). ∎

Observe that equation (9) implies

$$\tau_i^* \; > \; \tau_j^* \; \Leftrightarrow \; \theta_i \; > \; \theta_j,$$

i.e., optimal tax rates are heterogeneous among states whenever state yields differ due to inequality differences. The optimal tax formula is an inverse elasticity rule that reflects the competing needs to differentiate tax rates among states to reflect local conditions and to limit tax differences in order to control interstate shifting incentives. The first term in equation (9) is the inverse elasticity rule that would apply for the state if there were no interstate shifting. The second term is an adjustment factor that decreases the tax rate in high-yield states (increases in low-yield states), relative to this simple heuristic, to offset the shifting pressures. Observe that this "shifting adjustment" is larger, so that optimal tax rate differentials are smaller when the share of shifting e_s in the total tax base elasticity $e_a + e_s$ is larger.

B. *Optimal Tax Assignment in the Federation*

Proposition 2 characterizes tax policies that maximize national revenues. The problem is that a national government attempting to implement these policies

---

[7] This notwithstanding, federal and state tax rates are not strategic complements in general. See Keen (1998).

would violate the Oatesian constraint that national policies be uniform across the nation. Respecting the Oatesian uniform-policy constraint, what can be achieved under alternative assignments of tax powers in a federal system?

If taxation were fully decentralized to the states, then $T = 0$ and states would choose tax rates to maximize state tax revenues. Setting marginal state revenue to zero in equation (6), it follows that Nash equilibrium tax rates under full decentralization satisfy

$$(10) \qquad \tau_i^D = \frac{\theta_i}{e_a + e_s}.$$

This is again an inverse elasticity rule, reflecting that the response to a unilateral tax increase from the state's own perspective is proportional to the sum of the avoidance and shifting semi-elasticities. But from the national perspective, the shifting effect is (largely)[8] a transfer from one state treasury to another, which does not affect national revenues. This is the horizontal externality from interstate tax competition. In this sense, state taxation taken in isolation tends to result in tax competition and equilibrium tax rates that are too low from a national perspective. Comparing equations (9)– (10), we see that

$$\tau_i^* - \tau_i^D = \frac{\bar{\tau}}{e_a} \frac{e_s}{e_a + e_s} \geq 0,$$

that is, full decentralization with $T = 0$ results in Nash equilibrium tax rates in all states that are lower than the levels that maximize national revenues.

Unitary Taxation.—Under unitary taxation, a national government sets a uniform tax rate in all states to maximize national revenues. This is the Oatesian perspective on what can be achieved under centralization.[9] With unitary taxation, state tax differentials are zero, there is no cross-state shifting, and the single tax rate τ is chosen to maximize

$$(11) \qquad \max \sum_j \tau \left[y_j(\tau, 0) - k_j \right].$$

Differentiating equation (11) immediately establishes the following result.

PROPOSTITION 3: *The optimal tax rate τ^U in the unitary case can be expressed as the inverse elasticity rule*

$$(12) \qquad \tau^U = \frac{\bar{\theta}}{e_a}.$$

---

[8] It is not a pure transfer between treasuries, except in the case where all state tax rates are uniform, so that revenue losses by one state are exactly offset by revenue gains in others.

[9] In this case, for simplicity, we imagine that the national tax system adopts the same bracket thresholds k_i, but tax rates are constrained to be uniform across the country.

Recall, the state-yield parameters θ_j are the inverse Pareto parameters that measure the degree of inequality of top incomes in each state. The revenue yield of the tax will depend on how much income is above the top tax bracket threshold, which is determined by the state's Pareto parameter.

Unitary taxation "solves" the tax-competition problem in the sense that the tax rate is set in response to (national) avoidance responses, but not to interstate shifting. But comparison of equations (9) and (12) shows that the unitary tax policy is suboptimal whenever tax yield parameters θ_j differ among states.[10] In particular, the unitary tax rate is the optimal tax rate for a state with the average degree of inequality $\bar{\theta}$, but too high from the perspective of a low-θ state, and too low from the perspective of a high-θ state.

Optimal Federalism.—Neither full centralization nor full decentralization achieves the optimum in equation (9). Can we do better? Consider instead a federal system in which there is a uniform federal tax rate consistent with the unitary model but also decentralized state tax rates applied to the same base. The federal government taxes in this model can be said to "piggyback" on the state taxes since they share the same tax base.

Consider a two-stage "Stackelberg" game in which the federal government first chooses T.[11] Each state government observes T and simultaneously chooses its own tax rate t_i to maximize state revenue. Formally then, in the federal system of taxation, state tax rates are functions

$$(\tau_1^*(T), \ldots, \tau_J^*(T))$$

that represent a fixed point of the state best-response functions.[12] Setting marginal revenue to zero in equation (6), the Nash equilibrium tax rates in the sub-game satisfy

$$(13) \qquad \tau_i^* - T = \frac{\theta_i}{e_a + (1 - \omega_i)e_s}.$$

Given the equilibrium state tax rates, what can be achieved through the federal tax setting? Tax rates chosen in the decentralized case tend to be lower than optimal because of the horizontal externality on revenues of other states (when $e_s > 0$), but higher than optimal because of the vertical externality on federal revenues (when $T > 0$). By setting the federal tax correctly and redistributing the revenues,

---

[10] This echoes the trade-off between horizontal externalities and diversity in Alesina and Spolaore (1997).

[11] See Keen (1998) for analysis of the alternative "Nash" assumption that federal and state governments act simultaneously, so that the federal government cannot commit to undo the horizontal externality through its own tax policies.

[12] For brevity, we set aside issues of existence and uniqueness of the Nash equilibrium in our formal analysis. Note, however, that state tax rates are strategic complements, and that state objective functions are single crossing in own tax rates and the federal tax rate. Therefore, we can apply the results of Milgrom and Roberts (1990) to establish that there exist largest and smallest Nash equilibria of the tax-competition sub-game that are increasing in the federal tax rate.

the federal government may therefore be able to implement the national optimum through decentralized taxation.

A federal authority seeking to maximize national revenues chooses T to maximize $R^N(\tau_1, \ldots, \tau_J)$ subject to equation (13). Proposition 1 showed that the federal tax rate serves to mitigate the horizontal externality of tax competition by introducing an offsetting vertical externality. In fact, when the number of states in the federation grows large, we can establish that the national optimum tax vector is in fact implementable through a shared tax system with a uniform federal rate. In particular, the following result holds.

PROPOSITION 4: *If the federal tax rate is*

$$(14) \qquad \hat{T} = \frac{\bar{\theta}}{e_a} \frac{e_s}{e_a + e_s},$$

then as $\omega_i \to 0$, the Nash equilibrium tax rates approach the national optimum tax rates.

In this model, a uniform federal tax rate in the nation, if set optimally, acts as a Pigouvian subsidy to state tax increases and gets the incentives right for each state, regardless of differences in state tax yields. So the national optimal tax system is implementable in the federation with piggybacking, even in the presence of the Oatesian constraint that federal policies be uniform throughout the nation.

Our result on the optimality of federalism is asymptotic in the sense that it holds only as the weights $\omega_i \to 0$. For finite states with $\omega_i > 0$, equation (14) shows that large states internalize more of the shifting externality, choosing lower tax rates ceteris paribus.[13] There is therefore some diversity in state best-response functions that cannot be corrected through a uniform federal tax rate.

As a corollary, the model also gives a simple heuristic to determine the optimal "vertical fiscal gap," i.e., the optimal share of the federal government in tax revenues, which may exceed its share in national government spending (cf. Keen 1998). At the national optimum, the average tax rate is $\bar{\tau}^N = \bar{\theta}/e_a$. Comparing equation (14), we see have the following result.

COROLLARY 1: *If the share of the federal taxes in total taxes is*

$$\frac{\hat{T}}{\bar{\tau}} = \frac{e_s}{e_a + e_s}$$

in the average state, then as $\omega_i \to 0$, the Nash equilibrium tax rates approach the national optimum tax rates.

---

[13] This echoes the result in Bucovetsky and Wilson that small-population states set lower tax rates in equilibrium and so "win" the tax-competition game in per capita terms.

Of course, the optimal tax rate is lower in a low-yield state $\theta_i < \bar{\theta}$ (and conversely), leading to a larger vertical gap there, which can be offset through greater-than-average per capita transfers to citizens of i.

II. The Case of Personal Taxation in Canada

We study the case of Canada to explore the implications of our model with the aim of providing estimates of the key avoidance and shifting elasticity parameters from the model. While falling short of a full empirical test of the model, this evidence can provide support for the empirical relevance and magnitudes of the avoidance and shifting mechanisms at the heart of the model. In this section, we describe the relevant institutional details of personal income taxation in Canada necessary to motivate our empirical strategy presented in the next section.

Canada is a federation in which income taxation powers are co-occupied by the federal government and the governments of the ten provinces. Constitutionally, the provinces have wide latitude in designing their personal income tax systems, and they collect substantial revenue from them. In 2016, provincial personal income tax revenues were \$101 billion or 42 percent of combined federal–provincial revenues.[14] Several Canadian provinces derive a significant portion of their budgets from nontax resource revenues, which relieves pressure on the personal income tax system.[15] For example, in the 1988–2013 time period included in our empirical analysis, the share of total provincial revenues contributed by natural resource levies for Alberta averages 30 percent, while 6 provinces have natural resource shares under 2.5 percent. This interplay between provincial resource revenue and personal taxes will form part of our empirical strategy described in the next section.

Provinces generally apply progressive rate structures to taxable incomes with top marginal tax rates in 2016 ranging from 10 percent to 21.0 percent in the various provinces. During our 1988–2013 sample period, tax rates are applied to a common (federal) definition of taxable income and collected on behalf of provinces by federal tax authorities in all provinces except Quebec.[16] Provincial taxes are not deductible for federal taxation.

A major reform to provincial taxation occurred in 2000. Previous to this reform, provinces (outside Quebec) set their income taxes as a fraction of "basic federal tax."[17] An increase in this provincial tax rate affected all taxpayers proportionately, which strongly limited the ability of provinces to shape the distribution of the tax burden.[18] Following a reform in 2000–2001, provinces could set their own brackets and rates

[14] Data are from CANSIM table 380-0080, available at http://www5.statcan.gc.ca/cansim/. For this calculation, we include the three (small) Canadian territories in the "provincial" category.

[15] These natural resource revenues are chiefly the sale of oil and gas leases on public lands and royalties derived from mining and other extractive industries.

[16] The Spanish system since 2009 is a close parallel with most regional income tax rates set using a common centrally defined tax base and administration. See Albert Solé-Ollé (2013) for details.

[17] The "basic federal tax" was the tax liability generated by the federal tax rate and tax bracket calculation. Basic federal tax excludes special federal surtaxes and abatements. Quebec had its own tax base, bracket, and rate structure. The differences in tax base for our purposes are fairly minor. These provincial rates ranged in 1995 from 69 percent in Newfoundland and Labrador to 45.5 percent in Alberta.

[18] Provinces at that time did have the ability to add income surtaxes for high earners through which they could manipulate the tax liability and marginal tax rates of those at the high end of the income distribution.

applied to federally determined taxable income. This "tax-on-income" system gave provinces more flexibility in redistribution—and particularly the ability to operate a tax system with less progressivity than the federal one.

Some provinces did indeed respond by dropping tax rates and reducing progressivity. In particular, Alberta adopted a flat-rate income tax with a marginal rate of 10 percent, substantially lower than the 16–20 percent top rates existing in other provinces at that time. These tax differentials appear to have led to new strategies for shifting taxable income. One strategy, widely promoted by tax advisers,[19] was for high-income taxpayers in other provinces to transfer personal assets to an inter vivos trust resident in lower taxed provinces; income received by the trust is then taxed at the lower tax rate. Because only some assets need to be shifted, it is the marginal tax rate that matters for this decision. The fixed cost of setting up such avoidance strategies may make them practical only for the highest-income taxpayers.

An alternative tax planning strategy is simply for the taxpayer to declare residency in a lower taxed province. Income taxes in Canada are payable in the province of residence of a taxpayer on December 31 of each year, irrespective of the location of employment. Moreover, federal tax authorities may not closely scrutinize provincial residency claims.[20] This situation may be contrasted to that of the US states, where nexus for individual income taxation typically reflects the location of employment as well as residence, and state tax authorities may aggressively pursue false claims of residency.[21] Since a residence change involves moving all of one's income to the new province, it is the average tax rate which would matter most for the decision. However, for the highest income individuals of interest to us in our study, this distinction is less important as the average tax rate approaches the marginal tax rate at high incomes.

In short, the Canadian federal system is a useful testing ground for our theory of subnational income shifting in response to personal taxation. Our empirical strategy incorporates a framework that allows both for pure tax avoidance and for income shifting across provinces.

III. Data and Estimation Strategy

We estimate a log-linear specification for the tax base functions (1) of the form

$$(15) \qquad \log y_{it} = \alpha_i + \delta_t - e_u \tau_{it} + e_s \bar{\tau}_{-it} + x_{it}' \beta + \epsilon_{it},$$

where τ_{it} (the "own" tax rate) is the top marginal tax rate in province i and year t; $\bar{\tau}_{-it}$ is the average of contemporaneous tax rates in the other provinces, weighted by the inverse of distances between provincial capital cities (the "neighbor" average tax rate); and x_{it} is a vector of control variables discussed later in the paper.[22] In

---

[19] See Rochwerg (2004) and Cestnick (2003).
[20] See McClure (2013).
[21] State taxes are residence based when the states have reciprocity agreements. When there is no agreement, both residence and employment matter.
[22] We use this log-linear form rather than the standard log-log used in the literature because our specification fits most closely with our model.

our empirical implementation, we have a panel of taxable incomes and marginal tax rates for taxpayers in 10 provinces and 26 years.

We adopt the "share analysis" approach, common in the empirical literature on taxable income elasticities,[23] in which the dependent variable is the share of income reported for tax purposes by taxpayers in a top quantile of the distribution of reported income (mostly we look at the top 1 percent). The share approach may be derived from the model of individual behavior equation (15) under the assumption that taxpayers in the top quantile are influenced by top marginal tax rates according to equation (15), while the reported taxable income of others is not correlated with top marginal tax rates. Then, the use of the top income share on the left-hand side of equation (15) allows us to control for arbitrary shocks to incomes in province i and year t that are correlated with tax rates, but which leave the distribution of incomes unchanged. That is, we control for the log of total income in each province and year.

The source of our income data is the CANSIM high-incomes database.[24] We take the series for total income (excluding capital gains) as our main data for analysis.[25] We can observe for each of several fractiles the threshold cutoff for the fractile and the share of total income. We focus in this paper mostly on the ninety-ninth percentile cutoffs. We later graph the evolution of these top income shares across provinces against tax rates after first describing the construction of the tax rates.

The tax rates for our analysis come from the Canadian Tax and Credit Simulator (CTaCS 2016), which provides a calculation of income tax liability given a province, year, and a vector of income and family structure inputs. To calculate the marginal tax rates, we perform each simulation twice—once with the actual income and then again with earned income incremented by \$100. We take the difference in tax liability between these two runs and divide by 100 to obtain the marginal tax rate. Our primary focus is the ninety-ninth percentile cutoff, which we use to obtain the tax rate for those with incomes in the top 1 percent.[26]

An advantage of the Canadian example is the stability of the tax base across provinces and through time. Slemrod and Kopczuk (2002) notes that the taxable income elasticity depends on the tax base. So, interpretation of taxable income elasticities is facilitated by a constant tax base. Nine of ten Canadian provinces use the federal tax base, and this base has remained fairly constant since 1988.[27] This motivates our selection of 1988 to 2013 as our sample period.

<hr>

[23] See Saez, Slemrod, and Giertz (2012) for a survey.

[24] CANSIM table 204-0002 provides high-income threshold cutoffs and income totals for several measures of income covering the years 1982 to 2013. The ultimate source of these data is the Longitudinal Administrative Databank, a 20 percent sample of Canadians drawn from tax records.

[25] The database also includes "market" income, which excludes transfer income. Both total and market income are available with and without capital gains included. Our results are little changed when one of these alternatives is used in place of total income without capital gains. See Milligan and Smart (2015).

[26] We use the national ninety-ninth percentile income cutoff and adjust by inflation for the other years. Using a common income source across province and time isolates the statutory variation from temporal and provincial variation in underlying incomes. Also, over the time period covered by our data, the bracket threshold for facing the highest tax rate was lower than the ninety-ninth percentile cutoff in almost all cases. The sole exceptions are Ontario, which introduced a new tax bracket starting at \$500,000 in 2012 while the top 1 percent threshold was \$225,600, and Nova Scotia, which implemented a bracket at \$150,000 in 2010 below the top 1 percent threshold. This means that using the national inflation-adjusted income cutoff is not consequential to our analysis.

[27] The exception is the province of Quebec, but the differences in the tax base are minimal. Since 1988, the federal treatment of capital gains has changed three times, motivating our choice to use income excluding capital gains.
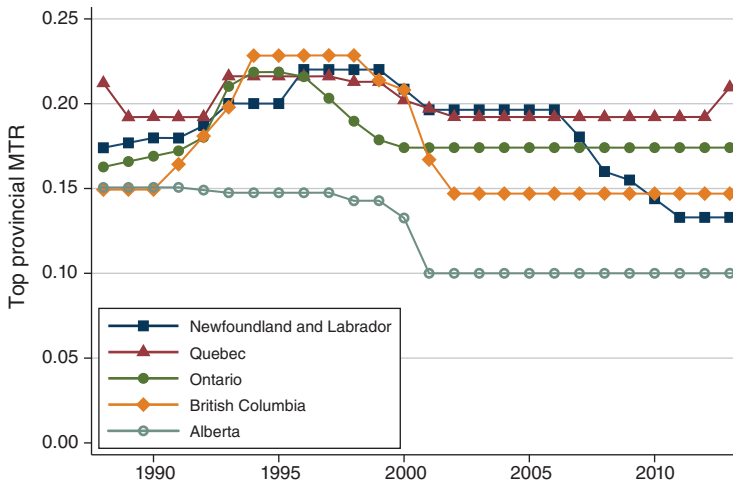
FIGURE 1. TOP PROVINCIAL TAX RATES, 1988–2013

Note: This figure shows the tax rate at the ninety-ninth percentile threshold for each province through time.
Source: Calculations using CTaCS.

We graph the provincial high-income tax rates by year for selected provinces in Figure 1, illustrating the variation in tax rates and cross-province differentials in our data.[28] For example, the top rate in Alberta hovered around 15 percent through most of the 1990s, but dropped to 10 percent in 2001 with the implementation of the flat-rate tax following the decentralization reform. As another example, British Columbia started with a tax slightly lower than Alberta's until 1991, then moved up to have the highest rate nearing 23 percent in the mid-1990s before falling back to the second lowest in the early 2000s at 14.7 percent.

A. *Preliminary Evidence*

Note that equation (15) includes jurisdiction and year fixed effects—it is a difference-in-difference estimator that allows for arbitrary fixed differences in income distribution among jurisdictions and nationally over time. This allows for a more robust inference of taxable income elasticities than is typically possible with the share analysis approach, which normally uses national-level data, and only time series variation in tax rates. To allow for within-province correlation, we cluster our standard errors by province. Because of the small number of cluster groups, we make our inferences based on a *t*-distribution with nine degrees of freedom.[29]

---

[28] In the regression analysis later in the paper, we use the combined federal-provincial marginal tax rate. However, in order to highlight the provincial variation, we graph here only the provincial component.

[29] This follows the method of Bester, Conley, and Hansen (2011), who find that with a fixed and small number of groups, the limiting distribution can be approximated using a *t*-distribution with the degrees of freedom set at the number of groups minus one.
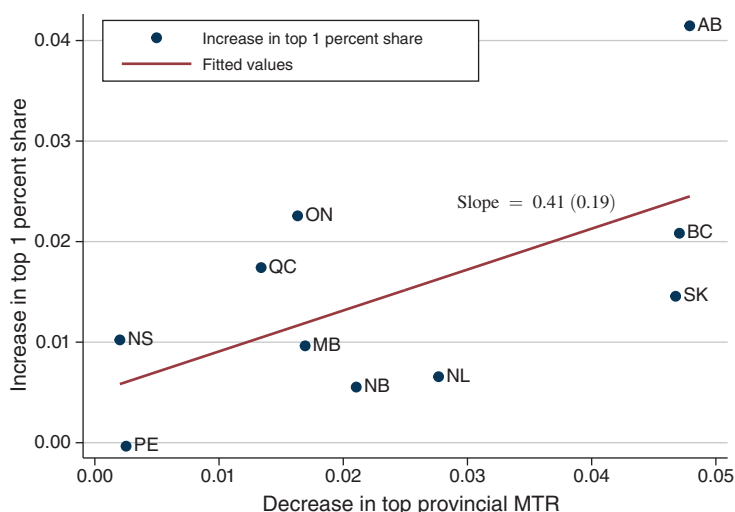
FIGURE 2. EFFECTS OF THE REFORM ON TAX RATES AND INCOME SHARES

To summarize the basic relationship of top income shares and tax rates, we look on either side of the 2000/2001 reform that switched from the old "tax-on-tax" to the current "tax-on-income" system. Figure 2 plots the change in top tax rates in each province from the pre-reform (1988–1999) to the post-reform (2001–2013) periods, against the corresponding change in the top 1 percent income share. The plot shows a clear positive correlation between tax reductions and reported income increases within provinces. The slope of the line of best fit is 0.41, which is the simple difference-in-difference estimate of the own tax elasticity derived from provincial tax changes around the reform. Of course, inferences based on this approach are subject to endogeneity concerns as the tax changes and top income shares may both be related to omitted determinants of the tax base. This motivates the need for the instrumental variables strategy we describe later in the paper.

Observe also that the neighbor average tax rate $\bar{\tau}_{-it}$ varies among provinces each year because it is a "leave-out" mean of contemporaneous tax rates that is weighted by the inverse of distances between provincial capital cities. Since we include year fixed effects in our regressions, the effect of the neighbor tax rate on tax bases is therefore identified from regional differences in how tax rates have evolved over time.

In a frictionless model, the minimum of the provincial tax rates for τ_{-it} might be considered as taxpayers would shift their income to the province with the most attractive rate. However, shifting income may require familiarity with legal and financial institutions in the recipient province, for which distance is a suitable proxy. Our use of distance here aligns with the logic of the gravity model used in empirical studies of international trade.[30]

---

[30] See, for example, Anderson and van Wincoop (2003) for a reconciliation of the use of the empirical gravity model with theory for international trade. We have also tried a simple other-province average. However, this yields

B. *Instrumental Variables*

We are interested in the relationship between reported top incomes and top tax rates at the province-year level. An obvious concern is the potential for bias in estimating elasticities resulting from omitted variables and other sources of endogeneity of tax rates. A jurisdiction's own tax rate may be endogenous if a local shock to high incomes leads to a change in the top tax rate there—as, for example, if the government responds to an increase in the tax base by reducing the rate to keep revenues relatively constant. More generally, any omitted province time-varying variables that are correlated with top tax rates and top income shares would lead to bias.

In Figure 2, the three provinces with the largest change in tax rates are ones with a high concentration of natural resource revenues. Alberta, British Columbia, and Saskatchewan clustered between a four and five point drop in the provincial marginal tax rate. Provincial governments' room to reduce tax rates following the reform likely depended on their other fiscal resources, which could insulate them against the risk of revenue declines following a tax cut. So provincial access to resource revenues—which is in large part determined by geological endowment— creates arguably exogenous variation in the extent to which an individual province responded to the decentralization reform by changing the top tax rate. In this sense, the more resource-dependent provinces are the "compliers" for which we can estimate a treatment effect of tax cuts on tax bases, using the exogenous variation in tax rates resulting from the 2000 reform.

We use this strategy to provide instrumental variables estimates of tax elasticities using as our main instrument each province's average share of resource revenues in total revenues over the sample period ($ResShare_i$), multiplied by a dummy variable for post-reform years ($POST_t$). The resource share component of this instrument is not time-varying, so our instrument does not rely on the time series trends in resource markets: the part of the instrument that is time varying is the decentralization reform. Our instrument captures the impact of the reform on provinces that typically have flexibility (owing to their preexisting resource revenues) compared to those that don't have this flexibility. To account for the possible direct effect of resource revenues on tax rates and income distribution, we include the time-varying variable $ResShare_i \times ResPrice_t$ as a control variable throughout the analysis, where $ResPrice_t$ is a national index of resource prices,[31] which captures the direct impact of resource price shocks on provincial government finances. In this way, our approach resembles the common Bartik (1991) strategy, combining a fixed characteristic (natural resource revenue share) with a time-varying factor (the 2000/2001 reform).

If the own tax rate τ_{it} is endogenous in equation (15), then contemporaneous spatial correlation in unobservable shocks could also cause the neighbor average tax rate $\bar{\tau}_{-it}$ to be endogenous. Analyzing a very similar issue in the context of estimating peer

---

much less variation across provinces—the within-year standard deviation is on average three times higher for the distance-weighted mean than the simple mean.

[31] We use the raw materials price index in CANSIM table 330-0008.

effects of educational attainment, Acemoglu and Angrist (2001) argues that what is required is a second instrument for the individual (i.e., own tax effect) that is not strongly correlated with the instrument for the peer (i.e., neighbor tax) effect. Since $ResShare_i \times POST_t$ is the "natural" instrument for τ_{it} given by the decentralization reform and $\bar{\tau}_{-it}$ is a weighted average of τ_{it}, it is tempting to treat the corresponding weighted average of $ResShare_i \times POST_t$ as the analogous additional instrument for $\bar{\tau}_{-it}$ necessary to identify equation (15) when both tax rates are endogenous. But the two instruments are too closely correlated to yield independent variation.

In our context, another possibility is idiosyncratic political variation that affects tax rates within a province. We therefore construct a dummy variable NDP_{it} equal to one when the provincial government is controlled by the New Democratic Party, a social democratic party that formed the government in approximately 18.8 percent of the province-year cells in our data.[32] We expect the NDP_{it} to be predictive of increases in provincial tax rates because of party ideology. This instrument will be successful only to the extent that there are not other mechanisms through which an NDP provincial government influences top income shares.

IV. Results

We begin in Table 1 by reporting ordinary least squares estimates of our difference-in-difference model (15). In the first two columns of the table, we exclude the neighbor average tax rate and report the own tax rate semi-elasticity alone. So, the parameter of interest is the combined shifting and avoidance elasticity e_u, derived from the coefficient on τ_{it}.

Looking initially at the own tax effect alone facilitates comparisons with the previous literature on taxable income elasticities, and it serves as a benchmark for our subsequent results including the neighbor tax effects.[33] For the simple difference-in-difference specification, the estimated semi-elasticity is -2.31. Elasticity of taxable income is typically reported in the literature as the elasticity of the tax base with respect to changes in one minus the tax rate. For comparison purposes, this is also reported in the table at the means of the data. The estimated ETI in this case is 1.21, which is rather high.

The next column includes province-specific economic conditions that may be correlated with tax changes. To control for the business cycle, we include the log of total income of all tax filers as reported in the tax records. To control for the effects of commodity prices on income distribution in resource provinces, we include the resource revenue share-resource price variable, $ResShare_i \times ResPrice_t$. Controlling for these variables markedly improves the fit of the regression, but it leaves the estimated own tax elasticity nearly unchanged.

The last column of Table 1 reports estimates from a specification including the neighbor average tax rate. At the bottom of the table, we report the avoidance and

---

[32] To account for the timing of the budget process, NDP_{it} is lagged one year.

[33] The results can be compared to the extensive analysis of the own-province relationship appearing in Milligan and Smart (2015), although there are some differences. First, we use a different specification here. Second, we have updated the data to include 2013 along with some improvements to the previous years' data. The goal here is to lay down a baseline result comparable to the literature before we allow for interprovincial shifting in the next table.

TABLE 1—OLS ESTIMATES

Variables	log 1 percent share (1)	log 1 percent share (2)	log 1 percent share (3)
Own tax rate	−2.31 [0.93]	−2.19 [0.46]	−2.17 [0.45]
Neighbor tax rate			0.62 [0.89]
log total income		0.94 [0.09]	0.94 [0.09]
ResShare × *ResPrice*		−0.17 [0.03]	−0.16 [0.03]
Observations	250	250	250
R^2	0.25	0.61	0.61
Number of provinces	10	10	10
Combined elasticity	1.21	1.14	
Avoidance elasticity			0.81
Shifting elasticity			0.32

Notes: All specifications include year and province fixed effects. Robust standard errors are clustered by province. We use a *t*-distribution with nine degrees of freedom.

shifting elasticities implied by our point estimates as appearing in equation (15), backing out the value for e_a using the combined estimate for e_u on the own tax rate. The neighbor tax effect is insignificantly different from zero in this specification, and other coefficients are essentially unchanged. As expected, therefore, the OLS estimates show little evidence of cross-jurisdiction income shifting.

A. *Instrumental Variables Results*

We now turn to instrumental variables estimates derived from the differential effects of the 2000 decentralizing reform, which are presented in Table 2. In column 1, we again return to the specification which excludes the neighbor average tax rate. As previously explained, the instrument for the own tax rate τ_{it} is $ResShare_i \times POST_t$. The *F*-statistic for significance of the excluded instrument in the first-stage regression is 12.2. (The coefficient on the instrument in the first stage, unreported in the table for brevity, is −0.0011 (0.0003), indicating that tax rates fell 1.1 percentage points more in provinces with a resource share of 10 percent of total revenues, compared to a province with no resource revenues.) The estimated semi-elasticity is −1.40, lower than the corresponding OLS estimate of column 2 of Table 1.

Recall that, if τ_{it} is endogenous in equation (15) and there is contemporaneous spatial correlation in tax rates, then $\bar{\tau}_{-it}$ is endogenous also.[34] If the own tax rate is negatively correlated with omitted variables increasing the tax base as suggested

---

[34] Indeed, $\bar{\tau}_{-it}$ might be a valid instrument for τ_{it} in equation (15), which complicates interpretation of the OLS coefficients. Tax rates of neighboring jurisdictions are in fact often used as instruments in empirical research on local and state public finance.

TABLE 2—IV ESTIMATES

Variables	log 1 percent share (1)	log 1 percent share (2)	log 1 percent share (3)
Own tax rate	−1.40 [0.50]	−2.21 [0.45]	−2.40 [0.48]
Neighbor tax rate			1.85 [0.86]
log total income	0.92 [0.10]	0.94 [0.09]	0.95 [0.10]
$ResShare \times ResPrice$	−0.14 [0.04]	−0.17 [0.03]	−0.16 [0.03]
R^2	0.58	0.61	0.60
Instruments	$ResShare \times POST$ -	$ResShare \times POST$ NDP	$ResShare \times POST$ NDP
First-stage F			
Own tax rate	12.2	49.3	49.3
Neighbor tax rate	-	-	114.4
Combined elasticity	0.73	1.16	
Avoidance elasticity			0.29
Shifting elasticity			0.97

Notes: All specifications include year and province fixed effects. Robust standard errors are clustered by province. We use a *t*-distribution with nine degrees of freedom.

by the results in column 1, then since the own tax rate and the neighbor average tax rate are necessarily positively correlated in aggregate, the OLS estimate of the coefficient on the average tax rate is biased downward in equation (15), so that we would tend to reject a cross-jurisdiction shifting effect even if one were present. We therefore treat the own and neighbor average tax rates as endogenous, and we seek instruments for both.

Column 2 presents two-stage least squares estimates of equation (15), where the neighbor tax rate is excluded and the weighted average of $ResShare_i \times POST_t$ and NDP_{it} are the excluded instruments for the own tax rate. In this case, the F-statistic on excluded instruments in the first stage is 49.3 and the first-stage coefficient on NDP_{it} is 0.0171 (0.0035), indicating that election of an NDP government is associated with a top tax rate that is 1.7 percentage points higher than other parties. The estimated tax semi-elasticity is −2.21, almost the same as in column 1, suggesting that the NDP instruments correctly overidentifies the own tax rate effect.

Column 3 then presents 2SLS estimates of the full model including the neighbor average tax rate, where the instruments are the same as in column 2. In this case, the point estimate of the shifting semi-elasticity (the coefficient on the neighbor average tax rate) is large and still significant at the 10 percent level of confidence even taking into account the grouped data. Working through the implications of this point estimate on the shifting and avoidance elasticities reveals a large impact—the avoidance elasticity is itself small, and the shifting elasticity is large. Thus, our results suggest that while provincial tax bases are highly responsive to unilateral tax changes, much of this elasticity may be accounted for by the shifting of income between provinces. In contrast, our estimates suggest that a federal or coordinated

TABLE 3—ALTERNATIVE INCOME THRESHOLDS

Variables	log 10 percent share (1)	log 5 percent share (2)	log 1 percent share (3)	log 0.1 percent share (4)
Own tax rate	−0.61 [0.10]	−1.53 [0.31]	−2.40 [0.48]	−3.08 [1.15]
Neighbor tax rate	1.05 [0.58]	0.81 [0.76]	1.85 [0.86]	4.08 [0.94]
log total income	0.27 [0.06]	0.54 [0.05]	0.95 [0.10]	1.11 [0.30]
ResShare × *ResPrice*	−0.01 [0.02]	−0.08 [0.02]	−0.16 [0.03]	−0.15 [0.10]
Observations	250	250	250	198
R^2	0.64	0.60	0.60	0.44
Number of provinces	10	10	10	8

Notes: All specifications include year and province fixed effects. Robust standard errors clustered by province. We use a *t*-distribution with nine degrees of freedom in columns 1, 2, and 3 and with seven degrees of freedom in column 4.

provincial tax rate increase would have a relatively small effect on the high-income tax base since there would be no cross-province shifting in response to a national tax change.

B. *Very High-Income Taxpayers*

Table 3 delves further into tax responsiveness by estimating the impact of top tax rate changes on top income shares for several different quantiles of the income distribution. The institutional tax-shifting channels described earlier are likely to be more prevalent among the highest earners. So, we expect to see stronger results among the highest earners.

In all columns reported in the table, the same two instruments for own and neighbor average tax rates are used as in the last column of Table 2. The only difference among specifications is therefore in the income threshold which defines the dependent variable, which ranges from the P90 threshold (i.e., the income share of the top 10 percent of taxpayers) in column 1 to P95, P99, and P99.9 successively in the remaining columns. The results show that estimated effects of own tax and neighbor average tax rates both rise as we examine taxpayers with higher incomes. Thus, the estimated own tax rate semi-elasticity is −0.61 for the top 10 percent threshold; the neighbor tax effect in this case while larger in magnitude is insignificantly different from 0. The results are starkest in the case of the P99.9 threshold, which captures the impact of tax changes on the top one-tenth of 1 percent of taxpayers—about 22,000 taxpayers in our data in a typical year. (In this case, our sample is somewhat smaller, due to masking of data in small provinces to meet confidentiality restrictions.) In this case, while both estimated semi-elasticities are large in magnitude, the neighbor tax effect in fact exceeds the own tax effect. On the basis of these estimates, we cannot reject the hypothesis that the elasticity of taxable income is all about interprovincial shifting. This evidence of increasing elasticities is consistent with a model featuring fixed costs of accessing avoidance technology.

V. Simulations

Top marginal rates in Canadian provinces currently range from 15 to 21 percent—roughly one-half to two-thirds of the federal rate. Moreover, several provinces have recently increased their top tax rates, even as the federal top rate remained stable in the 2000s.[35] However, our estimates of the shifting semi-elasticity are large, suggesting considerable potential for horizontal tax competition. These facts may appear at first glance to be inconsistent. In the Musgravian perspective, provinces would avoid increasing taxes at the top. However, there is substantial variation in top tax rates among provinces, suggesting prima facie an Oatesian case for decentralization. Can these facts be reconciled with our elasticity estimates in the light of our model? To weigh these considerations, we report here on numerical simulations of our formal model of Section I, using the estimated elasticities of Section IV.

The estimation model and formal welfare analysis mentioned earlier take as given a tax base function that embodies avoidance and shifting responses to taxation through the functional form $y_i(\tau_i, \tau_i - \bar{\tau})$. To get closed-form solutions for equilibrium tax rates and revenues, we simulate the model for the case of linear tax base functions:

$$(16) \qquad y_i(\tau_1, \ldots, \tau_J) = z_i - a\tau_i - s(\tau_i - \bar{\tau}),$$

where a, s are the avoidance and shifting parameters. In Appendix A, we show how equation (16) can be derived from a standard linear-quadratic model of individual taxpayer behavior, and we provide the formal derivations behind our simulation results.

The linear-quadratic model admits closed-form solutions for equilibrium tax rates and tax revenues, which facilitates comparisons among our three tax assignment alternatives—unitary taxation, full decentralization, and federalism with shared federal and state taxation. The key elements in the trade-offs among these systems are the degree of tax avoidance that is due to interstate shifting (which determines the cost of horizontal tax competition) and the degree of heterogeneity in tax base yields among states (which determines the welfare gains to differentiated taxation). For convenience, let $\delta_i = z_i - k_i$ denote the potential tax base per high-income taxpayer in each state.

In the Appendix, we show that tax revenues (per taxpayer) under a unitary tax system, with uniform rates in all states, optimal national tax revenues per taxpayer are

$$(17) \qquad R(\tau^U) = \frac{\bar{\delta}^2}{4a},$$

---

[35] Between 2010 and 2014, the top marginal tax rate is increasing in the provinces of Nova Scotia (19.2 percent to 21 percent), New Brunswick (14.3 percent to 17.8 percent), Quebec (24 percent to 25.75 percent), Ontario (17.4 percent to 20.5 percent), and British Columbia (14.7 percent to 16.8 percent). The federal top rate stayed at 29 percent throughout the 2000s, only changing upward in 2016.

where $\bar{\delta}$ is the national average yield parameter. Under a system of full decentralization, tax rates are lower on average due to tax competition, but they are differentiated among states in a way that responds to yield differentials. National revenues in equilibrium in the decentralized system are (see the Appendix)

$$(18) \qquad \bar{R}(\tau^D) \ = \ \frac{\bar{\delta}}{4a}\left[1 - \frac{s}{(2a+s)^2} + \frac{a}{a+s}\,\gamma^2\right],$$

where

$$\gamma \ = \ \frac{\mathrm{var}^{1/2}(\delta)}{\bar{\delta}}$$

is the coefficient of variation of δ_i among states, the relevant measure of heterogeneity in this linear-quadratic case. Finally, consider a federal system with federal and state co-occupancy of the tax base. From Proposition 2, the optimal national tax structure τ^N can be decentralized by a federal government acting as a Stackelberg leader and setting a federal tax rate that offsets the horizontal and vertical fiscal externalities. In the Appendix, we show that national revenues in the federal system are

$$(19) \qquad \bar{R}(\tau^N) \ = \ \frac{\bar{\delta}}{4a}\left[1 + \frac{2a}{a+s}\gamma^2\right].$$

Figure 3 graphs the three revenue functions against γ, the coefficient of variation in yield parameters. To construct the figure, we calibrate the tax base semi-elasticities to our preferred point estimates from column 3 of Table 2.[36] For clarity in the figure, we have normalized revenues from unitary taxation to one. We discuss each of the three options in turn.

With full Musgravian centralization, the revenue yield of the unitary tax $\bar{R}(\tau^U)$ is invariant to the dispersion in tax yields across jurisdictions, so the revenue line is flat at the normalized value of 1.0.

With full decentralization, the revenue yield $\bar{R}(\tau^D)$ is less than centralization $\bar{R}(\tau^U)$ at low levels of tax yield dispersion since the horizontal externality leads to own-revenue loss when states try to increase state taxes, so tax rates are lower in equilibrium. The losses due to full decentralization are increasing in the degree of inter-state tax competition (as measured by s/a) regardless of whether there is cross-state heterogeneity or not. However, it can be seen in the graph that decentralized revenues $\bar{R}(\tau^D)$ are increasing in the coefficient of variation γ since greater heterogeneity in yields creates more-than-proportionate gains in optimal differentiated taxation.

---

[36] To ensure an informative simulation, we chose these results as they have significant shifting and avoidance elasticities. The main impact of different point estimates in these simulations is the crossover point and steepness of the lines. Given the estimated standard errors in column 3 of Table 2, there is substantial overlap in the confidence intervals for the federalism and decentralization cases, but the point estimates used here convey how the three tax systems relate to each other.
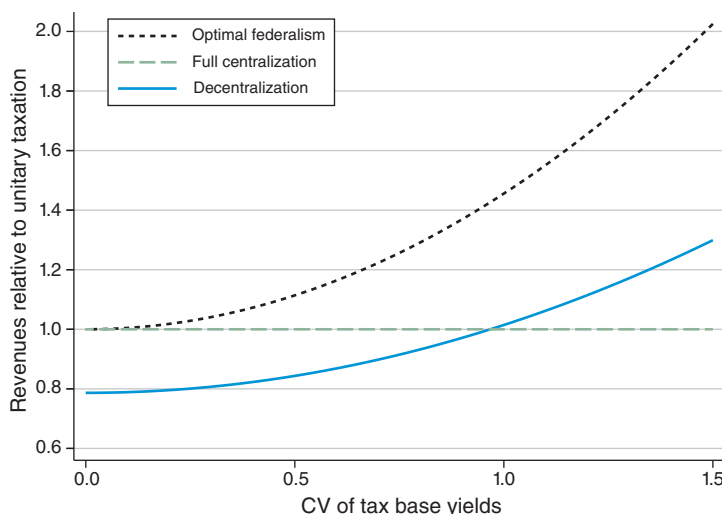
FIGURE 3. SIMULATED REVENUES UNDER ALTERNATIVE TAX SYSTEMS

Finally, at some point in γ (using our parameter estimates this point is just less than 1.0 in Figure 3), decentralization overtakes centralization. We can show that

$$\bar{R}(\tau^D) \geq \bar{R}(\tau^U) \iff \gamma \geq \hat{\gamma} \equiv \frac{1 + s/a}{(1 + 2a/s)^2}.$$

In words, full decentralization with interstate tax competition is preferred to Musgravian unitary taxation as long as the dispersion in tax base yields is large enough and the degree of interstate tax shifting (as measured by s/a) is not too large.

The third case is optimal federalism. Comparing the expressions (17)–(19), one sees that

$$\bar{R}(\tau^N) \geq \max\{\bar{R}(\tau^U), \bar{R}(\tau^D)\},$$

with strict inequality whenever $\gamma > 0$. Thus, optimal federalism is strictly superior to either unitary taxation or full decentralization whenever there is heterogeneity in tax base yields among states. This can be seen in the figure as the optimal federalism line lies everywhere above the other lines when $\gamma > 0$. This holds independent of parameter choices.

VI. Concluding Remarks

This paper studies a model of federal taxation featuring both interjurisdictional externalities and heterogeneity in tax yields across states and considers the efficiency of different taxation arrangements.

The traditional view in economics of fiscal federalism is that redistributive taxation should be assigned to the national government and not to subnational governments (e.g., Musgrave 1971). Given the potential for migration between jurisdictions within the federation, decentralized redistribution gives rise to two problems. First, interregional tax and transfer differences give rise to locational inefficiencies. Second, given migration responses, subnational governments are apt to undercut each other in redistributive taxes. Since both of these problems are absent under unitary government, centralization is held to be preferred for redistributive taxation of mobile factors. These considerations apply whether the tax-induced migration reflects mobility of real resources or the pure tax base shifting studied in this paper.

The empirical work in this paper however highlights the potential for heterogeneity in the optimal tax policies of states within a federation: states with greater inequality in top incomes face lower marginal excess burden of taxation per dollar of marginal revenue and so should optimally impose higher tax rates than others in the federation. If federal tax policies are constrained to be uniform in all states, this strengthens the case for decentralization in a manner that is reminiscent of Oates's (1972) "Decentralization Theorem." Furthermore, the case for decentralization is stronger still if the federal government piggybacks its own tax on the base nationally and adjusts the rate to limit the "race to the bottom" in state tax policies.

A question for future research is to evaluate the conflicting roles of subnational heterogeneity in the model: as regions diverge in income distribution, the "Oatesian" case for decentralization is strengthened, but the potential for tax base shifting and horizontal tax competition increases as well. It should be possible to use our model of avoidance and shifting and our estimates of the relevant elasticities to simulate the welfare gains (or losses) of centralizing income taxation in a federation.

Appendix A: A Structural Model of Income Shifting

A representative taxpayer in state i has potential income z_i and faces statutory marginal tax rate τ_i on taxable income above a fixed bracket threshold k_i. We note that the marginal tax rate τ_i comprises both a federal component T that is common to all states and a state component $t_i = \tau_i - T$ that may vary among states. The taxpayer can shelter α_i dollars in income from tax by engaging in a tax avoidance activity and can also shift σ_{ij} dollars of income to each state $j = 1, \ldots, J$, where it is taxed at the corresponding top rate τ_j.

The taxpayer resident in i maximizes

$$u_i(\tau) = z_i - \tau_i(z_i - k_i) - \tau_i \alpha_i - \sum_j (\tau_i - \tau_j) \sigma_{ij} - C_a(\alpha_i) - \sum_j C_s(\sigma_{ij}),$$

where the functions C_a and C_s measure the deadweight costs of avoidance and income shifting, respectively. Assuming that these functions are quadratic:[37]

$$C_a(\alpha) = \frac{1}{2a}\alpha^2,$$

$$C_s(\sigma) = \frac{1}{(s/J)}\sigma^2,$$

gives optimal avoidance and shifting rules

$$\alpha_i^* = a\tau_i, \quad \sigma_{ij}^* = -\sigma_{ji}^* = \frac{s/2}{J}(\tau_i - \tau_j),$$

and the tax base function (16).

Top-bracket revenues (per high-income taxpayer) are

$$(A1) \qquad R_i(\tau) = \tau_i(y_i(\tau) - k_i) = \tau_i(\delta_i - a\tau_i - s(\tau_i - \bar{\tau})).$$

Suppose there are J states with equal populations of high-income taxpayers. National revenues per taxpayer can be computed (summing equation (A1) and dividing by J) to be

$$(A2) \qquad \bar{R}(\tau) = \bar{\tau} \cdot \bar{\delta} + \text{cov}(\tau, \delta) - a\bar{\tau}^2 - (a + s)\text{var}(\tau),$$

where $\text{var}(\tau)$ and $\text{cov}(\tau, \delta)$ indicate the sample variance of τ_i among states and its sample covariance with δ_i, respectively.[38]

The revenue expressions (A1) and (A2) facilitate analysis of optimal taxation in the various cases of unitary taxation, optimal federalism, and full decentralization, as these are defined in the main text. In the case of unitary taxation, $\tau_i^U = \tau$ for all i, and the national average revenue reduces to

$$\bar{R}(\tau^U) = \tau\bar{\delta} - a\tau^2.$$

Straightforward differentiation yields the optimal unitary tax rate $\tau^U = \bar{\delta}/(2a)$ and optimal unitary revenues

$$(A3) \qquad R(\tau^U) = \frac{\bar{\delta}^2}{4a}.$$

In the case of a federal system with federal and state co-occupancy of the tax base, Proposition 2 shows that the national optimum is implementable when the

---

[37] This is a model of income shifting commonly used in the literature on international tax avoidance; see, e.g., Mintz and Smart (2004).
[38] That is, $\text{var}(\tau) = J^{-1}\sum \tau_j^2 - \bar{\tau}^2$ and $\text{cov}(\tau, \delta) = J^{-1}\sum \tau_j \delta_j - \bar{\tau} \cdot \bar{\delta}$.

federal government moves first and sets the optimal federal tax rate. Differentiating equation (A2) and rearranging, we find that in the linear-quadratic case the optimal tax rates are

$$(A4) \qquad \tau_i^N = \frac{\bar{\delta}}{2a} + \frac{\delta_i - \bar{\delta}}{2(a+s)}.$$

Substituting into equation (A2), national average revenues at the federal optimum are

$$(A5) \qquad \bar{R}(\tau^N) = \frac{\bar{\delta}}{4a}\left[1 + \frac{2a}{a+s}\gamma^2\right],$$

where

$$\gamma = \frac{\mathrm{var}^{1/2}(\delta)}{\bar{\delta}}.$$

Finally, in the case of a fully decentralized system, with no federal taxation and tax competition among states, we may obtain the best-response tax rate functions for each state $\tau_i^*(\bar{\tau})$ by maximizing own revenue as given in equation (A1), taking the average tax rate $\bar{\tau}$ as given.[39] These satisfy the first-order conditions

$$(A6) \qquad 2(a+s)\tau_i^* = \delta_i + s\bar{\tau}.$$

Solving for the fixed point of the best responses, the Nash equilibrium tax rates under full decentralization are

$$(A7) \qquad \tau_i^D = \frac{\bar{\delta}}{2a+s} + \frac{\delta_i - \bar{\delta}}{2(a+s)}.$$

Substituting again into equation (A2), national average tax revenues under full decentralization are

$$(A8) \qquad \bar{R}(\tau^D) = \frac{\bar{\delta}}{4a}\left[1 - \frac{s}{(2a+s)^2} + \frac{a}{a+s}\gamma^2\right].$$

APPENDIX B

PROOF OF PROPOSITION 1:

Differentiating the definition of $\bar{\tau}$ in equation (2) with respect to τ_i and using the definition of e_s yields

$$\frac{\partial \bar{\tau}}{\partial \tau_i}\sum_j y_j + \bar{\tau}e_s y_j \frac{\partial \bar{\tau}}{\partial \tau_i} = \sum_j \tau_j e_s y_j \frac{\partial \bar{\tau}}{\partial \tau_i} + y_i.$$

---

[39] In this case, we assume that each state is infinitesimally small, so that its effect on the national average tax rate is negligible. The case of a finite J is qualitatively the same but somewhat more tedious to compute.

Simplifying and noting that $\sum_j (\tau_j - \bar{\tau}) y_j = 0$, we have

$$
\text{(B1)} \qquad \frac{\partial \bar{\tau}}{\partial \tau_i} = \frac{y_i}{\sum_j y_j} \equiv \omega_i.
$$

Using equation (B1), it is straightforward to verify equation (6) by differentiation. Likewise, differentiating equation (5) and using equation (B1), the marginal effect on national revenue satisfies

$$
\frac{\partial R}{\partial \tau_i} = (y_i - k_i) + \tau_i(y_{i1} + y_{i2}) - \sum_j \tau_j y_{j2} \omega_i
$$

$$
= y_i[\theta_i - (e_a + e_s)\tau_i] + \sum_j \tau_j y_i \omega_i e_s.
$$

Substituting the definitions of $\bar{\tau}$ and ω_i then immediately yields equation (7). ∎

REFERENCES

Acemoglu, Daron, and Joshua Angrist. 2001. "How Large are Human-Capital Externalities? Evidence from Compulsory-Schooling Laws." In *NBER Macroeconomics Annual 2000*, Vol. 15, edited by Ben S. Bernanke and Kenneth Rogoff, 9–74. Cambridge, MA: MIT Press.

Akcigit, Ufuk, Salomé Baslandze, and Stefanie Stantcheva. 2016. "Taxation and the International Mobility of Inventors." *American Economic Review* 106 (10): 2930–81.

Alesina, Alberto, and Enrico Spolaore. 1997. "On the Number and Size of Nations." *Quarterly Journal of Economics* 112 (4): 1027–56.

Anderson, James E., and Eric van Wincoop. 2003. "Gravity with Gravitas: A Solution to the Border Puzzle." *American Economic Review* 93 (1): 170–92.

Atkinson, Anthony B., and Andrew Leigh. 2013. "The Distribution of Top Incomes in Five Anglo-Saxon Countries Over the Long Run." *Economic Record* 89 (S1): 31–47.

Bartik, Timothy J. 1991. *Who Benefits from State and Local Economic Development Policies?* Kalamazoo: W. E. Upjohn Institute for Employment Research.

Bester, C. Alan, Timothy G. Conley, and Christian B. Hansen. 2011. "Inference with dependent data using cluster covariance estimators." *Journal of Econometrics* 165 (2): 137–51.

Bucovetsky, Sam, and Andreas Haufler. 2007. "Preferential Tax Regimes with Asymmetric Countries." *National Tax Journal* 60 (4): 789–95.

Buettner, Thiess. 2003. "Tax base effects and fiscal externalities of local capital taxation: Evidence from a panel of German jurisdictions." *Journal of Urban Economics* 54 (1): 110–28.

Canadian Tax and Credit Simulator 2016–1.03. 2016. Vancouver School of Economics. http://faculty.arts.ubc.ca/kmilligan/ctacs/ (accessed March 31, 2016).

Cestnick, Tim. 2003 "Consider an Alberta trust even if you don't live there." *Globe and Mail*, June 21. https://www.theglobeandmail.com/globe-investor/investment-ideas/consider-an-alberta-trust-even-if-you-dont-live-there/article751349/.

Chetty, Raj. 2009. "Is the Taxable Income Elasticity Sufficient to Calculate Deadweight Loss? The Implications of Evasion and Avoidance." *American Economic Journal: Economic Policy* 1 (2): 31–52.

Day, Kathleen M., and Stanley L. Winer. 2006. "Policy-induced internal migration: An empirical investigation of the Canadian case." *International Tax and Public Finance* 13 (5): 535–64.

Diamond, Peter, and Emmanuel Saez. 2011. "The Case for a Progressive Tax: From Basic Research to Policy Recommendations." *Journal of Economic Perspectives* 25 (4): 165–90.

Epple, Dennis, and Thomas Romer. 1991. "Mobility and Redistribution." *Journal of Political Economy* 99 (4): 828–58.

Feldstein, Martin. 1999. "Tax Avoidance and the Deadweight Loss of the Income Tax." *Review of Economics and Statistics* 81 (4): 674–80.

Feldstein, Martin, and Marian Vaillant Wrobel. 1998. "Can state taxes redistribute income?" *Journal of Public Economics* 68 (3): 369–96.

Gagné, Robert, Jean-François Nadeau, and François Vaillancourt. 2001. "Taxpayers' Response to Tax Rate Changes: A Canadian Panel Study." http://www.iedm.org/uploaded/pdf/taxpayers.PDF.

Giertz, Seth H. 2007. "The Elasticity of Taxable Income over the 1980s and 1990s." *National Tax Journal* 60 (4): 743–68.

Gordon, Roger H., and Julie Berry Cullen. 2011. "Income redistribution in a Federal system of governments." *Journal of Public Economics* 96 (11–12): 1100–1109.

Gordon, Roger H., and Joel B. Slemrod. 2000. "Are 'Real' Responses to Taxes Simply Income Shifting between Corporate and Personal Tax Bases?" In *Does Atlas Shrug? The Economic Consequences of Taxing the Rich*, edited by Joel Slemrod, 240–80. Cambridge, MA: Harvard University Press.

Gruber, Jon, and Emmanuel Saez. 2002. "The elasticity of taxable income: Evidence and implications." *Journal of Public Economics* 84 (1): 1–32.

Johannesen, Niels. 2014. "Tax evasion and Swiss bank deposits." *Journal of Public Economics* 111: 46–62.

Keen, Michael. 1998. "Vertical Tax Externalities in the Theory of Fiscal Federalism." *IMF Economic Review* 45 (3): 454–85.

Keen, Michael. 2001. "Preferential Regimes Can Make Tax Competition Less Harmful." *National Tax Journal* 54 (4): 757–62.

Keen, Michael J., and Christos Kotsogiannis. 2002. "Does Federalism Lead to Excessively High Taxes?" *American Economic Review* 92 (1): 363–70.

Keen, Michael J., and Christos Kotsogiannis. 2004. "Tax competition in federations and the welfare consequences of decentralization." *Journal of Urban Economics* 56 (3): 397–407.

Kleven, Henrik Jacobsen, Camille Landais, and Emmanuel Saez. 2013. "Taxation and International Migration of Superstars: Evidence from the European Football Market." *American Economic Review* 103 (5): 1892–1924.

Kleven, Henrik Jacobsen, Camille Landais, Emmanuel Saez, and Esben Schultz. 2014. "Migration and Wage Effects of Taxing Top Earners: Evidence from the Foreigners' Tax Scheme in Denmark." *Quarterly Journal of Economics* 129 (1): 333–78.

Lehmann, Etienne, Laurent Simula, and Alain Trannoy. 2014. "Tax me if you can! Optimal Nonlinear Income Tax Between Competing Governments." *Quarterly Journal of Economics* 129 (4): 1995–2030.

McClure, Matt. 2013. "High-income earners use province to save on taxes." *Calgary Herald*, April 30. https://www.pressreader.com/canada/calgary-herald/20130430/281629597770562.

Milgrom, Paul, and John Roberts. 1990. "Rationalizability, Learning, and Equilibrium in Games with Strategic Complementarities." *Econometrica* 58 (6): 1255–77.

Milligan, Kevin, and Michael Smart. 2013. "The Devolution of the Revolution: Taxation of High Incomes in a Federation." http://faculty.arts.ubc.ca/kmilligan/research/papers/exeter-v2.pdf.

Milligan, Kevin, and Michael Smart. 2015. "Taxation and top incomes in Canada." *Canadian Journal of Economics* 48 (2): 655–81.

Milligan, Kevin, and Michael Smart. 2019. "An Estimable Model of Income Redistribution in a Federation: Musgrave Meets Oates: Dataset." *American Economic Journal: Economic Policy.* https://doi.org/10.1257/pol.20160600.

Mintz, Jack, and Michael Smart. 2004. "Income shifting, investment, and tax competition: Theory and evidence from provincial taxation in Canada." *Journal of Public Economics* 88 (6): 1149–68.

Moretti, Enrico, and Daniel J. Wilson. 2017. "The Effect of State Taxes on the Geographical Location of Top Earners: Evidence from Star Scientists." *American Economic Review* 107 (7): 1858–1903.

Musgrave, Richard A. 1959. *The Theory of Public Finance: A Study in Public Economy.* New York: McGraw-Hill.

Musgrave, Richard A. 1971. "Economics of Fiscal Federalism." *Nebraska Journal of Economics and Business* 10 (4): 3–13.

Oates, Wallace E. 1972. *Fiscal Federalism.* New York: Harcourt.

Rochwerg, Martin J. 2004. "Recent Developments in Estate Planning: The Alberta Advantage When Using Trusts." https://www.millerthomson.com/assets/files/article_attachments/Recent_Developments_in_Estate_Planning_The_Alberta_Advantage_when_using_Trusts.PDF.

Saez, Emmanuel. 2004. "Reported Incomes and Marginal Tax Rates, 1960–2000: Evidence and Policy Implications." In *Tax Policy and the Economy*, Vol. 18, edited by James Poterba, 117–74. Cambridge, MA: MIT Press.

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

Saez, Emmanuel, and Michael R. Veall. 2005. "The Evolution of High Incomes in Northern America: Lessons from Canadian Evidence." *American Economic Review* 95 (3): 831–49.

Slemrod, Joel, and Wojciech Kopczuk. 2002. "The optimal elasticity of taxable income." *Journal of Public Economics* 84 (1): 91–112.

Solé-Ollé, Albert. 2013. "Regional tax autonomy in Spain: 'Words' or 'deeds'?" https://english.oim.dk/media/13655/albert-sol%C3%A9-oll%C3%A9.pdf.

Veall, Michael R. 2012. "Top income shares in Canada: Recent trends and policy implications." *Canadian Journal of Economics* 45 (4): 1247–72.

Young, Cristobal, and Charles Varner. 2011. "Millionaire Migration and State Taxation of Top Incomes: Evidence from a Natural Experiment." *National Tax Journal* 64 (2): 255–83.

Young, Cristobal, Charles Varner, Ithai Z. Lurie, and Richard Prisinzano. 2016. "Millionaire Migration and Taxation of the Elite: Evidence from Administrative Data." *American Sociological Review* 81 (3): 421–46.