An estimable model of income redistribution in a federation: Musgrave meets Oates*

Kevin Milligan†
and
Michael Smart‡

August 27, 2017

Abstract

We develop a theory of cross-border income shifting in response to personal taxation, and examine its implications for the revenue potential and excess burden of personal taxes at the subnational level. We estimate the elasticity of tax avoidance and of cross-border tax base shifting using data on top income shares for Canadian provinces, finding that interprovincial shifting accounts for about two-thirds of total tax avoidance. We then propose a model demonstrating that a properly-chosen federal tax rate can offset the horizontal fiscal externality, allowing decentralized subnational tax rates to replicate the national welfare optimum.

*Thanks for comments from Kory Kroft, Michael Devereux, members of the Oxford Centre for Business Taxation, and participants at the 2016 PEUK conference.
†University of British Columbia, kevin.milligan@ubc.ca
‡University of Toronto, msmart@chass.utoronto.ca

1 Introduction

According to the received wisdom in public finance since Musgrave (1959), national governments should be assigned responsibility for redistribution of income. Decentralized income taxation is apt to 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 inter-state spillovers are small.

In this paper, we apply an Oatesian perspective to the 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. On the other hand, taxpayers may migrate or shift income between states of the federation in response to internal tax differentials, which favors national taxation. To analyze these effects, we develop a simple model of top-bracket income taxation where tax base is mobile amog states of a federation. We use the model to derive 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 round a decentralizing reform in Canada.

The paper proceeds as follows. Section 2 introduces our model of tax avoidance and tax shifting, shows how optimal tax policies relate to observable elasticities, and examines the implications of the model for fiscal federalism. Economists since Feldstein (1999) have understood that elasticities of taxable income (or semi-elasticities in this case) are informative about the marginal excess burden of taxation. On the other hand, Gordon and Slemrod (1998) and Chetty (2009) have emphasized that when the ETI reflects a transfer of income between two revenue sources, rather than simple tax avoidance behavior, then it is not generally a sufficient statistic for social welfare. 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, then explore the implications of this framework in a formal extension of our model that considers optimal tax setting within a federation.

Our findings at first suggest a Musgravian perspective on fiscal federalism: since a tax increase in one state leads to tax base shifting into others, there is a positive fiscal externality among state governments that favors national redistribution. On the other hand, according to the Oatesian perspective, differences in the yield of top bracket taxes (which are large in our data) require differentiated tax rates and so favor decentralization over uniform national taxation.

We proceed to analyze these tradeoffs in a formal model. As benchmarks, we consider a

[1] In this paper we mostly refer to sub-national jurisdictions as 'states' to keep the case general, but use the term 'provinces' when referring to our specific empirical exercise involving Canada.

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 (2012) 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.

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 asymmetries among states, in which the "Oatesian" tradeoff 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.

Section 3 and 4 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 non-tax revenues—but not in others. Indeed, the reform arguably created an internal tax haven with substantially greater potential for tax-motivated income shifting for high-income taxpayers in all provinces than before. We exploit this variation in order to estimate the shifting and avoidance components of the aggregate ETI. Our preliminary estimates suggest that interstate income shifting is large. However, we find that about two-thirds 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. In this vein, Kleven Landais and Saez (2013) and Kleven et al. (2014) find substantial effects of high-income tax rates on migration of "superstars" in Europe. These results complement and reinforce the earlier conclusions of Feldstein and Wrobel (1998) that state income tax differences lead to offsetting differences in equilibrium pre-tax 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 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.

[Section 5: simulation results. Section 6 concludes.]

2 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

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

where

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

is the average tax rate for all states in the nation, weighted by income shares, 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, (1)–(2) implicitly define the response of incomes to tax rates in the federation. In (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 *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 *shifting behavior*, and define the corresponding shifting semi-elasticity

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

which we assume are common for all states i.

The assumption that tax bases depend only on own tax rates and absolute tax differentials is of course restrictive, but it is a common assumption in the tax competition literature.[2] Appendix 1 to this paper shows how such tax base functions can be derived from an optimizing model of individual taxpayer behavior with shifting and avoidance technologies. Our model could be generalized, e.g. to one featuring asymmetric state tax competition interactions within the federation, but identifying such a model with our data would be difficult.

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 hetreogeneous tax rates $t_i = \tau_i - T$. So state government tax revenues are

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

and federal tax revenues are

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

Summing these expressions, national tax revenues are

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

2.1 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 will show that (e_a, e_s) are "sufficient statistics" for welfare analysis (Chetty, 2009) in this model. We will then contrast what can be achieved in a (Musgravian) unitary tax system, compared to an (Oatesian) equilibrium decentralized system.

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;

[2] This approach was for example used in a two-country model by Keen (2001). We adopt the same approach, but extend it to $J \geq 2$ jurisdictions.

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 $\tau_i = t_i + T$.

We assume that state governments maximize state tax revenues, disregarding the impact of their decisions on revenues of other states and the federal government.[3] In contrast, the welfare objective from the national perspective is the sum of revenues of the federal and all state governments. To understand better the difference between these two objectives in the presence of avoidance and shifting effects, we will characterize the marginal revenue of a tax increase in state i, from the state and national perspectives. As usual (Saez, 2002), 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 (1), (2) and (5), we can establish:

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

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

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

Proof. See appendix.

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 (6) from (7), state and national effects of tax increases differ by

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

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

[3] 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 disussion in Diamond and Saez (2011).

are too low from a national perspective, the vertical externality serves as a corrective, raising equilibrium tax rates.[4] It is this interplay between horizontal and vertical externalities that is the key to our results on optimal tax assignment in what follows.

Using (7), we can characterize optimal tax rates as follows.

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

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

Proof. See appendix.

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 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.

2.2 Optimal tax assignment in the federation

Proposition 2 characterizes tax policies that maximize national revenues. The problem is thatb a national government attempting to implement these policies would violate the Oatesian constraint that national policies be uniform across the nation. Respecting the Oatesian constraint, what can be achieved be 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 (6), it follwos that Nash equilibrium tax rates under full decentralization satisfy

$$\tau_i^D = \frac{\theta_i}{e_a + e_s} \tag{10}$$

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)[5] 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. Indeed, when $T = 0$, Proposition 2 shows that Nash equilibrium state tax rates must be lower than the levels that maximize national revenues.

[4]This despite the general result that federal and state tax rates ar enot strategic complemenets - see Keen (2002).

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

Unitary taxation. In contrast, consider unitary taxation, in which a national government sets a uniform tax rate in all states to maximize national revenues: the Oatesian perspective on centralization.[6]. Then tax differentials between states are identically zero, there is no cross-state shifting, and the single tax rate τ is chosen to maximize

$$\max \sum_j \tau [y_j(\tau, 0) - k_j] \tag{11}$$

Differentiating (11) immediately establishes:

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

$$\tau^U = \frac{\bar{\theta}}{e_a} \tag{12}$$

where

$$\bar{\theta} = \sum_j \omega_j(\tau^U, \ldots, \tau^U)\theta_j$$

is the weighted average of the individual state yield parameters, evaluated at the unitary optimum.

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 (9) and (12) shows that the unitary tax policy is suboptimal whenever tax yield parameters θ_j differ among states. 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 (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.

Consider a two-stage "Stackelberg" game in which the federal government first chooses T.[7] 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

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

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

[7][Comment on Stackleberg vs Nash].

that represent a fixed point of the state best response functions.[8] Setting marginal revenue to zero in (6), the Nash equilibrium tax rates in the subgame satisfy

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

Given the equilibrium state tax rates, what can be achieved through 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, 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 $R^N(\tau_1,\ldots,\tau_J)$ subject to (15). 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:

Proposition 4 *If the federal tax rate is*

$$\hat{T} = \bar{\theta}\frac{e_s/e_a}{e_a + e_s} \tag{14}$$

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 $\omega_i \to 0$. For finite states with $\omega_i > 0$, (13) shows that large states internalize more of the shifting externality, choosing lower tax rates ceteris paribus.[9] 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 (14) we see that:

[8] 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 a largest and smallest Nash equilibrium of the tax competition subgame that are increasing in the federal tax rate.

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

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.

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.

3 Personal taxation in Canada

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 2014, provincial personal income tax revenues were \$81 billion, or 39% of combined federal–provincial revenues. Provinces generally apply progressive rate structures to taxable incomes, with top marginal tax rates currently (2016) ranging from 14.7% to 21.0% in the various provinces—although rates had been as low as 10.0% until 2015. In all provinces except Quebec, tax rates are applied to a common (federal) definition of taxable income, and taxes are collected on behalf of provinces by federal tax authorities.

A major reform to provincial taxation occurred in 2000/2001. Previous to this reform, provinces (outside Quebec) set their income taxes as a fraction of "basic federal tax".[10] These rates ranged in 1995 from 69 percent in Newfoundland and Labrador to 45.5 percent in Alberta. An increase in this provincial tax rate affected all taxpayers proportionately. Provinces at that time also had the ability to add income surtaxes for high earners in order to further manipulate the tax liability and marginal tax rates of those at the high end of the income distribution. A reform from this previous "tax-on-tax" to the current "tax-on-income" system was implemented over 2000-2001. Under the new system, provinces could set their own brackets and rates, given the federally-determined taxable income. Thus the tax-on-income reform afforded provinces more flexibility in redistribution, and particularly in the ability to operate a tax system with less progressivity than the federal one.

While some provinces implemented the new tax-on-income system by choosing tax brackets and rates that produced tax liabilities identical (or very close) to the previous tax-on-tax system, the province of Alberta did not. Instead, Alberta adopted a flat-rate income tax with a top marginal rate of 10%, eschewing brackets altogether. The top rates implemented by other provinces at that time ranged between 16 and 20%.

[10] The "basic federal tax" was the tax liability generated by the federal tax rate and tax bracket calculation. Basic Federal Tax excludes special 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 tax differentials appear to have led to new strategies for shifting taxable income to Alberta. One strategy, widely promoted by tax advisers,[11] was for high-income taxpayers in other provinces to transfer personal assets to an inter vivos trust resident in Alberta; income received by the trust is taxed at the lower tax rate. The scheme appears especially popular in cases where a closely held corporation has substantial undistributed earnings or unrealized capital gains, so that tax savings exceed transaction costs of the trust. A trust is deemed resident in Alberta if a majority of its trustees reside there. In the words of one tax adviser marketing the scheme, "if the taxpayer does not know any Albertan residents they can choose an Alberta law firm or financial institution to act as trustee."[12]

For years, these trust arrangements attracted little notice from federal tax authorities, perhaps because provincial residency affects provincial but not federal tax revenues. In 2010, however, following a critical report from the federal Office of the Auditor-General, the Canada Revenue Agency announced a new initiative to verify the residency of Alberta trusts and to scrutinize certain distributions from trusts to beneficiaries. As well, the province of Quebec, which administers an independent income tax law, requires resident taxpayers to self-report income received from non-resident trusts and to pay tax on it at home. Most recently, a ruling of the Supreme Court of Canada[13] has tightened rules on residency of trusts in a way that seems likely to restrict the future use of Alberta trusts for interprovincial tax avoidance. These recent developments however seem unlikely to affect behavior much during the 1988-2013 period covered by our data set.

While Alberta trusts appear to have been used extensively to shift capital income to Alberta, they are practical only for the highest-income taxpayers, and they were not in themselves effective for shifting employment income out of high-tax provinces. An alternative tax planning strategy is simply for the taxpayer to declare residency in Alberta. Income taxes in Canada are payable in the province of residence of a taxpayer, irrespective of the location of employment. Residency for tax purposes is determined based on the taxpayer's principal residence on December 31 of each year. Moreover, federal tax authorities may not closely scrutinize provincial residency claims.[14] 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.

In short, the Canadian federal system is a useful testing ground for our theory of subnational income shifting in response to personal taxation. The anecdotal evidence points to the province of Alberta as a likely "onshore tax haven" towards which most domestic tax planning strategies were directed, at least in the period covered by our data set. Our empirical strategy incorporates this possibility in a framework that allows both for pure tax avoidance and for income shifting across provinces.

The tax rates for our analysis come from the Canadian Tax and Credit Simulator (CTaCS; see

[11] See, e.g., Martin Rochwerg, 2004, "Recent developments in estate planning: The Alberta advantage when using trusts," mimeo, Miller Thomson LLP, and Tim Cestnick, "Consider an Alberta trust even if you don't live there," *The Globe and Mail* (Toronto), June 21, 2003.

[12] Quoted from Cunningham Chartered Accountants, "Alberta and the use of Alberta trusts – the next tax haven?"

[13] Fundy Settlement v. Canada, 2012 SCC 14

[14] See, e.g., "High-income earners use Alberta to save on taxes," *Calgary Herald*, April 30, 2013.

Milligan 2016), which provides a calculation of income tax liability given a province, year, and a vector of income and family structure inputs. The CTaCS calculator is available for the years 1962 to 2016, which spans the years 1988 to 2013 that we use from the income data. Because our focus for this paper is top incomes, Canada's vast system of targeted refundable tax credits does not affect our estimates, as individuals at income levels observed in the top fractiles are out of the eligible range for these credits. We are interested in marginal tax rates rather than tax liabilities. 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.

Observed tax rates at the provincial level can vary both because incomes differ across provinces and because statutory tax rates differ. In order to isolate the effect of the statutory tax rates, we perform our tax simulations using a common 'synthetic' income distribution, rather than the observed incomes for each given province and year. Specifically, we use income cutoffs from the national distribution for the year 2000, and then adjust these income cutoffs according to CPI to create data for each year. This common set of incomes is then put through CTaCS for each province and year combination.

Our primary focus is the 99th percentile cutoff which we use to obtain the tax rate for those with incomes in the top one percent. Over the time period covered by our data the bracket threshold for facing the highest tax rate was lower than the 99th percentile cutoff in almost all cases. So, we do not need to confront concerns about estimating the tax bracket of top one percent earners as they are almost all in the top bracket.[15]

Kopczuk and Slemrod (2002) caution against thinking of taxable income elasticities as invariant parameters. In particular, we should expect the elasticity to depend on the definition of taxable income—which changes when income definitions or exemptions and deductions are changed. In Canada, a major tax reform in 1988 which changed the federal personal income tax base substantially. Since 1988, the federal tax base has been quite stable. Nine of the ten provinces allow the federal government to collect taxes on their behalf through a tax collection agreement. A core element of these tax collection agreements is the use of the federal tax base for calculating provincial tax liabilities. The one exception is Quebec, although the tax base differences in practice are modest. Taking this into account, our analysis focuses on the years of stability for the tax base, spanning 1988 to 2013.

We graph the provincial high income tax rates by province and year in Figure 1. In the regression analysis we use the combined federal-provincial marginal tax rate. However, in order to highlight the provincial variation we graph here only the provincial component. The bottom line in the graph shows the high tax rate for Alberta. The tax rate in Alberta hovered around 15 percent through most of the 1990s, shifting only slightly. Heading out of the 1990s, Alberta first dropped its overall tax rate slightly in 1998, and then removed a surtax in 2000.[16] The sharp drop

[15] The sole exception is Ontario, which introduced a new tax bracket starting at $500,000 in 2012, while the top one percent threshold was $225,600.

[16] The Alberta tax decrease in 1998 was from 45.5 percent of basic federal tax to 44.0 percent. Given the top federal rate of 29 percent, this dropped the effective Alberta basic rate by less than half of a percentage point. The surtax removed in 2000 was 8 percent of basic Alberta tax, which works out to a drop in the effective marginal rate of just over one percentage point.

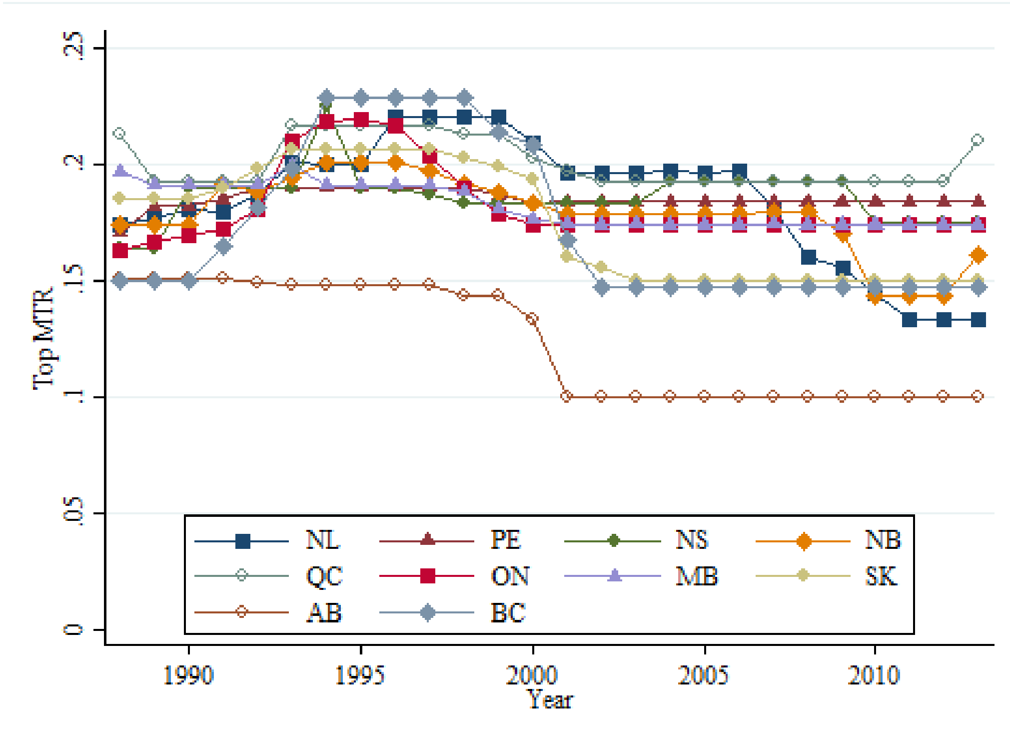

Figure 1: Top tax rates, 1988-2013

Source: Calculations using CTaCS. Shown is the tax rate at the 99th percentile threshold for each province through time.

in 2001 occurred as the province shifted to a flat 10 percent tax rate. This tax rate has remained unchanged through the subsequent years (although it has moved upward since 2015).

The top tax rates in the other provinces are less distinct in Figure 1. This is a virtue for our analysis, as the tangle of lines reflect substantial changes in tax policy at the provincial level during this 26 year period. As one example, BC 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. As another example, the top rate in Newfoundland and Labrador fell from the highest in the first half of the 1990s to the second lowest by 2011.

The substantial variation in provincial tax rates can be summarized using a regression of the top marginal tax rate on dummies for each province and each year in our data. The R-squared from this regression is 0.846, which suggests that the within-province through time variation accounts for around one sixth of the total variation in tax rates. This helps to justify our empirical strategy which exploits this residual within-province through time variation—if the within-province across time variation were smaller, it would be hard for our regressions to identify the impact of tax rates on reported incomes using this empirical strategy.

4 Empirical strategy

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

$$\log b_{it} = \alpha_i + \delta_t - e_u \tau_{it} + e_s \bar{\tau}_{-it} + x'_{it}\beta + \epsilon_{it} \tag{15}$$

where τ_{it} is the top marginal tax rate in province i and year t (the "own" tax rate), $\bar{\tau}_{-it}$ 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 below. In our empirical implementation, we have a panel of taxable incomes and marginal tax rates for taxpayers in ten provinces and 26 years.

We adopt the "share analysis" approach, common in the empirical literature on taxable income elasticities,[17] 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 one per cent). The share approach may be derived from the model of individual behavior (15) under the assumption that taxpayers in the top quantile are influenced by top marginal tax rates according to (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 (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.

Furthermore, (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 more robust inference of taxable income elasticities than is typically possible with the share analysis approach, which typically uses national-level data, and only time series variation in tax rates. 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.

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 in (15) 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 the difference-in-difference estimates from (15). Furthermore, if the own tax rate τ_{it} is endogenous, then contemporaneous spatial correlation in unobservable shocks could also cause the neighbor average tax rate $\bar{\tau}_{-it}$ to be endogenous in (15).

For this reason, while we report OLS estimates below, we focus on instrumental variables estimates of (15), derived from the effects of the decentralizing reform in 2000, described in Section 3 above. As noted, the reform afforded provinces more control over the progressivity of

[17] See Saez, Slemrod and Giertz (2010) for a survey.

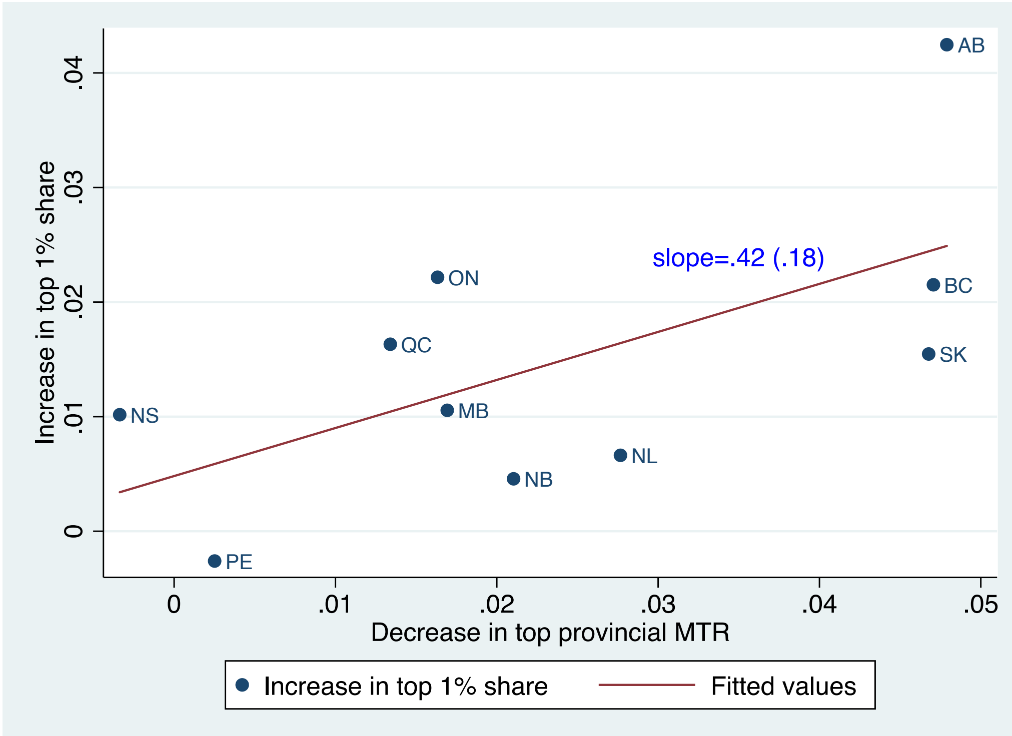

Figure 2: Effects of the reform on tax rates and income shares

the income tax schedule and led to declines in top tax rates in some provinces – notably Alberta – but not in others. Because the post-2000 tax changes were asymmetric, the resulting variation in tax rates is distinct from the year fixed effects and permits identification of tax elasticities in (15).

To give a sense of the impacts of the reform, Figure 2 plots the change in top tax rates in each province from the pre-reform (1988-99) to the post-reform (2001-13) periods, against the corresponding change in the top 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. Inferences based on this approach are still subject to endogeneity concerns, however, because post-reform tax changes may again be correlated with omitted determinants of the tax base.

We therefore look for a quasi-experimental approach to help pin down the counterfactual evolution of tax bases following the reform. 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. Several Canadian provinces (including Alberta) derive a significant portion of their budgets from non-tax resource revenues – chiefly, the sale of oil and gas leases on public lands and royalties derived from mining and other extractive industries. So access to resource revenues of some provinces creates arguably exogenous

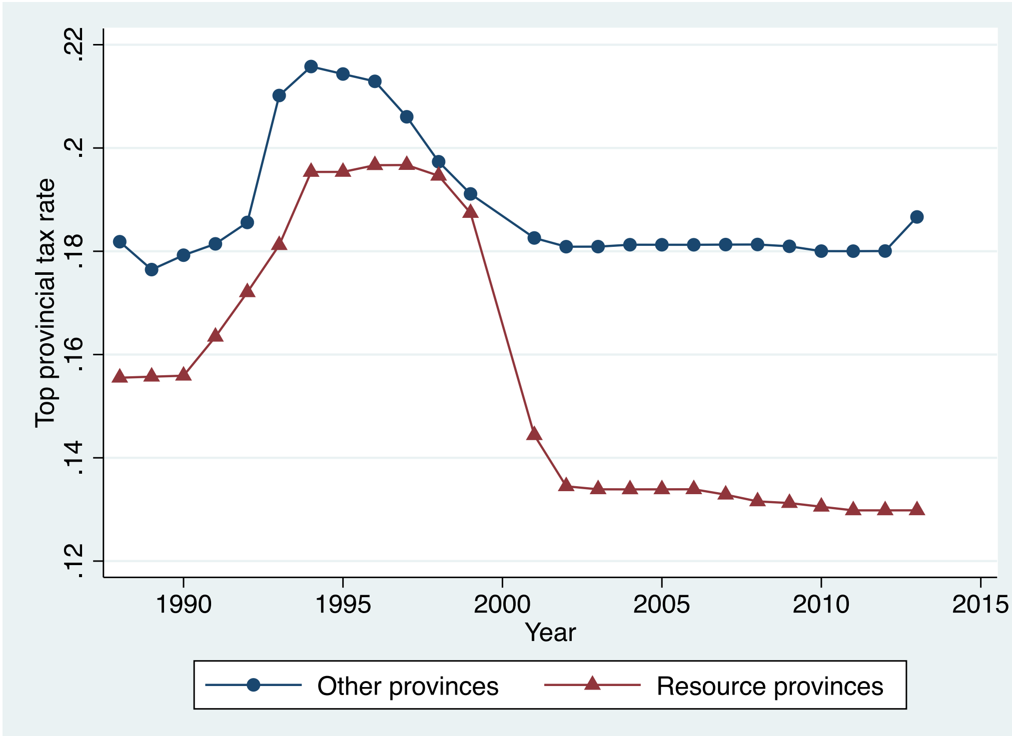

Figure 3: The evolution of top tax rates

variation in the extent to which an individual province responded to the decentralization reform by changing the top tax rate. That is, our instrumental variables strategy is based on the notion that the extra revenues sources in the more resource-intensive provinces afforded them more freedom to change tax rates in the new "tax-on-income" system.

Consistent with this view, we note that correlation between tax reductions in Figure 2 and the share of resource revenues in provincial budgets is high. We (somewhat arbitrarily) define "resource provinces" as those where resource revenues were at least 10 per cent of total revenues on average in the 1982-2013 period. There are four such provinces by this definition,[18] and they exhibited the largest post-reform tax rate declines among the 10 Canadian provinces, as shown in Figure 2. In this sense, the resource 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.

In Figure 3 we depict the time series of top provincial tax rates on average for resource and other provinces, which serve as "treatment" and "control" groups for our analysis. Observe that tax rates evolved similarly in the two groups prior to the reform, although somewhat lower in resource provinces, but fell substantially in resource provinces relative to the control group following the reform. In Figure 4 we depict the corresponding evolution of log top income shares

[18]They are Alberta (AB), British Columbia (BC), Saskatchewan (SK), and Newfoundland and Labrador (NL). These are the only four provinces with non-trivial oil and gas production.

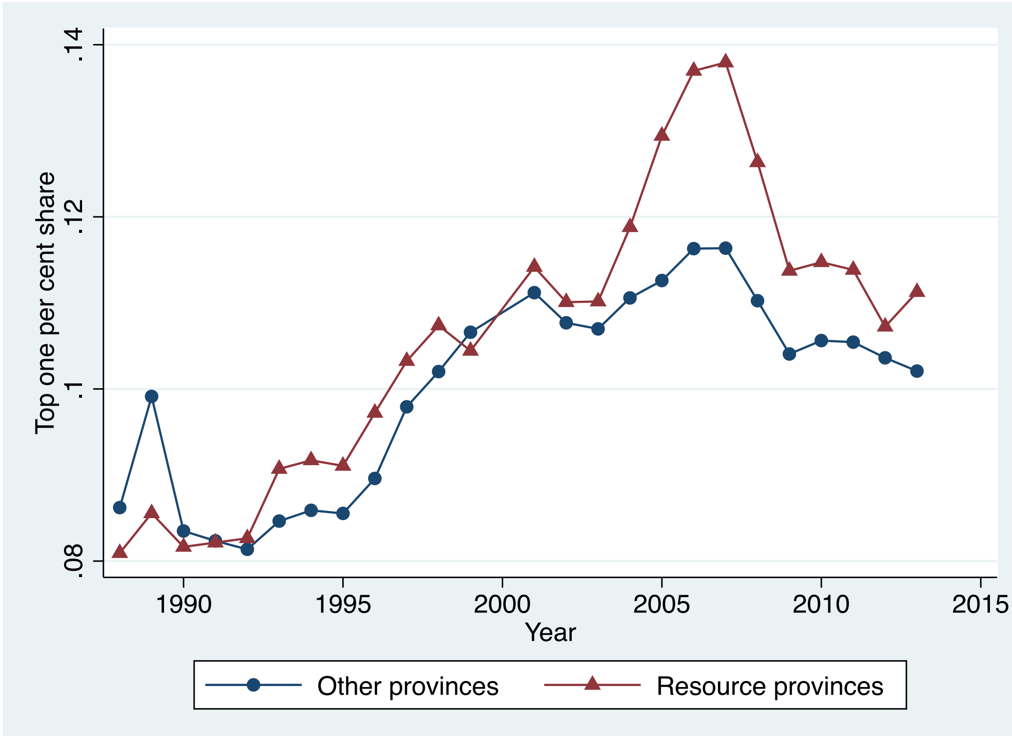

Figure 4: The evolution of top income shares

in the treatment and control provinces. Again, the two series evolve very similarly in the pre-reform period, giving support for the validity of our approach. By 2003, the two series begin to diverge, with top income shares rising in resource provinces relative to other provinces. The difference begins to decline following 2007, suggesting some influence of commodity booms and busts on the income distribution, but the difference remains positive throughout the post reform period.

In what follows, we therefore report instrumental variables estimates of tax elasticities using as our main instrument each province's average share of resource revenues in total revenues, multiplied by a dummy variable for post-reform years ($\text{RES}_i \times \text{POST}_t$). To account for the possible direct effect of resource revenues on tax rates and income distribution, as suggested by Figure 3, we include RES_{it} as a control variable throughout the analysis. Thus the instrument captures the differential effect of the 2000 reform on provinces depending on the average fiscal room created by resource revenues, while we control for the direct effect of resource revenues themselves on top income shares.

5 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

VARIABLES	(1) lshare99	(2) lshare99	(3) lshare99
Own tax rate	-2.31	-2.18	-2.18
	[0.93]	[0.43]	[0.42]
Neighbour tax rate			0.08
			[0.92]
Log total income		0.93	0.93
		[0.09]	[0.09]
RES*Energy price		-0.12	-0.11
		[0.02]	[0.02]
Observations	250	250	250
R-squared	0.25	0.61	0.61
Number of prov	10	10	10
Avoidance Elasticity	1.21	1.14	1.10
Shifting Elasticity			.04

All specifications include year and province fixed effects. Robust standard errors clustered by province.

Table 1: OLS estimates

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.[19] For the simple difference in difference specification, the estimated semi-elasticity is -2.31. Elasticity of taxable income(ETI) 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 about 1.2, 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 Statistics Canada's energy price index, multiplied by the share of resource royalties in total provincial revenues on average for the 1982-2013 period. This resource price index therefore captures the asymmetric way that exogenous resource

[19] 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. Our implementation of the shares approach here is simpler, but we now include energy price controls absent in our previous work. 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 inter-provincial shifting in the next table.

VARIABLES	(1) lshare99	(2) lshare99	(3) lshare99
Own tax rate	-2.32	-2.26	-2.41
	[0.78]	[0.41]	[0.44]
Neighbour tax rate			1.70
			[0.94]
Log total income	0.93	0.93	0.92
	[0.09]	[0.09]	[0.10]
RES*Energy price	-0.12	-0.12	-0.11
	[0.02]	[0.02]	[0.02]
R-squared	0.61	0.61	0.59
Instruments	RES*POST	RES*POST	RES*POST
	-	NDP	NDP
First stage F: Own tax rate	30.45	39.52	39.52
Neighbour tax rate	-	-	135.11
Avoidance Elasticity	1.21	1.18	.37
Shifting Elasticity			.89

All specifications include year and province fixed effects. Robust standard errors clustered by province.

Table 2: IV estimates

price shocks affect provincial economies, depending on their resource endowments. Controlling for these variables markedly improves the fit of the regression, but it leaves the estimated own tax elasticity nearly unchanged.

The next two columns of Table 1 report estimates from specifications including the neighbor average tax rate, with and without the parametric control variables. In both cases, the neighbor tax effect is insignificantly different from zero, and other coefficients are essentially unchanged. As expected, therefore, the OLS estimates show little evidence of cross-jurisdiction income shifting.

5.1 Instrumental Variables

We therefore 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 explained above, the instrument for the own tax rate τ_{it} is $RES_i * POST_t$. The F statistic for significance of the excluded instrument in the first stage regression is 30.4. (The coefficient on the instrument in the first stage, unreported in the table for brevity, is -0.0011, indicating that tax rates fell 1.1 percentage points more in provinces with a resource share of ten per cent of total revenues, compared to a province with no resource revenues.) The estimated semi-elasticity is -2.32, similar to the corresponding OLS estimate of column (2) of Table 1.

Recall that, if τ_{it} is endogenous in (15), and there is contemporaneous spatial correlation in tax rates, then $\bar{\tau}_{-it}$ is endogenous also.[20] If the own tax rate is negatively correlated with omitted variables increasing the tax base, as suggested 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 (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. Since $RES_i \times POST_t$ 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 $RES_i \times POST_t$ as the additional instrument necessary to identify (15) when both tax rates are endogenous. But the two instruments are too closely correlated to yield independent variation.

Analyzing a very similar issue in the context of estimating peer effects of educational attainment, Acemoglu and Angrist (2001) argue 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. In our context, one 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 per cent of the province–year cells in our data.[21] Column (2) presents two-stage least squares estimates of (15), where the neighbor tax rate is excluded, and the weighted average of $RES_i \times POST_t$ 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 39.5 and the first stage coefficient on NDP_{it} is 0.017, 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 -2.26, 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 neighbor average tax rate, where the instruments are the same as in column (2). In this case, the estimated shifting semi-elasticity (the coefficient on the neighbor average tax rate) is large and significant. The implied shifting elasticity (at the mean tax rate in the sample) is 0.89, whereas the pure avoidance elasticity at 0.37 represents only about one-third of the overall elasticity of taxable income. Thus our results suggest that while provincial tax bases are highly responsive to unilateral tax changes, much of this elasticity can be accounted for by the shifting of income between provinces. In contrast, our estimates suggest that a federal or coordinated provincial tax rate increase would have a relatively small effect on the high income tax base.

	(1)	(2)	(3)	(4)
VARIABLES	lshare90	lshare95	lshare99	lshare999
Own tax rate	-0.65	-1.53	-2.41	-3.11
	[0.11]	[0.31]	[0.44]	[1.14]
Neighbour tax rate	0.87	0.75	1.70	3.37
	[0.62]	[0.81]	[0.94]	[1.21]
Log total income	0.30	0.53	0.92	1.16
	[0.05]	[0.05]	[0.10]	[0.31]
RES*Energy price	-0.01	-0.05	-0.11	-0.12
	[0.01]	[0.01]	[0.02]	[0.08]
Observations	250	250	250	198
R-squared	0.65	0.60	0.59	0.45
Number of prov	10	10	10	8

All specifications include year and province fixed effects. Robust standard errors clustered by province.

Table 3: Alternative income thresholds

5.2 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 a variety of different quantiles of the income distribution. 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 ten per cent 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 e_u is -0.66 for the top ten percent threshold; the neighbor tax effect in this case while larger in magnitude is insignificantly different from zero. 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 one per cent 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.

[20] Indeed, $\bar{\tau}_{-it}$ might be a valid instrument for τ_{it} in (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.

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

5.3 Simulations

Our estimates of the shifting elasticity our large, which may appear inconsistent with the relatively high levels of redistribution in provincial personal tax systems, where top marginal rates currently range from 14.7 to 21 per cent—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.[22] These facts appear at first glance to be inconsistent with large shifting responses to taxation—in the Musgravian perspective, provinces would avoid increasing taxes at the top. On the other hand, 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?

The estimation model and formal welfare analysis above take as given a tax base function that embodies avoidance and shifting responses to taxation through the functional form $b_i(\tau_i, \tau_i - \bar{\tau})$. To illustrate its key properties and their quantitative significance, we simulate the model for the case of linear tax base functions:

$$b_i(\tau) = z_i - a\tau_i - s(\tau_i - \bar{\tau}) \tag{16}$$

where a, s are the avoidance and shifting parameters. (In Appendix 1, we show how (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 form of the tax base functions (16) make it easy to illustrate key features of the mode, and to derive closed-form solutions for tax rates and tax revenues.

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 tradeoffs 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

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

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

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

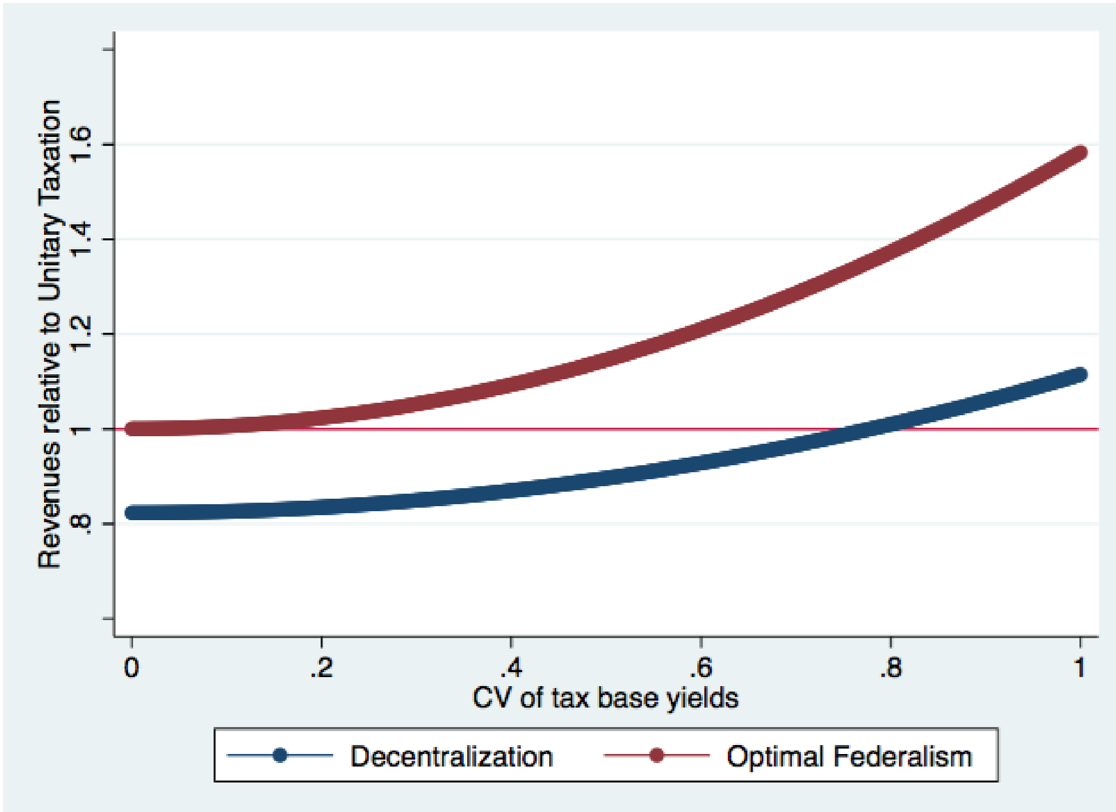

Figure 5: Simulated revenues under alternative tax systems

system are (see the appendix):

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

where

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

is the coefficient of variation of δ_i among states, the "right" 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 4, 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

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

Figure 5 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 estimates from column (3) of Table 2, i.e.

$$e_a = 0.7 \qquad e_s = 1.7$$

and normalize revenues from unitary taxation to one.

Examining the figure, and 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. Moreover, $\bar{R}(\tau^N)$ and $\bar{R}(\tau^D)$ are both increasing in the coefficient of variation γ (the former at a faster rate), since greater heterogeneity in yields creates more-than-proportionate gains in optimal differentiated taxation. The losses due to full decentralization are increasing in the degree of interstate tax competition (as measured by s/a) regardless of whether there is cross-state heterogeneity or not. Finally, 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.

6 Concluding remarks

This paper studies a model of tax avoidance that incorporates interstate shifting of income for tax purposes. The model yields estimable equations which we take to data on tax rates and reported incomes from Canadian provinces. We exploit subnational variation in income tax rates to identify both a shifting and an avoidance elasticity. Using a traditional approach that considers only the own-province tax rate, we find elasticity estimates that are consistent with the previous literature. When we allow for both inter-provincial shifting and for avoidance, we find evidence that both components exist, with statistically and economically relevant responses. Our preferred specification reveals a shifting response that accounts for about two-thirds of total tax avoidance.

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 engage in a "race to the bottom" 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 1: A structural model of taxpayer avoidance and 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 $\tau_i^s = \tau_i - T$ 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 $j = 1, \dots, J$, 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:[23]

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

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

$$R_i(\tau) = \tau_i(b_i(\tau) - k_i) = \tau_i(\delta_i - a\tau_i - s(\tau_i - \bar{\tau})) \tag{20}$$

Suppose there are J states with equal populations of high-income taxpayers. National revenues per taxpayer can be computed (summing (20) and dividing by J) to be

$$\bar{R}(\tau) = \bar{\tau} \cdot \bar{\delta} + \text{cov}(\tau, \delta) - a\bar{\tau}^2 - (a+s)\,\text{var}(\tau) \tag{21}$$

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.[24]

The revenue expressions (20) and (21) facilitate analysis of optimal taxation in the various cases of unitary taxation, optimal federalism, and full decentralization, as these are defined in

[23] This is a model of income shifting commonly used in the literature on international tax avoidance; see e.g. Mintz and Smart (2004).

[24] That is, $\text{var}(\tau) = J^{-1}\sum \tau_j^2 - \bar{\tau}^2$, and $\text{cov}(\tau, \delta) = J^{-1}\sum \tau_j \delta_j - \bar{\tau} \cdot \bar{\delta}$.

the main text. In the case of unitary taxation, $\tau_i^U = \tau$ for all i, and 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

$$R(\tau^U) = \frac{\bar{\delta}^2}{4a} \tag{22}$$

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 federal government moves first and sets the optimal federal tax rate. Differentiating (21) and rearranging, we find that in the linear–quadratic case the optimal tax rates are

$$\tau_i^N = \frac{\bar{\delta}}{2a} + \frac{\delta_i - \bar{\delta}}{2(a+s)} \tag{23}$$

Substituting into (21), national average revenues at the federal optimum are

$$\bar{R}(\tau^N) = \frac{\bar{\delta}}{4a}\left[1 + \frac{2a}{a+s}\gamma^2\right] \tag{24}$$

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 (20), taking the average tax rate $\bar{\tau}$ as given.[25] These satisfy the first-order conditions

$$2(a+s)\tau_i^* = \delta_i + s\bar{\tau} \tag{25}$$

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

$$\tau_i^D = \frac{\bar{\delta}}{2a+s} + \frac{\delta_i - \bar{\delta}}{2(a+s)} \tag{26}$$

Substituting again into (21), national average tax revenues under full decentralization are

$$\bar{R}(\tau^D) = \frac{\bar{\delta}}{4a}\left[1 - \frac{s}{(2a+s)^2} + \frac{a}{a+s}\gamma^2\right] \tag{27}$$

[25] 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.

Appendix 2: Proof of Proposition 2.

Differentiating (5), the optimal tax vector $(\tau_1^N, \ldots, \tau_J^N)$ satisfies

$$\frac{\partial R}{\partial \tau_j} = [(y_j - k_j) + \tau_j^N (y_{j1} + y_{j2}) - \frac{y_j}{Y} \sum_i \tau_i^N y_{i2}] = 0 \quad j = 1, \ldots, J$$

where $Y = \sum_i y_i$. Applying the definition of the parameters (e_a, e_s, θ_j) and defining the national optimal weighted average tax rate

$$\bar{\tau}^N = \sum_i \frac{y_i}{Y} \tau_i^N$$

we can express the first order condition as

$$\frac{\partial R}{\partial \tau_j} = y_j [\theta_j - \tau_j^N e_a - (\tau_j^N - \bar{\tau}^N) e_s] = 0 \quad j = 1, \ldots, J$$

Solving this expression for τ_j^N and then averaging yields marginal revenue from a tax increase in j is

$$\frac{\partial R}{\partial \tau_j} = y_j [\theta_j - \tau_j^N e_a - (\tau_j^N - \bar{\tau}^N) e_s] \tag{28}$$

Setting $\partial R^N / \partial \tau_j = 0$ for all j and rearranging then yields (9).

References

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

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), 535–564

Diamond, Peter, and Emmanuel Saez (2011) 'The case for a progressive tax: From basic research to policy recommendations.' *Journal of Economic Perspectives* 25(4), 165–190

Feldstein, Martin, and Marian Vaillant Wrobel (1998) 'Can state taxes redistribute income?' *Journal of Public Economics* 68(3), 369–396

Gagné, Robert, Jean-François Nadeau, and François Vaillancourt (2010) 'Taxpayers' response to tax rate changes: A Canadian panel study.' Scientific Series, CIRANO, December. Number 2000s-59

Giertz, Seth H (2007) 'The elasticity of taxable income over the 1980s and 1990s.' *National Tax Journal* 60(4), 743–768

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

Gruber, Jonathan, 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 Staff Papers* 45(3), 454–485

__ (2001) 'Preferential regimes can make tax competition less harmful.' *National Tax Journal* 54(4), 757–762

Keen, Michael J, and Christos Kotsogiannis (2002) 'Does federalism lead to excessively high taxes?' *The American Economic Review* 92(1), 363–370

__ (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*

Milgrom, Paul, and John Roberts (1990) 'Rationalizability, learning, and equilibrium in games with strategic complementarities.' *Econometrica* 58(6), 1255–1277

Milligan, Kevin (2016) *Canadian Tax and Credit Simulator.* Version 2016-1

Milligan, Kevin, and Michael Smart (2013) 'Devolution of the revolution: Taxation of high incomes in a federation.' Technical Report, University of British Columbia and University of Toronto

Milligan, Kevin, and Michael Smart (2015) 'Taxation and top incomes in Canada.' *Canadian Journal of Economics* 48(2), 655–681

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), 1149–1168

Musgrave, Richard A. (1959) *The theory of public finance: A study in public economy* (New York: McGraw-Hill)

Oates, Wallace E. (1972) *Fiscal Federalism* (New York: Harcourt Brace Jovanovich)

Saez, Emmanuel (2004) 'Reported incomes and marginal tax rates, 1960-2000: evidence and policy implications.' In 'Tax Policy and the Economy, Volume 18' (MIT Press) pp. 117–174

Saez, Emmanuel, and Michael R Veall (2005) 'The evolution of high incomes in Northern America: Lessons from Canadian evidence.' *The American Economic Review* 95(3), 831–849

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

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

Veall, Michael R (2012) 'Top income shares in Canada: recent trends and policy implications.' *Canadian Journal of Economics* 45(4), 1247–1272

Young, Cristobal, and Charles Varner (2011) 'Millionaire migration and state taxation of top incomes: Evidence from a natural experiment.' *National Tax Journal* 64(2), 255–284