Non-Technical Summary

REVEALING VALUES: APPLYING THE INVERSE-OPTIMUM METHOD TO US STATE TAXES

By Robert Embree

Systems of tax rates represent a major mechanism by which societies implement their social preferences about inequality and redistribution. Observing different income tax rates in two jurisdictions may be a consequence of differences in social preferences; however, it may also reflect differences in income distributions, or forms of progressivity present in other parts of the tax system. Optimal income tax policy tells us that having higher income inequality suggests higher tax rates on the rich for any given social welfare preferences.

The inverse-optimum approach to tax policy reverses the optimal income tax method, and uses existing tax rates to infer marginal social welfare weights for each income level. These weights reveal information about underlying philosophical preferences, and are obtained from a calculation that considers both tax rates and income inequality. Lockwood and Weinzierl have used the inverse-optimum to examine US federal income taxes.

However, existing work using the inverse-optimum method has not been used to examine state-level income taxes, does not include sales and property taxes, and does not examine single and joint filers separately. To analyze state-level differences in preferences, the method must include the sales and property taxes which are so significant at the state and local level.

In this paper, I apply the inverse-optimum method to US state tax data. Using IRS tax data, I calculate the implied weights on each filer type in each income group in every US state, and I make three main contributions to the literature. Firstly, I extend the theory underlying the inverse-optimum method to include the effects of commodity taxation, such as sales and property taxes. Secondly, I calculate effective marginal income tax and commodity tax rates for each state and income level. I use federal, state, and local income taxes, federal payroll

taxes, state and local sales taxes, and state and local property taxes. I consider tax deductions and the tax treatment of different sources of income. Thirdly, I apply my inverse-optimum methodology to both single and joint filers, and to all 50 US states and the District of Columbia in the 2016 tax year.

I find decreasing weights for both types, and differences between the weights for single and joint filers that vary substantially with income. I show how state income tax rates can be explained by differences in both preferences and income inequality. I observe substantial differences in progressivity across states, finding that the ratio of the top-to-median weights varying from a low of 0.48 in New Jersey to a high of 0.72 in Nevada. The inverse-optimum method reveals that tax rates are explained both by the references for redistribution in a state, and by the effect of income inequality on the ability to raise revenue.

REVEALING VALUES:

APPLYING THE INVERSE-OPTIMUM METHOD TO US STATE TAXES

Robert Embree

The inverse-optimum income tax method quantifies social preferences by assigning weights to income groups. However, existing work examines only national-level income taxes, does not include sales and property taxes, and does not examine single and joint filers separately. I extend the theory underlying the inverse-optimum method to include sales and property taxes. Using IRS data, I calculate effective marginal income tax and commodity tax rates for each state and income level. I calculate the implied weights on each filer type in each income group in every US state, finding non-monotonically decreasing weights and substantial differences in social preferences across states.

Keywords: tax policy, marginal social welfare weights, inverse-optimum, state taxes, optimal taxation

JEL Codes: H21, H24, H27, H71, E62

Robert Embree: Department of Economics, University of Toronto, Toronto, ON, Canada (robert.embree@mail.utoronto.ca)

I. INTRODUCTION AND CONTRIBUTION

Systems of tax rates represent a major mechanism by which societies implement their social preferences about inequality and redistribution. Observing different income tax rates in two jurisdictions may be a consequence of differences in social preferences; however, it may also reflect differences in income distributions, or forms of progressivity present in other parts of the tax system. Optimal income tax policy tells us that having higher income inequality suggests higher tax rates on the rich for any given social welfare preferences.

The inverse-optimum approach to tax policy reverses the optimal income tax method, and uses existing tax rates to infer marginal social welfare weights for each income level. These weights reveal information about underlying philosophical preferences, and are obtained from a calculation that considers both tax rates and income inequality.

However, existing work using the inverse-optimum method examines only national level income taxes, does not include sales and property taxes, and does not examine single and joint filers separately. To analyze state-level differences in preferences, the method must include the sales and property taxes which are so significant at the state and local level.

In this paper, I apply the inverse-optimum method to US state tax data. Using IRS tax data, I calculate the implied weights on each filer type in each income group in every US state, and I make three main contributions to the literature. Firstly, I extend the theory underlying the inverse-optimum method to include the effects of commodity taxation, such as sales and property taxes. Secondly, I calculate effective marginal income tax and commodity tax rates for each state and income level. I use federal, state, and local income taxes, federal payroll taxes, state and local sales taxes, and state and local property taxes. I consider tax deductions and the tax treatment of different sources of income. Thirdly, I apply my inverse-optimum methodology to both single and joint filers, and to all 50 US states and the District of Columbia in the 2016 tax year. (Although DC is not a state, I use the collective noun ”states” throughout this paper to mean all 51 jurisdictions.)

I find non-monotonically decreasing weights for both types, and differences between the weights for single and joint filers that vary substantially with income. I show how state income tax rates can be explained by differences in both preferences and income inequality. I observe substantial differences in progressivity across states, finding that the ratio of the top-to-median weights varying from a low of 0.48 in New Jersey to a high of 0.72 in Nevada. The inverse-optimum method reveals that tax rates are explained both by the preferences for redistribution in a state, and by the effect of income inequality on the ability to raise revenue.

This work has substantial policy ramifications. Unusual consequences of the tax code - and possible areas for tax reform - can be revealed when we observe non-monotonic progressions of weights, negative weights, or other atypical patterns. Furthermore, the calculation of marginal social welfare weights provides a better way of comparing preferences across states than does a simple examination of marginal or average rates. This paper thus illuminates the relationship between tax rates, preferences, and inequality.

II. LITERATURE REVIEW

This work extends the literature on the inverse-optimum method of determine marginal social welfare weights. This inverse-optimum method is in turn based on the optimal income tax literature, beginning with Mirrlees (1971). Mirrlees’ method allows us to consider the ways in which tax changes affect both mechanical revenue and individual behaviour. The Mirrleesian approach is common throughout the optimal income tax literature (Tuomala, 1990; Diamond, 1998; Diamond and Saez, 2011; Piketty and Saez, 2013; Dahan and Strawczynski, 2000). The Mirrleesian approach, unlike the Ramsey approach, does not require us to restrict the set of instruments that can be used by the tax policy planner. Saez (2001) incorporates elasticities into the calculation more explicitly; his derivation is more transparent than that of Mirrlees and shows a clearer decomposition of the relevant economic effects. He also uses a system of social marginal welfare weights, saying that “if the government has redistributive tastes, then these weights are decreasing in income” (Saez, 2001).

Several papers attempt to find these marginal social welfare weights starting from real-world income data and a particular set of tax rates. Bourgignon and Spadaro (2012) are the first to outline the methodology of the inverse-optimum method, and they apply it to French income taxes. They derive an inverse-optimum tax formula using continuous distributions of both income and taxes. Lockwood and Weinzierl (2016) apply the inverse optimum method to the US federal tax code in 4 distinct tax years: 1980, 1990, 2000, and 2010. They can calculate the marginal social welfare weights (MSWWs) for each of those years and display them at key percentiles. They show that the MSWWs at the high end dropped substantially between 1980 and 1990 (a result of the 1981 and 1986 tax changes which together reduced top marginal federal rates on income from 70 percent to 28 percent). They solve the Saez function inversely, with the US statutory rates taken as a continuous function. This lets them solve for the weights as a continuous function. Their paper focuses on the high end, omitting consideration of the EITC.

The authors use CBO income data, which does not include the breakdown of income by different sources. Furthermore, the CBO data does not include tax deductions. They make several simplifying assumptions along the way. Lockwood and Weinzierl decide to treat all households as if they are joint filers with two children. Rather than fitting data in each bin, they first use a continuous distribution that they fit to their moments (bin averages) in the data. They use a Pareto-lognormal distribution, which imposes a shape assumption on the data. They do not consider any commodity taxes, nor do they consider deductions or state and local taxes.

They do not break income down by source or consider the effect of different capital gains or dividend taxation.

Lockwood and Weinzierl note that the weights found depend highly on the assumption made about elasticity. They assume a uniform elasticity of taxable income of 0.3. They are interested in the pattern of the weights, and how it compares to their baseline assumption about the pattern you would expect from standard political philosphies. They find that, for reasonable assumptions about elasticities, the revealed welfare weights of the US tax code are inconsistent with the standard guess of $1/c$, where c is after-tax income. Lockwood and Weinzierl suggest that excessively high elasticity assumptions would be required to fit the pattern of weights into common patterns used in the optimum income tax literature. Their work exposes the potential gap between what is expected and what occurs in the tax code in the US. Lockwood and Weinzierl also use their paper to estimate the costs of recessions in terms of the distributional effects relative to the estimated weights, an interesting subject that is beyond the scope of the present paper.

Other recent papers using the inverse-optimum method also examine national-level income taxes. Zoutman et al. (2013) have two papers, one of which deals with the revealed preferences of Dutch political parties. Bargain and Keane (2010) look at implied aversion to inequality over time in Ireland and the UK. Bargain et al. (2014) compare inequality aversion in the US to several European countries using country-specific elasticities. They find that the US is close to identical-weight utilitarianism, and that other countries are more progressive in the pattern of weights.

Gordon and Cullen (2012) present a variant of this method and include the effects of state income taxes, but consider only the effect of US state policy on average rather than going state-by-state. Their paper is primarily interested in migration elasticities, and not in the weights themselves. They use the Gruber and Saez (2002) labour supply elasticities in some of their calculations. They estimate weights and migration elasticities in 5 quintile bins, and get some atypical welfare weights because of the way the EITC phases out. The mobility elasticities that would explain the results are also estimated, and they note that they are also atypical. Milligan and Smart (2018) present a method of decomposing ETI into a pure avoidance and a cross-border-shifting component using an instrumental variables approach, rather than the Gordon and Cullen inverse-optimum approach. Their paper quantifies explicitly the tradeoff between the externality-avoidance benefits of centralizing redistribution and the local-preference benefits of decentralizing it.

Hoynes and Luttmer (2011) examine the insurance effects of state and federal government taxes and transfers. They include the effect of state sales taxes, and use CEX data to estimate household propensities to spend in different categories. They use the general state tax rate and don't consider local tax rates or grocery exemptions from state sales taxes.

III. METHODOLOGY

A. The Consumer's Problem

To include the effects of commodity taxes in the inverse-optimum method, I construct a model where consumers have quasi-linear utility, as in Diamond (1998) and Saez (2001), but also have a second-stage expenditure problem which they solve.

All consumers face income taxes as well as sales taxes, with the sales taxes set differently for each good. Each consumer has a type w and supplies labour L at wage corresponding to their type, and they spend a vector of expenditure shares x on commodity goods at a tax-inclusive commodity price vector q. Index the different commodities by i. Preferences are of the standard separable type, with no income effects on labour supply.

Consumers have a subutility function H, which determines utility from the bundle of consumption goods and is homothetic in x. This means that H is additively separable in commodites, as in Gorman (1959), and that there are homothetic demands for the individual commodities. H depends not just on the goods purchased but on productivity w, which means that different consumers may have different consumption preferences. C is total expenditure, which is equal to income after payment of income tax. There is a constant parameter e, which with this type of utility function will be exactly equal to the elasticity of labour supply (Diamond, 1998). Utility is

(1) $$U(x, C, L) = H(x, w)C - \frac{L^{1+1/e}}{1+1/e}$$

The consumer's problem can then be solved through the standard two-stage budgeting method. We want to find optimal labour supply and pre-tax income at that level. For the first stage, let

(2) $$v(q, w) = \max H\,(x, w) \text{ subject to } q \cdot x = 1$$

be the consumer's indirect utility function, per dollar of after-tax income spent on consumption. Observe that commodity demand functions per unit of expenditure are, by Roy's identity, given by

(3) $$x_i^*(q, w) = \frac{\partial v(q,w)}{\partial q_i} = v_i(q, w)$$

We will denote total income tax payments as T and total commodity tax payments as R^c. Given an income tax function $T(z)$, consumption will be $wl - T(wl)$ as in Diamond (1998). Then for the second stage, the consumer's labour supply function solves the higher-level nested problem

(4) $$U^*(q, w) = \max_L v\,(q, w)[wL - T(wL)] - \frac{L^{1+1/e}}{1+1/e}$$

The solution to this problem will depend on $T(wL)$. Assume, as in real-world income tax schedules, that the consumer faces a linear tax rate over each range of income. Then, it is notationally convenient to consider consumption as having the form $C = (1-\tau)z+I$ for local marginal tax rate τ, net effect of other tax brackets I, and pre-tax income $z = wL$. In other words, at the optimal labour supply L^*, pre-tax income wL^* is taxed at marginal rate $T^0(wL^*) = \tau$. Then the FOC for labour supply is

(5) $$L^*(1-\tau, q, w) = [(1-\tau)v(q,w)w]^e.$$

As in Diamond (1998, equation 7), the elasticity of labour supply with respect to w is e regardless of the income tax rates or commodity tax rates. Pre-tax income is

(6) $$z^*(1-\tau, q, w) = wL^*(1-\tau, q, w)$$

Observe that there are no income effects on labour supply. As well, commodity taxes overall are equivalent to a type-specific increase in the marginal tax rate on labour income in this model, in the sense that a uniform increase in the price level of purchased commodities has the same impact as an increase in the marginal tax rate through the income tax. At the optimum, each consumer pays labour income taxes $T(z^*(1-\tau, q, w))$ plus commodity taxes

(7) $$R^c(\tau, q, w) = \sum_i (q_i - p_i) v_i(q, w)[(1-\tau)z^*(1-\tau, q, w) + I]$$

where p_i are the pre-tax producer prices.

B. The Planner's Problem

Atkinson and Stiglitz (1976) show that, under certain reasonable assumptions, if both income taxes and commodity taxes can be controlled, it is optimal to set those commodity taxes to zero. Kaplow (2008) further shows that if there are any distortionary commodity taxes, the social planner can do a revenue-neutral Pareto-improving tax reform that lowers them and raises income taxes. Kaplow (2006) shows that a differential commodity tax is not optimal even if income taxes are non-optimal.

However, this is not an Atkinson-Stiglitz environment because consumers at different income levels may have different preferences over which bundle to consume, and thus about which commodity taxes to pay. This preference heterogeneity means that the optimal tax system could include commodity tax differentials.

Furthermore, in practice, all US states have a tax environment which features income and payroll taxes, sales taxes, and property taxes. The federal government sets federal income taxes, and local governments set some of the property taxes, as well as some sales and income taxes. The state is setting state income and state sales taxes, and sometimes state property taxes, taking the other taxes as given. If all levels of government were working together to set tax policy optimally, we would not have differential commodity taxation - yet we do observe that.

Thus, I take sales and property taxes as given, and treat the social planner's problem as a second-best problem or a constrained optimization. There are two main reasons to consider the problem in this way. Firstly, I want to perform an inverse-optimum calculation. The objective of this calculation is to find the marginal social welfare weights that would make the actual tax code optimal, including the effects of income, sales, and property taxes. Wealthier individuals have a different propensity to consume housing and taxable goods, so sales and property taxes affect different income groups differently (Casperson and Metcalf, 1993). To omit them from the calculation entirely would ignore this effect. To attempt to determine the marginal social welfare weights that would make combined commodity and income taxes optimal is impossible: no such policy would be optimal because of the Atkinson-Stiglitz and Kaplow results discussed above. Imposing a constraint on the social planner, so they can only control income taxes, provides a better way of finding the value placed on people at different income levels.

Secondly, political considerations make commodity taxes more difficult to change than income taxes. Because US states don't have VATs, the sales tax is highly visible and salient for taxpayers, making changes unpopular. Property taxes are usually set at the state level, so state lawmakers can’t change them easily. Many sales taxes are also set locally. Accordingly, I use a second-best conception of what governments are doing, with commodity taxes as given.

Each state also features tax competition from other states, so theoretically the social planner is choosing a best response function given the other states’ responses. Lehmann, Simula, and Trannoy (2014) describe an extension of the Diamond-Saez formula to find the optimal income tax rates at the Nash equilibirum of two governments with migration effects. Their work highlights the importance of the semi-elasticity of migration.

As well, with state-level income tax increases there is both a vertical fiscal externality (reducing federal income tax revenue due to overall tax avoidance behaviour) and a horizontal fiscal externality (increasing income tax revenue in other states due to the shifting of income and economic activity). Milligan and Smart (2018) show that, although taxes are set separately and are typically uncoordinated, horizontal and vertical fiscal externalities are nearly offsetting. Relatedly, they discuss how it is possible to decompose the elasticity of taxable income into an avoidance effect and a shifting effect, with the later depending on tax rates in other states (Milligan and Smart, 2018). Gordon and Cullen also consider migration elasticities separately, because that is their main object of interest.

However, to scale up the 2-actor model of Lehman, Simula, and Trannoy is beyond the scope of the current work, and we lack reliable estimates of state-specific semielasticites of migration. Thus, in this paper, I make several simplifying assumptions about how the unitary social planner interacts with other jurisdictions. I firstly make the strong simplifying assumption that all other jurisdictions (other states and localities, and for that matter other countries) do not change their tax rates in response to the social planner’s decisions. This is assumption is much more justifiable for small states, such as Wyoming, than for larger ones such as Texas or California. If a large jurisdiction raises taxes, then in a Nash equilibrium framework other jurisdictions will raise their taxes somewhat as well. Thus, the net migration effect will be smaller than a static analysis would suggest, implying a lower true elasticity of taxable income.

I thus consider the combined elasticity of taxable income, which includes both the avoidance and shifting effects. Thus, the elasticity of taxable income parameter is the value of the elasticity taking the tax rates of other states as given. Giertz (2007) has a method of estimating this combined elasticity, as discussed in Section 5. Note that in Giertz's data set, the responses of other jurisdictions to state-level tax changes are not controlled away, and thus are implicitly included in his estimate of the elasticity of taxable income. Thus, his estimates are appropriate for use with these assumptions.

I also assume that the horizontal and vertical fiscal externalities are internalized, and I model government as unitary, so it captures all revenues from federal, state, and local taxes in that region. I consider federal and local income taxes as given along with the commodity taxes, and conceive of the social planner as changing state income taxes. Unlike the method in Gordon and Cullen (2012), this is robust to different preferences by state.

Overall, we should conceive of the unitary social planner as representing a theoretical mechanism for implementing social preferences by controlling income taxes – not necessarily as simply a stand-in for particular levels of government.

The planner has some exogenous function $G(\cdot)$, which represents preferences over income. This is a form of generalized utilitarianism, and is widely used in the optimal income tax literature (Mirrlees, 1971). I assume that the social planner has preferences over household income rather than over type w. The planner chooses the income tax function $T(\cdot)$ optimally, given the commodity tax rates in q, to maximize

$$(8) \qquad \int G\big(U^*(q,w)\big)dF(w) \text{ subject to } \int\Big[T\big(z^*(q,w)\big)+R^c(q,w)\Big]dF(w)\geq 0$$

Note that the planner takes commodity tax payment rates as given, but considers the effect of income tax changes on commodity tax revenues. In addition to the commodity tax payment rates, the planner is exogenously given social preferences $G(\cdot)$ and the distribution $F(w)$. Each social planner represents a combination of the federal government and one particular state government, choosing tax rates on a one filer type, so there are separate planner's problems for each state and filer type.

In this model, homothetic preferences mean the commodity taxes are a perfect substitute for additional progressive labour taxes. Thus, to apply the inverse-optimum method and learn the underlying marginal social welfare weights, I need data not just on income taxes but on sales and property taxes as well. This motivates the approach in the next subsection.

C. Optimal Income Tax Rates

I want to consider the optimal piecewise linear labour income tax system with B tax brackets. Each bracket is associated with a marginal tax rate τ_b and a threshold (minimum) income level z_b at which that marginal tax rate first applies. This builds off of the discrete inverse-optimum method applied in Gruber and Saez (2000, 2002). I number tax brackets beginning at the highest (b = 1) and proceeding downwards.

Proposition 1. *At the optimum, the following equation will hold for each bracket b.*

$$(z_{b-1} - z_b)\sum_{j<b}\left(1 - \frac{g_j}{\mu}\right)(1 - \overline{t_j})P_j + \left(1 - \frac{g_b}{\mu}\right)(\overline{z_b} - z_b)(1 - \overline{t_b})P_b$$

$$= \frac{\tau_b + \overline{t_b}(1 - \tau_b)}{1 - \tau_b} e\overline{z_b}P_b$$

The proof of this proposition is found in Appendix A, along with the procedure for normalizing the weights.

We have b of these equations, one for each bin, and b unknowns, one for each g_b. I then solve this system of equations separately for each state and filer type. Note that if $t_b = 0$, then this proposition collapses to the income-tax only case as in Gruber and Saez (2002).

This Proposition goes beyond Lockwood and Weinzierl (2016) or other papers in the literature to include commodity taxes. In the standard method, when a dollar is returned to the consumer from income taxes, it is treated as staying with the consumer - but in reality, they then spend that dollar, so there is a fiscal externality to the government of higher commodity tax revenue. This fiscal externality is extremely important, as without it there is an incomplete picture of the revenue effects of tax changes. This is now captured in the methodology.

Commodity taxes here are equivalent to a scaling up of the marginal tax rate from labour income taxes. They affect behaviour in the same way since only relative prices matter. Moreover, because of assumption of separable homothetic demands, the marginal effect on commodity tax revenue is proportional to the *average* commodity tax rate faced by agents in that tax bracket. In this model, I am defining the commodity tax rate as the fraction of a marginal dollar in after-income-tax income. Thus, when I calculate the commodity tax rates for each state and income bin, I must use the different types of commodity taxes, such as sales and property taxes, as well as the marginal propensities to spend income in the categories to which those taxes apply.

Now, to solve this system of *n* equations, I want to find $1 - \frac{g_j}{\mu}$ for every bracket, which gives us *n* equations and *n* unknowns. For the inverse-optimum calculation, I know all τ_b, P_b, z_b, and $\overline{z_b}$ values, and I make an assumption about e. I create a matrix version of this system of equations and solve it to find the weights.

When I talk about marginal tax rates in this context, this is the combined effective marginal tax rate on a dollar in pre-tax income. This includes federal and state income taxes and relevant payroll taxes, but is after deductions. We should conceive of the marginal tax rate schedule as mapping one-to-one onto average tax rates.

I make the assumption that there are no household-type-specific transfers of income. In the data I use, TAXSIM automatically considers differences in standard deductions or exemptions as well as the EITC, but I do not consider other transfers. I do not consider the fact that single and joint households may receive different benefits from government spending. Furthermore, different taxes at each level of government (federal, state, and local) may fund different packages of government benefits. However, I make the simplifying assumptions that all public goods are provided by the unitary social planner, that the marginal value of public funds is uniformly 1, and that public goods do not have a differential effect across income groups.

Taxes may have some additional benefits if they are Pigouvian taxes which are designed to correct for a negative externality. However, I assume that there are no externalities resulting from or corrected for by taxes in this model. I do not include potentially Pigouvian taxes such as those on gasoline or alcohol, other than the general sales taxes which fall on all goods.

Lastly, one critique of this inverse-optimum method is that, in fact, these state governments are not acting optimally, so the assumption that they are doing so is misleading. However, in economics, we often make this type of assumption. If we observe a consumer purchasing cereal when the price is $3 but not at $3.50, we infer that the value he places on the box is somewhere in between. This assumes rationality, and in doing so we can learn something about preferences. The consumer may not in fact be rational at all they may have made a mistake, or been driven by some unknown external factor - but we still frequently apply this way of thinking. In the same way, if a government applies the same marginal rate to groups making $60 000 and $80 000 a year, we can still infer something about their underlying preferences, even if they are not truly acting optimally. Conceptually, the marginal social welfare weights calculated in this paper represent the preferences that, if held, would make the combined federal, state, and local tax code optimal. I do not claim that these revealed preferences are the actual preferences of voters, but rather represent some combination of democratic and institutional factors. The nature of this combination is beyond the scope of the current work.

V. TAX RATES

A. Property and Sales Taxes

I first obtain data on tax rates and income from the IRS Statistics of Income data set, and from other sources described in the online appendix. Commodity taxes come in two main forms: property taxes, and sales taxes (including those on groceries). These must be aggregated into a single rate of the payment of commodity taxes in order to solve equation (27). I find the rates to pay property and sales taxes out of a marginal dollar in expenditure, and then aggregate them into a single number.

Different income groups have different propensities to spend on housing, groceries, and non-grocery income, as well as to save. In this paper, I calculate propensities as shares of expenditure, not of income. I do not consider savings. Using current period income may be deceptive, as expenditure is more closely linked to permanent income. Bird and Smart (2016) argue that the progressivity of sales taxes is better understood using consumption rather than income. Thus, for both property and sales taxes I consider the marginal propensity to pay those taxes out of a marginal dollar of expenditure.

As a result, the commodity tax payment rates used in this paper vary somewhat from other methods of calculating the burden of commodity taxes on different income groups, such as the method used by the Institute on Taxation and Economic Policy (ITEP, 2018), which is applied to the income distribution in Piketty et al. (2018). Because higher-income households save a larger fraction of their income, their rate of paying commodity taxes out of income is much lower than their rate of paying commodity taxes out of expenditure.

The payment of property taxes out of income was found by first estimating the value of housing for each bin and state. ACS data on owned housing value was used, but an estimate of imputed value is needed for renters. Monthly rents were annualized and then multiplied by the state price-to-rent ratio in that year, obtained from Zillow. For each state, the price-to-rent ratio is different. This gives either the actual value or an estimate of the imputed value for each home.

I estimate the below equation, where *IMV* is the imputed or actual value of the home, *Inc* is total household income that year, b_i is a dummy variable for each income bin, corresponding to the 9 income bins in the IRS SOI data, and s_j is a dummy variable for each of the 51 states. The log of this value of the home was regressed on an interaction term of log of household income and each bin, as well as an interaction term of the log of household income and the code for each state. I did not use a constant term. Each observation is a single household, and ACS data features weights which show the representativeness of each household within the population, so observations were weighted in the regression.

$$(10) \qquad log(IMV) = \sum_{i=1}^{9} log(Inc) * b_i + \sum_{j=1}^{51} log(Inc) * s_j + e$$

By adding the estimated coefficient on income for each bin and state interaction term, I obtain the marginal propensity for a household in a particular bin and state to increase home value with an additional dollar in income.[1] This was then multiplied by the average local property tax rate, as found in the Tax Foundation data, to find the marginal rate of paying property taxes out of a dollar of income. I am assuming that all property tax incidence falls 100 percent onto the owner and onto the renter.

One possible limitation of this measure is that there may be selection bias in where different homeowners live. Wealthier homeowners may live in areas with lower taxes as a share of owner-occupied housing value, because a larger tax base makes lower tax rates possible in those areas. However, the ACS data on property tax payments has three limitations which make a more detailed calculation difficult: It breaks property tax payments into categories rather than

[1] This is not quite the same as an elasticity, as this is not in percentage terms.

giving the exact amount, it only reports property taxes for owner housing (not implied taxes for renters), and it lumps all annual property tax payments above $10 000 into one category. Also note that I do not adjust for the low-income property tax credits which exist in some states.

For sales taxes, I used CEX data to find the share of non-housing expenditure by households on groceries, services, and all other categories. Attanasio et al. (2004) explain the different technical considerations involved in using CEX data. I found the state sales tax rates and grocery exemptions from the Tax Foundation. I weighted the grocery sales tax rate and the non-grocery goods sales tax rate proportionally to spending by consumers in that income bin. Although some states exempt clothing purchases, I assume that all clothing expenditure is taxed as I have not yet located comprehensive data on state-level clothing tax exemptions. This weighting of sales tax rates by spending categories gives an overall rate of payment of sales taxes out of a marginal dollar in income for each bin in each state.

While services may not be subject to the sales taxes in many states, businesses use inputs which may themselves be subject to sales taxes, raising the question of passthrough of input prices. In this paper, I assume that there is no such pass-through, and that sales taxes do not apply to services. I do not consider additional sales or excise taxes, such as those on alcohol, tobacco, gasoline, or cell phones. I also do not consider tariffs, vehicle taxes, personal property taxes other than those on housing, user fees, or the impact of transfers such as social assistance or Medicaid. Very few filers deduct sales taxes on their federal returns, so that is ignored.

As in the theory section above, these estimates of paying property and sales taxes were combined into a single rate of payment of commodity taxes. I assume that consumers make a two-stage budgeting decision for choosing expenditures as is often done in the literature on expenditures (Gorman, 1959; Molina, 1997). First, consumers allocate a share of income to housing and to non-housing. Secondly, out of the non-housing share they allocate sub-shares to groceries, services, and goods subject to sales taxes. As in Gorman (1959), this is possible if we have made the assumption, as discussed in section 3, of additive separability in the subutility function *H*. I follow that structure in aggregating my tax payments. I calculate the share of marginal expenditure spent on housing (including property taxes) by taking the marginal propensity to spend on housing found above, dividing by the price-to-rent ratio for that state and adding property tax payments. Then, property taxes and sales taxes (as a share of expenditure in that category) are multiplied by the share of overall expenditure to which those taxes apply. The CEX data I have on non-housing spending is national data, so I assume that while housing spending varies from state to state, the sub-shares of non-housing spending on groceries, services, and other goods remains constant across states. This gives a single commodity tax payment rate for consumers in that state and bin.

The estimated commodity tax payment rates, as well as several of the inputs, are presented in Appendix B. Note that these commodity tax payment rates are non-monotonic in income in most states; these are marginal rates of payment of commodity taxes out of a marginal dollar of income, which may not be the same as the average rate of paying commodity taxes.

B. Marginal Tax Rates on Income

I used TAXSIM to find marginal tax rates for each income level and state, as described in Appendix C. TAXSIM automatically considers marginal rates given any level of income and deductions; however, it assumes that deductions (other than the deduction of state or federal taxes) are fixed.

If income is wL, we can define a function $X(wL)$ which is the total taxable income as a function of income. For each bracket b, the effective marginal tax rate on income is then

$$\tau_b = \frac{\partial T\big(X(wL)\big)}{\partial wL} = T'\big(X(wL)\big)X'(wL) \quad (11)$$

In the data, this is the statutory marginal rate T^0 times the propensity for income to be taxable. X^0 is one minus the propensity to take deductions. Feldstein (1999) argues that the impact of deductions is vital for understanding the behavioural effects of a tax change, and here it is important for understanding the impact of marginal rates in each bin.

I adjust for the marginal propensity to take additional deductions out of a dollar of income. This corresponds to the *X*() function in equation (11). To find a simple estimate of the marginal propensity to take additional non-SALT federal deductions, I look at the increase in such deductions taken relative to the increase in income from bin to bin. This estimate is a simple back-of-the envelope calculation, and is likely an upper bound for the true marginal propensity to deduct. (This method may be flawed, because taking average changes between bins may not be the same as estimating the behaviour of the households in that bin.) This provided estimates for the 7 middle bins, and I assume the extreme bin propensities are the same as the nearest bin. Because I am considering all deductions together, I do not separately calculate a propensity to itemize (Benzarti, 2017; Pitt and Slemrod, 1988).

The Tax Foundation calculates average local tax rates by averaging the rates of the largest city and the capital city in each state. While this is a simplification, it is preferable to excluding them entirely. This method may underestimate marginal rates if, as in New York State, more high-end taxpayers are concentrated in local areas with higher marginal rates. TAXSIM does not automatically consider the effect of the deductibility of local income taxes, so while average local income tax rates are included, their deductibility is not considered in this paper.

For some US states with high state taxes, the federal marginal rate on wage income for the second highest bin was the 28 percent rate of the Alternative Minimum Tax, rather than the standard statutory rate (which would typically be 39.6 percent, minus the effect of state income tax deductibility, plus relevant deduction phase-outs). The AMT rate applies when deductions taken are sufficiently large that the total tax payment using the AMT method is larger. In those cases, I assume that the AMT rate applies to the income of the entire bin, but the marginal propensity to take deductions is set to 0 for that bracket.

Payroll taxes are added. I exclude Social Security contributions, which in some ways resemble a tax on income, from this calculation. This is because Social Security benefits depend on income, so it is a payment in exchange for a benefit. Were this included as a tax, it would sharply reduce lower and middle income weights, because the US has payroll taxes only on income up to $90,000 and not thereafter, and only on payroll income. I exclude state and federal unemployment insurance contributions and state disability insurance contributions for similar reasons.

The federal government has Medicare payroll taxes which are applied to all wage income. The base Medicare payroll tax rate is 2.9 percent, with a 0.9 percent surtax on high earners. The surtax also applies to investment income, unlike the regular payroll tax. However, half of the 2.9 percent is an employer contribution that I include, making an assumption that the tax incidence is shifted 100 percent forward onto workers. To find total marginal rates on wages, I want to know marginal tax payments as a share of total employer spending, so I divide the total wage income marginal rate by 1.0145 (Prante and John, 2013). Although self-employed individuals are responsible for both parts of the tax, I apply this adjustment method to all income as a simplifying assumption. The marginal propensity to take deductions is applied to federal, state, and local income marginal rates, but not to payroll taxes as no deductions are possible. I exclude any local payroll taxes.

Thus, the procedure is as follows. I calculate the combined federal, state, and local marginal rates, performing the adjustment for payroll taxes and for the marginal propensity to deduct. The use of the US average deductions for finding marginal rates in TAXSIM means that I don’t consider state-level differences in deductions or propensities to deduct. This gives combined marginal rates on wage income or on capital gains income for each bin. I average those rates, based on the share of income coming from either wage and regular income or capital gains and dividends in each state-bin pair. From the point of view of the social planner at the state level, they have the freedom to raise the marginal tax rate overall - and they could do that by raising either the capital gains or the wage component, or both. For the purposes of finding the weights, we only care about the average.

This gives the marginal tax rate on the typical dollar of new income in that bin. The results for single filers are presented in tables 6 and 7 in Appendix B, and for joint filers in tables 8 and 9. Note that these do not include the effects of commodity taxation, as those taxes are treated differently in equation (27).

VI. RESULTS: MARGINAL SOCIAL WELFARE WEIGHTS

A. Solving for the Weights

I use IRS SOI data to solve for the marginal social welfare weight for each bin-state pair, representing the value placed on a transfer of $1 to members of that group by the social planner. I calculate the weights separately for single and joint filers, as the marginal rate thresholds are different. The IRS data shows the different probability distributions over the two filing types, and is preferable to the CBO data used by Lockwood and Weinzierl because of this separation.

The IRS SOI data reports a ‘less than $1’ category, which includes people reporting tax losses (often business owners) as well as those with zero income. This unusual combination category is separated from the ‘$1 to 10000’ bin, and so it is easy to exclude it in the discrete calculations. I use this category for the normalization of the weights as discussed in Appendix A.

I make an assumption about the splitting of total income among filers. I only observe total income of all filers in a bin, so I assume that single and joint filers who have income in a given bracket have the same average income within that bracket. Thus, I can divide the average gross income (AGI) by the total number of filers of all types in that bin. This may exclude the effect of married filers filing separately or dependent filers (such as children with income), but these are very small categories that are not separately reported in the IRS SOI data. Because of the very small number of Head of Household (HOH) filers (single parents) at the high end of the income spectrum, this data was extremely noisy. Thus, I made the decision not to pursue analysis of filers of this type, although I may in future work.

The Lockwood and Weinzierl method is to calibrate the binned income data to an income distribution, then use that distribution to calculate a continuous function of MSWWs. Instead, I perform this calculation in a discrete way, following the methodology above, and calculate the weight for each bin. The advantage of this method is that, rather than discarding the binned information and using the fitted distribution, I can use the actual data given. This has significant ramifications for some states with unusual income distributions, such as the District of Columbia.

One possible concern would be that the bins do not line up well with the federal or state marginal tax brackets. The inverse optimum method works by assuming that the bin marginal rate is imposed on the whole bin - in other words, that the bin and the bracket are the same. I have used the discrete method, rather than forcing the income data to match a continuous distribution, because it does not discard any of the data, which may be more nuanced than the fitted distribution. Furthermore, the federal marginal rates on single filers line up well with the bin thresholds. This is less true for the joint marginal rate thresholds. State rates may vary more.

I take the TAXSIM output, showing marginal tax rates in each bin and for each state and filer type, along with the state-specific estimates of commodity taxes for each bin, and solve equation (27) for each state. This gives the MSWWs.

This paper uses the elasticity of taxable income (ETI) rather than the elasticity of broad income because the option of shifting income or taking additional deductions is relevant to determining the revenue effects of tax changes. As discussed in section 3.2, I use an ETI with respect to changes in states taxes which takes responses of other state governments and the federal government as given. Auten and Carroll (1999) look at 1980s data and use a method which includes state income taxes, and find an elasticity of taxable income of 0.57. Giertz (2007) updates the elasticity literature using 1990s data, and also performs an estimate including state data. Using the combined federal and state rates as an independent variable, and instrumenting with the predicted change in federal rates, allows him to isolate the effect of changes in the state rate. Giertz finds that the elasticity of taxable income with respect to state income tax changes is 0.347 for the period 1979-1998 and 0.212 for the period 1988-1998. This justifies following Lockwood and Weinzierl (2016) in assuming a constant elasticity of taxable income of 0.3. Moretti and Wilson (2017) find suggestive evidence that migration elasticities are higher for high earners, however I make the simplifying assumption that the combined elasticity of taxable income is constant across income groups.

Furthermore, I assume that this elasticity is constant across states. This may have the effect of biasing the results if in fact the elasticity varies across states. For example, if a state, such as New York, has desirable amenities for wealthy people, the migration component of the elasticity of taxable income may be lower in high brackets. Thus, the marginal social welfare weight calculated with a uniform elasticity of 0.3 may be an underestimate compared to the weight if calculated with a lower elasticity. Thus, states with desirable amenities likely have higher welfare weights on the wealthy than suggested by my calculations below.

One additional issue is whether the elasticity assumption should be adjusted for the component of income that comes from capital gains taxes and dividends. I have made the simplifying assumption that the elasticity for all forms of income is the same.

Several robustness checks were performed. Combining bins at the top, which obviously affects the weight of the combined bin, changes weights at the low end by less than 5 percent - in other words, there is no substantial change in the pattern or level of weights. I also observe that a change in elasticity has a large effect in weights at the top, but a small effect on weights at the bottom. This is expected because for lower income bins, the revenue lost from behavioural changes is much smaller proportional to the effect of a tax change on revenue in higher bins.

B. Overview of the Average Weights

The results are broadly consistent with the findings of Lockwood and Weinzierl (2016). The analysis of the average pattern of weights reveals three major eccentricities in the US tax code: the non-monotonically-declining pattern of weights; the widely differing and sometimes negative weights for some joint filers; and the inconsistent difference between single and joint filers.

If we value people with higher incomes less, we would expect weights to be highest on poor taxpayers and to decline monotonically with income. However, in none of the 51 states do the weights decline monotonically with income for either single or joint filers. Figure 1 shows the simple average of 51 states, unweighted, in red, and the average for joint filers in green. This shows the generally declining pattern of weights, and that none of the weights are negative for single filers. As in Lockwood and Weinzierl, we don't see lower weights for incomes over $1 000 000 compared to those over $200 000. This pattern is a consequence of the shape of the federal brackets, and the fact that two people making $500 000 and $1 000 000 will face the same federal marginal rates, and typically the same state marginal rates. The inverse-optimum formula works by assuming those rates are optimal. Because the elasticities are assumed to be the same, this can only be optimal if the weights on the wealthiest individuals are somewhat higher. In other words, taxes on the very rich aren't high enough to give them a lower weight.

The joint weights (green line in Figure 1) have a very different pattern, in part because of the relationship between the brackets and bins. The single and joint marginal rate schedules are constrained to be the same. Policymakers can adjust the bracket thresholds, and the EITC and some other tax credits are differentiated, but the rates are identical. Thus, to achieve desired policy effects within the bracket constraint is part of what produces the perhaps odd-looking oscillating pattern for joint filers. One finding of this paper is that some of the highest weights are in the 75-100k bin for joint filers. These filers are still subject to the 15 percent marginal rate in the federal bracket, which can only be optimal if their weight is very high. The US federal tax code has a very interesting feature, the large 15 percent bracket, which is especially large for joint filers. This bracket is somewhat of an historical accident. The 1986 federal tax reform reduced the number of brackets to just two: 15 percent and 28 percent. The 1990 and 1993 tax changes added higher rates, but only on wealthy taxpayers above 28 percent. The 2001 tax cut created a 10 percent bracket, and a 25 percent bracket, but left the large 15 percent in place. Because families making $100 000 in pre-deduction income can still be in the 15 percent joint bracket, along with much poorer families, those poorer families have a lower weight assigned by the inverse-optimum method.

For joint filers, some of the state-level weights are negative, notably the large negative weight on the lowest bin found in almost all states. The negative weights may seem incongruous but do have a possible interpretation. The inverse-optimum method assumes optimality, and infers the weight which would make the tax rates optimal. For example, suppose there are two bins which are in the same bracket and face the same marginal income tax rate. The households in the lower bin have less income, so taxes on them raise less money and have smaller negative behavioural change effects. If the marginal rate is the same it must be because they are weighted much less. In the extreme, this logic can create cases where the weight must be negative to be optimal. This means that from the point of view of the social planner, giving a dollar of income to that group reduces the overall social utility. This should be understood in the following way: it means that, with a more conventional social welfare function, the tax code cannot possibly be optimal. It is a strong indicator of an unusual feature in the tax code.

However, this methodology does not consider a separate participation effect for the EITC, and thus, the weight estimate for the lowest bin may not accurately reflect the effects of the EITC. Furthermore, the optimal income tax method of Saez (2001) always gives tax rates which are non-negative, and does redistribution through a lump-sum rebate. Since negative marginal weights will never be optimal, putting them into the inverse-optimum framework is potentially a misspecification of model and problem. Thus, the bottom two weights for joint filers should not be considered to be as accurate as the other weights. I have included the discussion of the negative weights in the current draft for completeness, but going forward I would either need to use a participation elasticity explicitly, or not consider the bottom 2 income bins for joint filers. Note that because the inverse-optimum method is recursive, any issues in calculating the bottom weights does not affect the calculation of the higher weights.

One issue highlighted by this paper is the potential equity consequence of having a small number of large tax brackets. The inverse-optimum calculation shows that to make large brackets optimal, we must have large differences in the MSWWs placed on taxpayers at the high and low ends of those brackets.

The average of joint filers is negative for the lowest bin. This is a consequence of the nearly identical marginal rates for the 1-10k and 10-25k bins. This can only be optimal if the MSWW for the lowest bin is actually negative. There is a significant difference between states with and without an income tax, as seen in the section below. The pattern of joint weights is more variable across states.

Looking again at Figure 1, we can see how much more the social planner is implied to value a joint filer with a spouse and two children relative to a childless single filer. We can see that for the 75-100k bin this difference is very large and the difference is not consistent across bins. This suggests that the difference between single and joint tax schedules is not, in fact, optimal. It should be the case that adding a spouse and children would, with more dependents, consistently increase the implied value of a transfer to that group. Furthermore, we might expect the value of that difference to be increasing in income, but that is not observed.

C. A Geographic View of Taxes and Progressivity

Figure 2 is a map of US states by their top state marginal income tax rate on wage income.[2] The states are broken into 6 quantiles, with the deepest red-coloured states having no state income taxes at all.

A more complete measure of income taxation is the combined federal, state, and local marginal income tax rate (excluding commodity taxes and weighting for capital gains) that I have calculated, as described above. This would consider income from all sources, including capital gains and qualifying dividends. Some states have a much larger fraction of income from those sources than others do, so the income of their taxpayers is taxed at a lower average rate due to the tax-advantaged treatment of those sources of income.

Next, we can combine this with the rate of paying commodity taxes out of a dollar of after-tax-income in the top bin, represented in the same geographic way. Note that some states with very low income taxes have relatively high commodity taxes, such as Washington, Tennessee, Nevada, and Texas; other states with low income taxes also have low commodity taxes, such as Wyoming and New Hampshire. Thus, in some states a low income tax does not necessarily mean a lower tax burden - and thus, considering commodity taxation is extremely important for making relative judgements about progressivity.

We construct a simple combined metric of both income and commodity taxes. Commodity taxes are paid out of after tax income, so if the income tax rate is τ and the commodity tax rate is $\bar{t}$ then total tax payments from a marginal dollar of pre-tax income will be $\tau + (1-\tau)\bar{t}$. A map of US states by this combined rate in the top bracket is shown in Figure 3.

[2] Maps were made using Michael Stepner's Maptile in Stata, which builds off of the spmap package.

What other factors are important for explaining why some states have higher taxes? Income inequality is important: at any given level of weights, more income inequality will lead to higher levels of taxation. Note that some zero income tax states such as Nevada, Wyoming, New Hampshire, Washington, and Florida are among those with the highest average income in the top bin. Those states may be attracting wealthy taxpayers from other states, may have wealthy people using influence to keep taxes low, or may encourage higher income through encouraging risk-taking or through compensation bargaining effects. The other states in that top-income category are high-tax states New York and California, as well as Arkansas, a likely outlier due to the presence of several billionaire Wal-Mart heirs in that small state.

These factors are all combined by the inverse-optimum method to show the weights on the top bin. To examine progressivity, we need to consider not just the top weight but how it relates to other weights. The top weight alone being higher than in other states doesn't differentiate between cases where all weights are relatively high or low.

In Figure 4, I show the ratio of the top (9th) to US median income (3rd) bin for each state, for both single and joint filers. We can interpret the red-coloured states as having the lowest decline in the weights from the median to the top, and thus being least progressive in their revealed preferences. Of the 8 states in red, 5 have no state income tax (Nevada, New Hampshire, Wyoming, Florida, South Dakota), but 2 others actually have an above-median top income tax rate (Arkansas, Idaho). We find that states such as Ohio and West Virginia are more progressive than their tax rates alone might suggest, due to a much lower level of income inequality.

Table 1 presents the same top-to-single ratio, ranked in order of the ratio for single filers, from most to least redistributive. Two states are notable outliers for joint filers: DC, which has a negative implied MSWW on the median income due to high income inequality; and Virginia, which has a very low MSWW on median joint filers.

These weights reveal progressivity in terms of both the income tax code, and inequality. A core principle of the optimal income tax literature is that higher inequality, holding preferences constant, should imply higher taxes on the rich. Thus, if higher inequality is not paired with higher taxes on the rich, the underlying preferences found from this inverse-optimum method are not actually that progressive. For example, Hawaii has what appears to be a very progressive schedule of income tax rates, but inequality is so high that its top-to-median ratio is actually relatively low. States such as West Virginia or Ohio, with relatively few wealthy people, don't have as much redistribution, but overall are relatively progressive by this measure. This measure of progressivity looks at the revealed values of the state, which depends on the tax code relative to the income distribution. Nevada, which has both very little redistribution and very high inequality, is the least progressive state by this measure.

There were a few unusual patterns of weights. Arkansas is a notable outlier with an unusually high income in the top bracket and high capital gains revenue. These bins are analyzed using an average, not a median, so the small number of wealthy heirs of the Walton family may exert a disproportionate effect on average tax revenues in a small economy. The District of Columbia has a unique income distribution over the bins reflecting a high degree of income inequality. In particular, DC has the most taxpayers of any state in the third highest number of filers in the 100-200k bin. Because they are not taxed more highly despite representing a large amount of potential revenue, the inverse-optimum method will assign them a high weight, and the next group down a low weight.

Jurisdictions like California, where wealthy people have a large fraction of their incomes coming from capital gains and dividends, have a lower effective marginal rate on a dollar of income, and thus have higher weights. This is because while California has the highest rate on regular wage income at the top, 14.1 percent, it benefits from the favourable federal tax treatment of capital gains. The extent to which states have capital gains income is thus a significant driver of this result.

D. States With and Without State Income Taxes

Some of the state level variation is explained by whether a state has income taxes or not. In 2016, 9 US states had no state income tax on wages. (Two of these 9, Tennessee and New Hampshire, have income taxes on dividends and interest, but not on wages). These 9 US states exhibit substantially different patterns of weights than the 42 states without.

Here are the different patterns for single weights, again with a 51-state unweighted average in Figure 5. This shows that the income-tax states, with more redistribution, have higher average weights in the bottom 4 bins, and lower average weights in the top 5 bins.

There is also a substantial difference in pattern for joint weights. In the absence of a state income tax, the weight on joint filers in the 75-100k bracket is mostly determined by the federal marginal rate, so it is higher, which forces the lower-bin weights to change.

VII. CONCLUSION

This paper determines the marginal social welfare weights for both single and joint filers that are implied by the tax systems of all 50 US states and the District of Columbia. The unusual pattern of weights, particularly for joint filers, may suggest possible areas of improvement in the tax code. Analysis of the top-to-median weight ratio reveals that progressivity is poorly explained by a simple examination of the rate schedule of US states.

VIII. ACKNOWLEDGEMENTS

I’d like to thank Michael Smart for his guidance and expertise as the supervisor of this paper. He was extremely generous with his time as I pursued this project. My Ph.D. supervisor, Shari Eli, and committee member Kory Kroft were also very helpful and supportive. Thank-you also to many seminar participants at the University of Toronto.

IX. DISCLOSURES

The author has no financial or other arrangements that might give rise to conflicts of interest with respect to the research reported in this paper

X. REFERENCES

Alesina, A., Glaeser, E., and Sacerdote, B., 2001. “Why doesn’t the US have a European-style welfare system?” (No. w8524). National Bureau of Economic Research.

Alesina, A., Stantcheva, S. and Teso, E., 2018. “Intergenerational mobility and preferences for redistribution”. American Economic Review, 108(2), pp.521-54.

Atkinson, A.B. and Stiglitz, J.E., 1976. “The design of tax structure: direct versus indirect taxation”. Journal of Public Economics, 6(1-2), pp.55-75.

Attanasio, O., Battistin, E. and Ichimura, H., 2004. “What really happened to consumption inequality in the US?” (No. w10338). National Bureau of Economic Research.

Bargain, O., Dolls, M., Neumann, D., Peichl, A., and Siegloch, S., 2014. “Comparing inequality aversion across countries when labor supply responses differ.” International Tax and Public Finance, 21(5), 845-873.

Bargain, O., and Keane, C., 2010. “Tax–Benefit-revealed Redistributive Preferences Over Time: Ireland 1987–2005.” Labour, 24, 141-167.

Benzarti, Y., 2017. “How Taxing Is Tax Filing? Using Revealed Preferences to Estimate Compliance Costs” (No. w23903). National Bureau of Economic Research.

Bird, R. and Smart, M., 2016. “Finances of the Nation.” Canadian Tax Journal, 64(2), pp.417-442.

Bourguignon, F. and Spadaro, A., 2012. “Tax–benefit revealed social preferences.” The Journal of Economic Inequality, 10(1), pp.75-108.

Caspersen, E. and Metcalf, G., 1993. “Is a value added tax progressive? Annual versus lifetime incidence measures” (No. w4387). National Bureau of Economic Research.

Chetty, R., Hendren, N., Kline, P. and Saez, E., 2014. “Where is the land of opportunity? The geography of intergenerational mobility in the United States.” The Quarterly Journal of Economics, 129(4), pp.1553-1623.

Diamond, P., 1998. “Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates.” The American Economic Review, 88(1), 83-95.

Diamond, P., and Saez, E., 2011. “The Case for a Progressive Tax.” Journal of Economic Perspectives, 25(4), 165-190.

Feenberg, Daniel. and Coutts, E., 1993. “An Introduction to the TAXSIM Model.” Journal of Policy Analysis and Management, vol 12 no 1, Winter 1993, pages 189-194. (https://www.nber.org/taxsim/).

Feenberg, D.R. and Rosen, H.S., 1986. “The Interaction of State and Federal Tax Systems: The Impact of State and Local Tax Deductibility.” The American Economic Review, 76(2), pp.126-131.

Feldstein, M., 1999. “Tax Avoidance And The Deadweight Loss Of The Income Tax.” The Review of Economics and Statistics, MIT Press, vol. 81(4), pages 674-680, November.

Gordon, R.H. and Cullen, J.B., 2012. “Income redistribution in a federal system of governments.” Journal of Public Economics, 96(11-12), pp.1100-1109.

Gorman, W.M., 1959. “Separable utility and aggregation.” Econometrica: Journal of the Econometric Society, pp.469-481.

Gruber, J. and Saez, E., 2000. “The Elasticity of Taxable Income: Evidence and Implications.” NBER Working Paper, (w7512). (Contains some content not in below published version.)

Gruber, J., and Saez, E., 2002. “The Elasticity of Taxable Income: Evidence and Implications.” Journal of Public Economics, 84, 1-32.

Hacker, J., and Pierson, P., 2005. “Off Center: The Republican revolution and the erosion of American democracy.” Yale University Press.

Hacker, J., and Pierson, P., 2010. “Winner-take-all Politics: How Washington made the rich richer–and turned its back on the middle class.” Simon and Schuster.

Hermle, J. and Peichl, A., 2018. “Jointly Optimal Taxes for Different Types of Income.” CESifo Working Paper Series 7248, CESifo Group Munich.

Hoynes, H.W. and Luttmer, E.F., 2011. “The insurance value of state tax-and-transfer programs.” Journal of Public Economics, 95(11-12), pp.1466-1484.

Internal Revenue Service, 2018. “Statistics of Income Historical Table 2.” IRS, Washington, DC. https://www.irs.gov/statistics/soi-tax-stats-historic-table-2

ITEP (Institute on Taxation and Economic Policy), 2018. ”Who Pays? A Distribution Analysis of the Tax Systems in All 50 States, 6th Edition.” Washington, DC. https://itep.org/wp-content/uploads/whopays-ITEP-2018.pdf

Kaplow, L., 2006. “On the undesirability of commodity taxation even when income taxation is not optimal.” Journal of Public Economics, 90(6-7), 1235-1250.

Kaplow, L., 2008. “The Theory of Taxation and Public Economics.” Princeton University Press.

Lehmann, E., Simula, L. and Trannoy, A., 2014. “Tax me if you can! Optimal nonlinear income tax between competing governments.” The Quarterly Journal of Economics, 129(4), pp.1995-2030.

Lockwood, B. B., and Weinzierl, M., 2016. “Positive and normative judgments implicit in US tax policy, and the costs of unequal growth and recessions.” Journal of Monetary Economics, 77, 30-47.

Molina, J.A., 1997. “Two-stage Budgeting as an Economic Decision-making Process for Spanish Consumers.” Managerial and Decision Economics, 18(1), pp.27-31.

Moretti, E. and Wilson, D.J., 2017. “The effect of state taxes on the geographical location of top earners: evidence from star scientists.” American Economic Review, 107(7), pp.1858-1903.

Milligan, K. and Smart, M., 2018. “An Estimable Model of Income Redistribution in a Federation: Musgrave Meets Oates”. (No. 7043). CESifo Working Paper.

Mirrlees, J., 1971. “An Exploration in the Theory of Optimal Income Taxation.” Review of Economic Studies(38), 175-208.

Persson, T., and Tabellini, G. E. (2005). “The economic effects of constitutions.” MIT press.

Piketty, T., Saez, E., and Zucman, G., 2018. “Distributional National Accounts: Methods and Estimates for the United States.” The Quarterly Journal of Economics, Oxford University Press, vol. 133(2), pages 553-609.

Pitt, M.M. and Slemrod, J., 1988. “The compliance cost of itemizing deductions: Evidence from individual tax returns.” NBER Working Paper, 1988.

Poterba, J. and Sinai, T., 2008. “Tax expenditures for owner-occupied housing: deductions for property taxes and mortgage interest and the exclusion of imputed rental income.” American Economic Review, 98(2), pp.84-89.

Prante, G.T. and John, A., 2013. “Top Marginal Effective Tax Rates by State and by Source of Income, 2012 Tax Law vs. 2013 Tax Law (as enacted in ATRA).” Available at SSRN 2176526.

Ruggles, S., Flood, S., Goeken, R., Grover, J., Meyer, E., Pacas, J., and Sobek, M., 1993. “IPUMS USA: Version 9.0 [dataset].” Minneapolis, MN: IPUMS, 2019. https://doi.org/10.18128/D010.V9.0

Saez, E. 2001. “Using Elasticities to Derive Optimal Income Tax Rates.” The American Economic Review, 68, 205-229.

Saez, E., Slemrod, J., and Giertz, S. 2012. “The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review.” Journal of Economic Literature, 50(1), 3-50.

Saez, E. and S. Stantcheva, 2016. “Generalized Social Marginal Welfare Weights for Optimal Tax Theory.” American Economic Review, American Economic Association, vol. 106(1), pp. 24-45.

Stepner, Michael, 2015. “MAPTILE: Stata module to map a variable”. Statistical Software Components S457986, Boston College Department of Economics, revised 23 Jun 2017.

Tax Foundation, 2018. “Facts and Figures 2018”. Tax Foundation, Washington, DC.

Tax Foundation, 2018. “2018 State Business Tax Climate Index.” Tax Foundation, Washington, DC. https://taxfoundation.org/state-business-tax-climate-index-2018/.

Weinzierl, M., 2014. “The promise of positive optimal taxation: normative diversity and a role for equal sacrifice.” Journal of Public Economics, Elsevier, vol. 118(C), pages 128-142.

Zillow, 2019. “State Price-to-Rent ratios, All Homes, 2016 calendar year.” https://www.zillow.com/research/data/ . Retrieved on May 16, 2019. Aggregated data on this page is made freely available by Zillow for non-commercial use.

Zoutman, F. T., Jacobs, B., and Jongen, E. L., 2013. “Revealed Social Preferences of Dutch Political Parties.” Working Paper.

APPENDIX A: THE INVERSE-OPTIMUM EQUATION

Proposition 2. *At the optimum, the following equation will hold for each bracket b.*

$$(z_{b-1}-z_b)\sum_{j<b}\left(1-\frac{g_j}{\mu}\right)\left(1-\overline{t_j}\right)P_j+\left(1-\frac{g_b}{\mu}\right)\left(\overline{z_b}-z_b\right)\left(1-\overline{t_b}\right)P_b$$

$$=\frac{\tau_b+\overline{t_b}(1-\tau_b)}{1-\tau_b}e\overline{z_b}P_b$$

Proof. The portion of a taxpayer’s income z that is in bracket b is the lesser of $z-z_b$ if that difference is positive, and $z_{b-1}-z_b$, or

(13) $\widetilde{z_b}(z)=\min\{\max\{z-z_b,0\},z_{b-1}-z_b\}$

Total income tax payments by that taxpayer are therefore

(14)

$$T(z)=\sum_{b=1}^{B}\tau_b\widetilde{z_b}(z)$$

To solve the optimal income tax problem, we want to consider averages of the different variables within each bracket.

Pre-tax income will be uniquely determined at any given wage level. Given any such tax system, consumers may be partitioned into subintervals of productivity $W_b=[w_b,w_{b-1}]$ that choose to locate in bracket *b* for *b* = 1,...,*B*. Let $P_b = \int_{W_b} dF(w)$ be the associated probability mass of consumers in that partition, and

(15) $$\bar{z_b} = \frac{1}{P_b}\int_{W_b} z^*(1-\tau_b, w)dF(w)$$

be average labour income in the bracket.

To characterize the planner's problem, we want to be able to aggregate the different commodity tax rates q −1 on each good *i* into a single rate of paying commodity tax on a marginal dollar of income. The same partition of consumers above governs commodity demands. I make the simplifying assumption that consumers within each bracket have identical per-dollar demands over commodity goods x^*_i, and thus that the within-bracket covariance between x^*_i and *C* is 0. Using equation (3), average total demand *D* for commodity *i* by consumers in bracket *b* is therefore:

(16) $$D_{bi} = \frac{1}{P_b}\int_{W_b} x_i^*(q,w)C(1-\tau_b, w)dF(w) = \overline{x_{bi}^*}(q)C_b$$

where

(17) $$\overline{x_{bi}^*}(q) = \frac{1}{P_b}\int_{W_b} x_i^*(q,w)dF(w)$$

and

(18) $$C_b = \frac{1}{P_b}\int_{W_b} C(1-\tau_b, w)dF(w)$$

We normalize producer prices to 1 and define $t_i = q_i - 1$ as the percentage tax on each commodity. Then average commodity tax payments by a consumer in bracket b are

(19) $R_b^c = \sum_i t_i D_{bi} = \sum_i t_i \overline{x_{bi}^*}(q)C_b = \overline{t_b}(q)C_b$

(20) $$\overline{t_b}(q) = \sum_i t_i \overline{x_{bi}^*}(q)$$

To characterize the optimal income tax system in this environment, I employ a version of the variational arguments developed by Saez (2001). In particular, beginning from the piecewise linear tax system just described, consider an increase $\Delta\tau$ in the marginal tax rate on labour incomes in bracket b. This perturbation results in a mechanical effect and a behavioural effect on tax revenue in bracket b, and an inframarginal change in total tax payments for all agents locating in brackets $j < b$. The change in labour income tax payments in bracket j is approximately

(21) $$\Delta T_j \approx \tau_b \Delta \overline{z_b} + \Delta\tau(\overline{z_b} - z_b) \; for \; j = b$$

$$\Delta T_j \approx \Delta\tau(z_{b-1} - z_b) \; for \; j < b$$

The change in commodity tax payments in bracket j is

(22) $$\Delta R_j^c = \overline{t_j} \Delta C_j$$

Note that for the bracket $j = b$, commodity tax payments change in two ways: mechanically as $\Delta\tau$ reduces after-tax income, and also from $\Delta \bar{z}_b$ as the consumer changes their labour supply. So

(23) $$\Delta R_b^c = \overline{t_b} \Delta C_b = \overline{t_b}\big((1 - \tau_b)\Delta \overline{z_b} - \Delta\tau(\overline{z_b} - z_b)\big)$$

An optimal tax system sets the marginal impact on welfare of any such perturbation $\Delta\tau$ to zero. Define g_j as the marginal social value of a dollar transferred to the average agent in bracket j. This is our marginal social welfare weight, which combines the effects of both individual marginal utility and the social planner's weighting of the welfare of that individual. Letting μ denote the marginal social value of public funds, this condition can be written as

(24) $$\Delta W = \sum_{j\leq b} g_j \Delta C_j P_j + \mu \sum_{j\leq b}\left(\Delta T_j + \Delta R_j^c\right) P_j = 0$$

This simplifies to

(25) $$\Delta W = \sum_{j<b}\left(\mu - g_j\right)\Delta\tau(z_{b-1} - z_b)\left(1 - \overline{t_j}\right)P_j + (\mu - g_b)\Delta\tau(\overline{z_b} - z_b)(1 - \overline{t_b})P_b + \mu\left(\tau_b + \overline{t_b}(1-\tau_b)\right)\Delta\overline{z_b}P_b = 0$$

Re-writing the definition of elasticity, we see that the behavioural change in average income in bracket b is

(26) $$\Delta\overline{z_b} = -\frac{1}{1-\tau_b} e\overline{z_b}\Delta$$

Substituting that in and dividing through by $\Delta\tau$ and μ, and re-arranging, we have

(27)

$$(z_{b-1} - z_b)\sum_{j<b}\left(1 - \frac{g_j}{\mu}\right)\left(1 - \overline{t_j}\right)P_j + \left(1 - \frac{g_b}{\mu}\right)(\overline{z_b} - z_b)(1 - \overline{t_b})P_b$$

$$= \frac{\tau_b + \overline{t_b}(1-\tau_b)}{1-\tau_b} e\overline{z_b}P_b$$

□

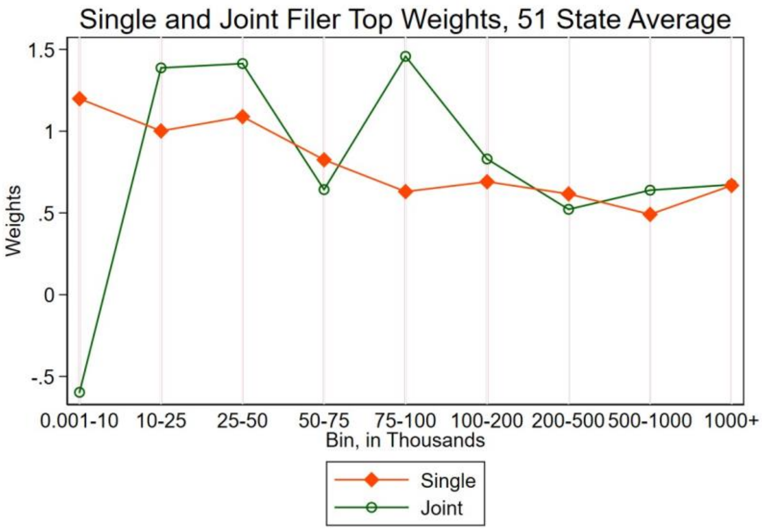
The inverse-optimum problem is solved separately for single and joint filers, giving values $\frac{g_j}{\mu}$ for both types. To finish the calculation, I assume that the value of μ (the marginal value of public funds) used in each state is the same for both filer types. This is normalized to 1. As in other inverse-optimum papers such as Lockwood and Weinzierl (2016) and Saez and Gruber (2002), I require that at the optimum, the social planner is indifferent between a one dollar lump sum subsidy across the entire population and a dollar in public spending. Thus, I calculate the weight g_0 that would, when placed on the share of filers with less than one dollar in income, satisfy $\sum_{i=0}^{n} P_i g_i = 1$. This normalization gives comparability of the weights across household types.

Table 1: Top-to-Median Weight Ratio, for Single and Joint Filers (Ranked by Single Filer Ratio)

State	Single	Joint		State	Single	Joint
New Jersey	0.48	0.75		Connecticut	0.61	0.34
Maryland	0.51	0.80		Kansas	0.61	0.40
District of Columbia	0.53	-0.17		US Average	0.61	0.46
West Virginia	0.55	0.43		North Dakota	0.61	0.28
New York	0.55	0.66		Mississippi	0.62	0.46
Ohio	0.56	0.42		Michigan	0.62	0.45
Alaska	0.56	0.27		Kentucky	0.62	0.48
Delaware	0.56	0.42		Oregon	0.62	0.44
Minnesota	0.57	0.50		Arizona	0.63	0.57
Iowa	0.57	0.38		Colorado	0.64	0.57
Virginia	0.57	1.02		Texas	0.64	0.42
Georgia	0.58	0.58		New Mexico	0.64	0.53

Louisiana	0.59	0.41		Oklahoma	0.65	0.54
Vermont	0.59	0.47		Montana	0.65	0.47
Pennsylvania	0.59	0.35		Hawaii	0.65	0.42
Rhode Island	0.59	0.37		Tennessee	0.66	0.49
Massachusetts	0.60	0.66		Missouri	0.66	0.47
Alabama	0.60	0.41		Washington	0.66	0.35
California	0.60	0.70		New Hampshire	0.66	0.29
South Carolina	0.60	0.44		South Dakota	0.67	0.36
Illinois	0.60	0.51		Idaho	0.67	0.50
North Carolina	0.61	0.59		Utah	0.68	0.46
Nebraska	0.61	0.42		Florida	0.68	0.43
Wisconsin	0.61	0.38		Wyoming	0.71	0.35
Maine	0.61	0.46		Arkansas	0.72	0.59
Indiana	0.61	0.38		Nevada	0.72	0.41

Figure 1


Single and Joint Filer Top Weights, 51 State Average
1.5
1
.5
0
-.5
Weights
0.001-10
10-25
25-50
50-75
75-100
100-200
200-500
500-1000
1000+
Bin, in Thousands
Single
Joint

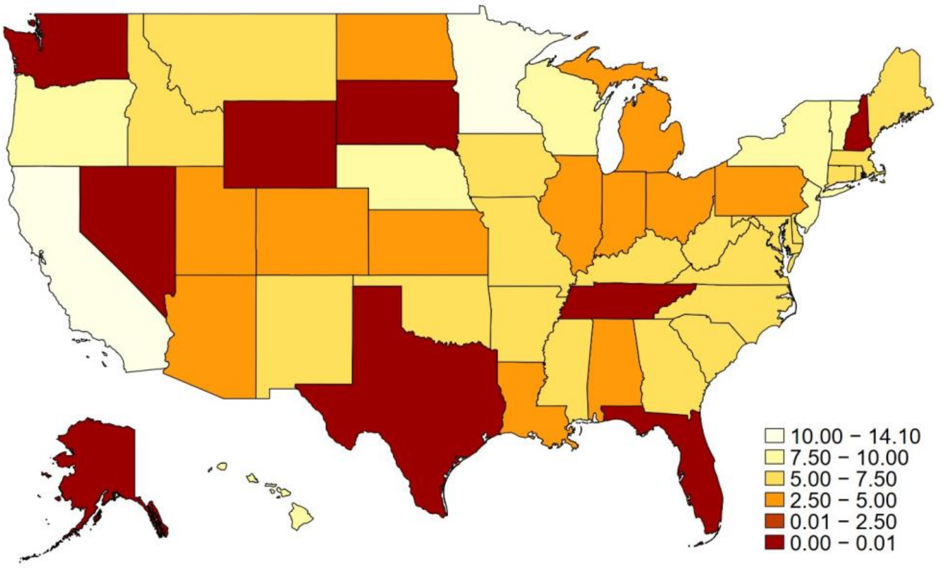


Figure 2: US States by Top State Marginal Rate on Wage Income

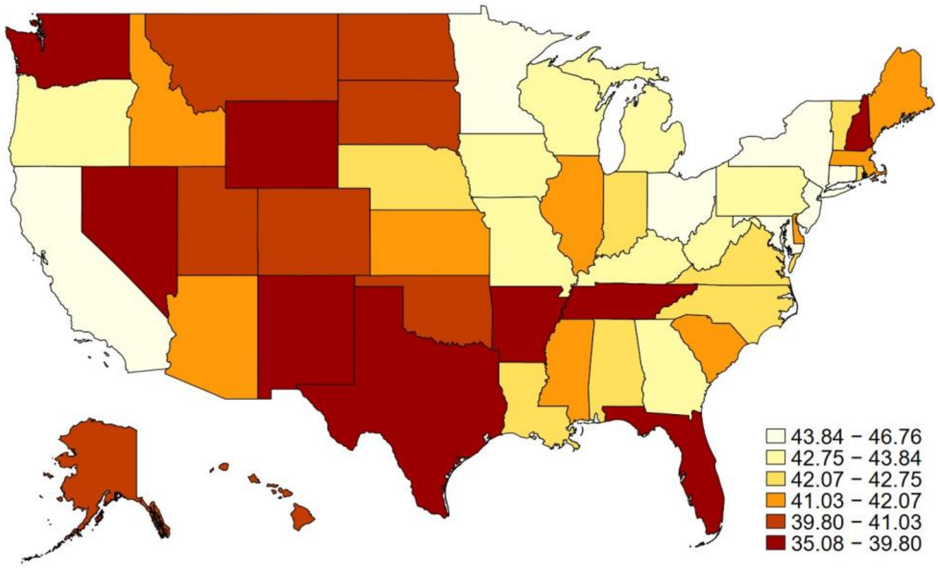


Figure 3: US States by Top Bin Combined Income and Commodity Marginal Tax Rate

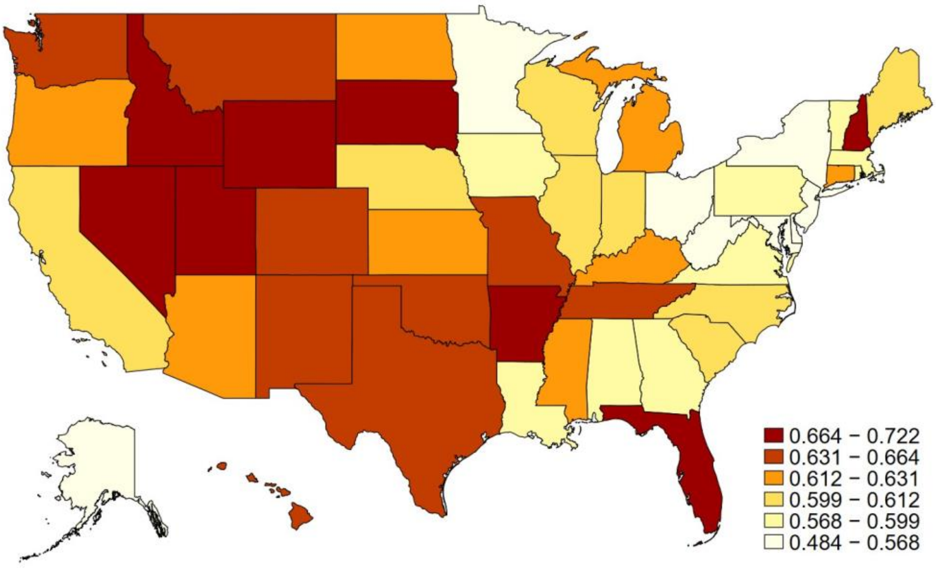
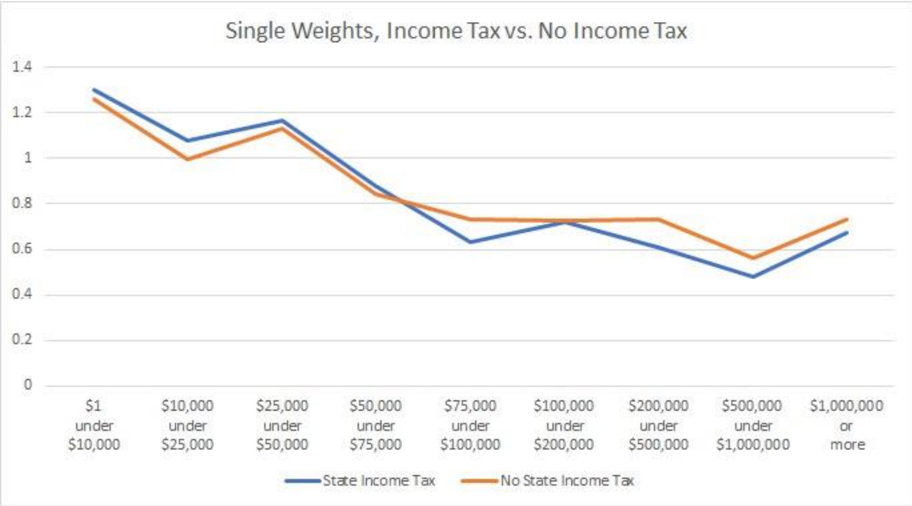


Figure 4: US States by Top-to-Median Weight Ratio



Single Weights, Income Tax vs. No Income Tax
1.4
1.2
1
0.8
0.6
0.4
0.2
0
$1 under $10,000
$10,000 under $25,000
$25,000 under $50,000
$50,000 under $75,000
$75,000 under $100,000
$100,000 under $200,000
$200,000 under $500,000
$500,000 under $1,000,000
$1,000,000 or more
State Income Tax
No State Income Tax

Figure 5