On The Empirics of Optimal Tax policy Under Parameter Uncertainty

Italo Morais Santos

First Version: February 2019
Current Version: June 2022

Abstract

In this paper, I investigate optimal income tax policy when policymakers don't know the elasticity of taxable income, the key structural parameters which is sufficient to determine optimal tax in a stylized model. In this setting, a question of how parameter uncertainty should be represented in the optimal tax policy arises. The standard practice is to derive optimal taxes as mappings of structural parameters and plug-in point and interval estimates in this mapping. I show the policy maker prior information is not fully incorporated in this case and propose an alternative — using welfare as a loss function when summarizing the posterior distribution of structural parameters. In a simulation exercise, I show the optimal policy interval estimates using the plug-in approach are wider than the ones using my proposed approach without any gain in coverage.

1 Introduction

The elasticity of taxable income (ETI) is a central parameter in public finances. It measures the individual response to changes in taxation on earned income. This parameter is central to the measurement of potential tax distortions and therefore central to the design of optimal taxation and transfer policies and other related concepts in public finances such as the Laffer Curve. In order to apply the ETI in optimal tax analysis, economists derive optimal tax formulas and plug-in empirical point and interval estimates of the ETI. There is however no theory, economic or econometric, that justify this practice.

In this paper I investigate optimal tax analysis when policymakers don't know the structural ETI parameter. I construct a simple model where the government must, besides designing the tax and transfer policy, estimate the ETI using the available data using Bayesian regressions. The main feature of this model is to incorporate statistical uncertainties into the policymaker's optimization problem. I show that in this setting the policymaker faces two separate decisions: one regarding the construction of optimal tax and transfer schedule and another statistical decision problem regarding the value of the optimal tax. The standard practice of using plug-in estimates in optimal tax formulas is not justified by statistical decision theory, although it delivers consistent point estimates. However, I show the plug-in approach does not incorporate all the prior information regarding the ETI parameter. To correct this problem, I propose the policymaker should incorporate prior beliefs by optimizing tax under parameter uncertainty. In this framework, both the optimal tax and the statistical decision are solved simultaneously. From the perspective of an econometrician, my proposed approach amounts to using welfare functions as the loss function in the statistical decision problem. To showcase my approach, I perform a simulation experiment. My proposal approach delivers much tighter credible sets without any loss of coverage, due to the incorporation of prior information in the tax optimization step. In addition to the optimal taxation problem, I also analyze other issues in the public finances that revolve around the ETI parameter and can benefit from my proposed framework, such as the Laffer Curve estimates.

Related Literature During the 1990s, empirical literature sought to estimate the ETI in order to analyze the optimality of the US income tax schedule. The initial attempts of Slemrod (1996) and Feldstein (1995) were quite different. Slemrod (1996) used time series data from high-income sharing to the US and tested a structural break during the Tax Reform Act of 1986. Feldstein (1995) also uses the 1986 tax reform to identify ETI, but using a differences-in-differences estimate. This second approach has proved to be more fruitful as many other studies have built in this framework. Auten and Carroll (1999), Moffitt and Wilhelm (1998), Gruber and Saez (2002), Giertz (2007) and Weber (2014) attempt to improve the identification of the ETI, taking into account the possibility of mean reversion [1]. From this point on, more structural literature emerged, trying to bridge the gap between theory and policy formulation instead of dealing with identification issues. For example, Chetty (2009) investigates the conditions for the reduced form parameters to be sufficient for policy making. This paper relates to this second branch of literature.

The remainder of this paper is organized as following: In Section 2, I establish the theoretical framework, both for the estimation of the optimal tax and for these alternative questions in public finances. In Section 3, I briefly discuss the econometric framework used to identify the ETI in Bayesian setting. In Section 4, I present the simulation experiment and discuss its results. Finally, in Section 5, I conclude the paper and discuss possible paths for future research.

2 Optimal Tax Design and Parameter Uncertainty

2.1 Benchmark

In this section, I present two cases in public finances, where the ETI is a central parameter. The optimal linear marginal income tax model as in Sheshinski (1972) and Mirrlees (1971). Although linear taxes are too simple compared to the reality of tax legislation, I believe that this is a good starting point for studying the design of tax laws under the uncertainty of the ETI. Next, I analyze the alternative problem of revenue change, given a tax reform proposal.

[1] To complete review of the literature during the 1990s and 2000s, see Saez et al. (2012)

Optimal Linear Taxation Suppose the economy is populated by a continuum of individuals i distributed across $I(i)$. Without loss of generality, I normalize the population size to 1. As in Saez, Slemrod and Giertz (2012) formulation of this problem, individuals choose between consumption c and taxable income y. I opt for this formulation exactly for the same reasons the empirical ETI literature do: many intensive margin factors such as individual skill are not observable and in our model the social planner must face the optimal taxation problem under the same informational constraints of econometricians. Individuals must pay a τ linear income tax. They also receive a fixed amount transfer of R. Thus, the disposable income of the individual is $c_i = (1-\tau)y_i + R_i$. The optimal marginal tax problem is given by:

$$\max_{\tau,R} \int \alpha_i W(u_i((1-t)y_i + R_i, y_i))dI(i) \tag{1}$$

$$\text{s.t. } tY(1-t,R) \geq G + R, \tag{2}$$

where α_i is the Pareto weights, $W(.)$ is the social welfare function and $Y = \int y_i dI(i)$ is aggregate revenue and G is a exogenous government expenditure. Note that, for the government budget set, R is implicitly defined by τ. For this reason, the entire fiscal policy, which is the pair τ, R, can be unambiguously referenced as just τ or R. In solving the optimal marginal fiscal problem in relation to only τ, I also determine optimal transfers R. If I solve for optimal transfers R instead I would find the same optimal τ [2].

Rewriting the problem using this fact and the envelope theorem, one can show that the first-order conditions of the social planner's problem are:

[2] Often this is not done because there is not a closed solution for $R(\tau)$ in the budget equation from the government. That would require a numerical method. Later, I will show that when we consider parameter uncertainty, this particular feature of the model is lost

$$\int_i \alpha_i W'(u_i) u_{ci}(Y - y_i - \tau \frac{dY}{d(1-\tau)}) = 0, \tag{3}$$

where u_{ci} is marginal utility of consumption. From 3 we may obtain the solution of the optimal tax problem, the well-known formula:

$$\tau^*(\epsilon) = \frac{1+g}{1+g-\epsilon}, \tag{4}$$

where ϵ is the aggregate elasticity of taxable income and $g = \int g_i y_i dI(i)/Y$. g_i is the marginal social welfare weights:

$$g_i = \frac{\alpha_i W'() u_{i,c}}{\int \alpha_i W'() u_{i,c} dI(i)}. \tag{5}$$

Empirically computing the optimal marginal tax comes down to computing these two values. Here I focus on the aggregate ETI ϵ as this is the parameter to be estimated.From this point onward, consider g is fixed and known ($g = g_0$).

The empirical ETI literature typically plug-in the estimates in some optimal tax formula like 4. Even in calibration exercises of advanced optimal tax models such as Heathcote et al. (2017), unknown elasticities are obtained from estimates of econometric models.

Suppose now that the government does not know the real ϵ value. I refer to the solution of this standard optimal tax problem as a mapping $\tau : \Theta \rightarrow [0, 1]$ where Θ is the parameter space for ϵ. I call this solution *optimal tax policy*. This is in contrast with the standard approach of fixing ϵ, where the optimal tax would be a fixed real number. Under this parameter uncertainty, solving the optimal tax problem alone is not sufficient to yield a decision about the optimal tax.

Definition 1. *The plug-in optimal decision* $\tau^{d1} = \tau(\epsilon_0)$ *where* ϵ_0 *is the policy maker point estimator of the unknown parameter* ϵ *and* $\tau(.)$ *is the optimal tax policy.*

A fairly general way of defining beliefs is using probabilities. Let $p(\epsilon)$ be the government's belief about the unknown parameter ϵ. If the government is absolutely certain about the value of ϵ, then $p(\epsilon = \epsilon_0) = 1$ and we call this a degenerate belief. Doubts about ϵ true value manifest in the form of a more defuse probability distribution for ϵ. If the government have defused beliefs about ϵ then the optimal tax value is a random variable with distribution $p(\tau(\epsilon))$.

One way to interpret the procedure of using plug-in point estimators is to compute the optimal tax value with optimal tax policy $\tau(.)$ and degenerate beliefs $p(\epsilon = \hat{\epsilon}) = 1$, where $\hat{\epsilon}$ is an ETI estimator. This is a good decision in the sense that if the estimator ϵ is consistent, the optimal tax value is also consistent by direct application of the continuous mapping theorem. However, this is an incoherent formulation as the government knows for a fact that $p(\epsilon = \hat{\epsilon}) < 1$ due to parameter uncertainty.

Bayesian Beliefs A natural approach to deal with parameter uncertainty and beliefs in the context of estimation is using a Bayesian estimator. In this approach, the government observes data [3] and making use of a statistical likelihood model $p(Y, X, Z, W|\epsilon)$, update its prior beliefs $p(\epsilon)$ into the posterior distribution $p(\epsilon|X, Y, W, Z)$. Draws from the posterior distribution $p(\epsilon|Y, X, Z, W)$ can be used to draw a distribution for $p(\tau(\epsilon)|Y, X, Z, W)$, even when the non-linearity of the optimal tax policy makes it impossible to derive a parametric form for the optimal tax value distribution.

According to Definition 1, the optimal tax value under the Bayesian belief is a random number generated from the $p(\tau(\epsilon)|Y, X, Z, W)$ distribution. This optimal tax "decision" has the advantage of adopting a non-degenerate belief regarding ϵ. An immediate consequence

[3] $\{Y, X, Z, W\}$ observed data is respectively, dependent variable, endogenous independent variables, possible control variables, and additional exogenous covariates. This is done because we showcase our results by simulating a 2SLS estimator for the ETI. See Sections 3 and 4 for more details.

of having a full distribution as an optimal tax decision is that we can now gauge the level of uncertainty with respect to the policy and possible asymmetries in its distribution.

Another advantage of Bayesian beliefs is that it adds a new layer of variation to the optimal tax debate. Policymakers often disagree on tax policy and alternative tax reforms. The welfare approach of the optimal taxation theory may explain these differences by varying the normative elements of the problem, the Pareto weights, and the social welfare function. For example, egalitarian politicians can design tax policies based on Rawlsian's welfare function.

Clearly, differences in normative beliefs are an important part of the public debate on taxation. However, we believe that disagreement is not limited to these. Politicians may also disagree because they have different views about the impact of taxation. When policymakers advocate higher or lower taxation based on the merits of efficiency, they do so because they believe in the value of the ETI, or a related concept, and not in their personal stance on distribution. In the standard approach of using point estimators, these differences can not be considered. If policymakers should stick to the particular estimation procedure, what explains their disagreement?

The Bayesian posterior offers a formal and rigorous alternative where differences in beliefs about how individuals respond to taxation, given by the ETI statistic, can be easily implemented via different prior distributions $p(\epsilon)$ between policymakers. The mean of the prior distribution represents the policy maker's expectation about the ETI before seeing any data. Policymakers who believe high taxation represents a large disincentive to work may be represented in this model by having much of its prior distribution mass around high ETI values. The posterior distribution is a function of both these prior beliefs and the likelihood function of the estimation which should be common to all policymakers - at least all who have access to the same data. Thus using the Bayesian estimation posterior I may account for these differences. An immediate consequence of the incorporation of Bayesian estimation is that, since the optimal tax is inversely proportional to the ETI values, policymakers with strong beliefs in low ETI values are likely to propose higher taxes, and vice versa, for a given preference for social welfare. I believe this particular advantage is of relevance when optimal tax theory is interpreted not as a normative tool to design tax policy but as an analytical tool to gauge government's underlying preferences given their tax policy choices.

Despite all these advantages, the Bayesian optimal tax value is of little practical use by itself. A single random drawn from $p(\tau(\epsilon)|Y, X, Z, W)$ will naturally have a high variance compared to plug-in an estimator, delivering a much less efficient estimator of the true optimal tax. However, this is to be expected, as in Bayesian statistics the posterior distribution is used to construct the statistics of interest via statistical decision theory.

The goal is to establish a loss function such that the decision rule is consistent. A natural candidate, for example, would be minimizing the quadratic loss $L(x) = (x - \tau(\epsilon))^2$. This wields the decision rule:

Definition 2. *The Bayesian optimal tax decision is* $\tau^{d2}_{Plug-In} = \mathbb{E}(\tau(\epsilon))$ *where* $\tau(.)$ *is the optimal tax policy and* $p(\epsilon)$ *is the government belief on* ϵ.

Surprisingly this alone doesn't yield a consistent decision rule for the government. To see this, first, notice that the optimal tax policy $\tau(\epsilon)$ is convex in ϵ, so by Jensen's inequality $\mathbb{E}(\tau(\epsilon)) > \tau(\mathbb{E}(\epsilon))$. Now, notice that the decision rule $\tau^* = \tau(\mathbb{E}(\epsilon))$ is consistent for as long as the Bayesian posterior estimate $p(\epsilon|Y, X, Z, W)$ is properly identified. In that case $\mathbb{E}(\epsilon|Y, X, Z, W) \rightarrow^p \epsilon$. By the the Continuous Mapping Theorem, $\tau(\mathbb{E}(\epsilon|Y, X, Z, W)) \rightarrow^p \tau(\epsilon)$. Pareto$\tau(\mathbb{E}(\epsilon))$ can't converge to $\tau(\epsilon)$.

Notice that an additional condition for convergence is that ϵ does not assume values where the optimal tax policy wouldn't be continuous. Making use of the posterior distribution, I can check the probability that ϵ is outside this bound. The above suggests that the decision rule we are looking for is:

Definition 3. *The Plug-In Bayesian optimal tax decision is* $\tau^{d3}_b(\epsilon) = \tau(\mathbb{E}(\epsilon))$ *where* $\tau(.)$ *is the optimal tax policy and* $\mathbb{E}()$ *is taken with respect to the posterior distribution of* ϵ.

Definition 3 yields a consistent decision rule when point estimates drawn from the posterior ϵ are consistent, a common feature of Bayesian regressions under standard regularity conditions, although it doesn't have apparent justification under Bayesian decision theory. In the following subsection, I will show the above decision can be derived by using welfare functions as loss functions in the statistical decision problem.

2.2 Expected Welfare Approach

Now I will develop a alternative approach to the optimal tax decisions under parameter uncertainty that rationalize the consistent decision rule. Instead of a candidate function on top of the welfare maximization problem, consider a problem where taxes are chosen to optimize expected welfare:

$$\max_{\tau} \int_{\epsilon} \int_{i} \alpha_i W(u_i((1-\tau)y + \tau Y(1-\tau) - G, y)) dI(i) p(\epsilon) d\epsilon. \tag{6}$$

The difference between optimal problems (1) and (6) is one of timing. Under problem (1) the government maximizes for every possible value of ϵ and then based on its beliefs will choose a particular marginal tax by following one of the decision rules. Under the optimal taxation problem (6) the government optimizes the expected welfare function, with expectations taken with respect to its beliefs regarding the parameter ϵ. Also note that (6) is a generalization of the optimal tax problem (1), as they are both equivalent under the degenerate belief. Finally, note that one may interpret problem (6) as a statistical decision problem of finding an optimal tax estimator with a loss function defined by the negative of the classical welfare problem (1). Using the same first order condition and envelop theorem I can derive the optimal linear marginal tax formula:

$$\tau^d = \frac{\mathbb{E}_{\epsilon}(\mathbb{E}_i(\alpha_i W'(.) u_{i,c}(1 - \frac{y_i}{Y})))}{\mathbb{E}_{\epsilon}(\mathbb{E}_i(\alpha_i W'(.) u_{i,c})(1 - \frac{y_i}{Y})) + \mathbb{E}_{\epsilon}(\mathbb{E}_i(\alpha_i W'(.) u_{i,c})\epsilon))} \tag{7}$$

The formula above wields a single value for the optimal tax even under parameter uncertainty, eliminating the need for additional decision criteria, as the problem (6) is simultaneously an optimal tax problem and the statistical decision problem. Because of the Jesen's Inequality, problems (1) and (6) do not generally wield the same solution. Under the Bayesian belief, the expectation $\mathbb{E}_\epsilon$ is conditional to data $\{Y, X, Z, W\}$. If we assume $u_{i,c}$ does not depend on the ETI parameter, such as in the standard case of quasi-linear preferences, and $W'()$ also does not depends on it by our own choice of welfare function [4], I may simplify (7).

$$\tau^* = \frac{1 - g_0}{1 - g_0 + \mathbb{E}[\epsilon|Y, X, Z, W]} \tag{8}$$

Note (8) is equivalent to solving (1) and using the consistent decision rule $\tau^{d3}()$, giving a theoretically-grounded rationalization of this consistent decision rule.

2.3 Notes on Government Budget

A problem that I have ignored so far is that parameter uncertainty does not only change problem (1) by making its solution a mapping $\tau(\epsilon)$ instead of a real number in the interval $[0, 1]$. If the policymaker is uncertain about the ETI, it doesn't know how to aggregate income Y will respond to different tax rates, so the government budget set is random.

Instead of solving (1) under a random budget what we do is discount R such that the budget set always binds. This is equivalent to substituting $R(\tau)$ on the objective function. To clarify this matter I will look into a particular case. For now assume $g = 0$ and a degenerate belief such that $\hat{\tau}^* = \frac{1}{1+\hat{\epsilon}}$. Also, assume no income effect. Our estimator is different from the real parameter value but the aggregate earning function $Y(1-\tau)$ will behave in accordance

[4] For a choice of welfare function such that is a function of utility and not only observable variables, I could still easily compute (8) using the Bayesian posterior but only by making using of a structural model for individuals preferences.

with the real ETI. Notice the optimal tax function is decreasing in ϵ. The marginal revenue function is:

$$D'(\tau) = Y \times (1 - \frac{t}{1-t}\epsilon). \tag{9}$$

Evaluating it on the perceived optimal tax yields

$$D'(\hat{\tau}^*) = Y \times (1 - \frac{\epsilon}{\hat{\epsilon}}), \tag{10}$$

which is negative for $\hat{\epsilon} < \epsilon$. So whenever the estimator $\hat{\epsilon}$ is smaller then the true parameter value, we choose an optimal tax such that we collect insuficient revenues to carry out the optimal transfer program. As consequence, the chosen transfers will not be optimized. The properties of optimal tax problem (1) where fiscal policy can be entirely reduced to either taxes or transfers do not hold under parameter uncertainty. Under the framework here developed, the government must make a choice between optimal taxes or optimal transfers, a practical fact that is commonly overlooked.

2.4 Tax Reform and Laffer Curve Slope

Now I turn my attention to a different problem in public finance where the ETI also plays an important role and the Bayesian posterior $p(\epsilon|Y, X, Z, W)$ can shed new light: the analyses of Laffer Curve slope. Again let $D(\tau) = \tau Y(1-\tau, R)$ be the government revenue function. For now, let's assume there are no income effects $Y(1-\tau, R) = Y(1-\tau)$ so I may focus solely on the impact of the tax change on tax revenue. The marginal revenue function given by (9) can be used to analyze whether for a given marginal tax rate, raising the marginal income tax across the population would raise or decrease tax revenue from a small change. Under the assumptions: (i) $Y(1-\tau)$ is strictly increasing, and (ii) $D(\tau)$ is strictly concave, one finds the revenue-maximizing point $\tau^*_{LC} = \frac{1}{1+\epsilon}$ — the same optimal tax as the previous welfare analysis under Rawlsian welfare function.

Note both considerations regarding the Laffer Curve — its current slope and the revenue-maximizing taxes — relies on estimates of the ETI parameter. This estimate, however, is uncertain and so are any considerations regarding those two. To account for this uncertainty one may use draws from the posterior distribution to simulate draws from the marginal revenue $d(\tau)_s = Y(1-\tau_s/(1-\tau_s)\epsilon)$ where τ_s are draws from the posterior revenue maximizing distribution, $\tau(\epsilon_s)$ and ϵ is fixed on the mean of $p(\epsilon|Y, X, W, Z)$. This plot gives a look at the slope of the Laffer Curve around the revenue-maximizing point, weighted by the probability mass of the parameter ϵ.

Tax Reform Evaluation The formula above only deals with local changes, however. Suppose now that a policymaker wishes to evaluate the impact of tax reform on government deficit. The Bayesian posterior will let us compute this impact taking into account the uncertainty surrounding the ETI estimate. This is useful to answer questions such as the probability that the tax change will raise or reduce a certain amount of funds.

To illustrate this, suppose we are evaluating the change in revenues given a change in one specific tax bracket. The variation in revenues is $\Delta D = \tau_1 Y_1 - \tau_0 Y_0$, where τ_j is the marginal income tax, Y_j are the aggregate specifically to that bracket [5] and $j = 0, 1$ are before and after the reform, respectively. I can rewrite this expression as $\Delta D = (\tau_1 - \tau_0)Y_0 - \tau_1(Y_1 - Y_0)$. The term $\Delta M = (\tau_1 - \tau_0)Y_0$ is the mechanical effect of the tax change over tax revenues and does not relies on individuals' response to taxation, thus it does not depend on the ETI.

The behavioral effect $\Delta B = \tau_1(Y_1 - Y_0)$ accounts for revenue change provoked by changes in incentives in an individual's income supply. Assuming the ETI is constant across individuals and taxation, I rewrite the behavioral effect as $-\frac{\tau_1}{1-\tau_0}Y_0(\tau_1 - \tau_0)\epsilon$. This term depends on the ETI estimator. With draws from the ETI posterior I may simulate the posterior bahevioral effect distribution. Add to the mechanical effect, we have a total revenue change

[5] $Y_j = n(Y^m_j - \bar{Y})$ where n is the number of individuals in that tax bracket, Y^m_j is the average income in that bracket and $\bar{Y}$ is the minimum income threshold for that specific bracket.

distribution

$$\Delta D^{(s)} = \Delta M - \frac{\tau_1}{1-\tau_0} Y_0(\tau_1 - \tau_0)\epsilon^{(s)} \tag{11}$$

Which is the probability distribution of revenues variation given the change in marginal tax policy.

3 Interval Estimation

In the previous section, I discussed the problem of making the optimal tax decision under the parameter uncertainty of the sufficient structural parameter. Despite the parameter uncertainty regarding the structural parameter, those decisions do not quantify uncertainty regarding the optimal tax rule. In this section, I discuss this second problem where policy maker must make a decision regarding an interval optimal tax interval $(\tau_l, \tau_u)_\alpha$ where α is the level of tightness of this interval [6].

Confidence intervals are the standard way to quantify uncertainty in the frequentist framework. This leads to the first alternative where one just plug-in the ETI confidence interval in the optimal tax formula:

Definition 4. *The Plug-In Optimal Tax Interval* $(\tau_l^{d1}, \tau_u^{d1})_\alpha$ *is given by* $\tau_l = \tau^*(\epsilon_l)$ *and* $\tau_u = \tau^*(\epsilon_u)$ *where* $\tau^*()$ *is the optimal tax function and* $(\epsilon_l, \epsilon_u)_\alpha$ *is the* $(1-\alpha)$ *confidence interval estimator for* ϵ. *"*

Which is justified by applying the delta method. One failing of this interval though, is that it doesn't incorporate prior information with respect to ϵ, which I remediate by using the Bayesian framework.

To construct an optimal tax using Bayesian posterior I consider a pair (τ_l, τ_u) such that: (i) $\tau_u > \tau_l$, (ii) $Prob(\tau^* \in [\tau_l, \tau_u]) = 1 - \alpha$ and (iii) the interval is small, that is $\tau_u - \tau_l$ incurs in a loss. This is achieved by the interval in the following definition:

[6] For example, the probability the optimal tax is within the interval

Definition 5. *The Bayesian pseudo optimal tax interval is given by* $(\tau_u^{d2}, \tau_l^{d2})_\alpha$ *such that it minimizes*

$$(\tau_u - \tau_l) + \frac{2}{\alpha}\int_{-\infty}^{\alpha}(\tau_l - \tau(\epsilon)p(\epsilon|Y, X, Z, W)d\epsilon + \frac{2}{\alpha}\int_{\alpha}^{\infty}(\tau(\epsilon) - \tau_u)p(\epsilon|Y, X, Z, W), d\epsilon, \quad (12)$$

which is equivalent to plug-in equal-tailed credible set of the posterior ϵ on τ^*. Using this method part of the prior information with respect to ϵ is incorporated through the posterior of ϵ. However, this simulated distribution is not the real posterior distribution of optimal taxes. As a consequence, the credible set $(\tau_u^{d2}, \tau_l^{d2})$ does not have the correct coverage and has excess width.

To simulate draws from the real I will implement a strategy similar to the one used to derive the point estimator τ^{d3} — to solve the optimal tax and statistical decision problem simultaneously.

Definition 6. *A optimal tax equal-tailed credible interval is given by* $(\tau_u^{d3}, \tau_l^{d3})_\alpha$ *such that it minimizes*

$$\tau_u - \tau_l + \frac{2}{\alpha}\int_{-\infty}^{\alpha}(\tau_l - \tau(\epsilon)p(\tau(\epsilon)|Y, X, Z, W)d\tau + \frac{2}{\alpha}\int_{\alpha}^{\infty}(\tau(\epsilon) - \tau_u)p(\tau(\epsilon)|Y, X, Z, W)d\tau, \quad (13)$$

Where $\tau(\epsilon)$ *is the posterior distribution of optimal taxes defined by transforming the posterior* $\epsilon|Y, X, W, Z$ *using the optimal tax formula* $\tau^* = \frac{1-g}{1-g-\epsilon|Y,X,W,Z}$.

I do not provide closed-formed solutions for $(\tau_u^{d3}, \tau_l^{d3})_\alpha$ but simulating it is trivial by just plug-in draws of the posterior ϵ on τ^* and getting quantiles of that distribution.

4 Simulation Study

Empirical Strategy I will now present the empirical model which is used to perform the simulations in the following section. First, note a direct regression of household income on $ln(1-\tau)$ does not identify the ETI because residual income ($y_{i,t}$) shocks may correlate to an individual's tax bracket. In order to overcome identification issues, I will use a two-way fixed effect difference-in-differences (DiD) research design common in the ETI literature. While DiD estimators of the ETI have been criticized, in this particular simulation they are unbiased and consistent by design. The identification issues in DiD model stem from the distribution of the (potential) income process which can be addressed in the data-generating process directly. Also, the Bayesian 2SLS which I use to estimate ETI can be directly applied to more recent IV-based estimates such as Weber (2014).

In order to identify the ETI, researchers use changes in tax brackets induced by tax reforms. Usually changes in tax laws are different across different brackets, which allows us to interpret individuals in different brackets as control and treatment groups. The regression equations are given by

$$\ln y_{i,t} = \theta_i + \gamma_t + \epsilon \ln(1-\tau_{i,t}) + e_{it} \tag{14}$$

$$\ln(1-\tau_{i,t}) = \lambda 1\{i \in T\} 1\{t \geq \bar{t}\} + \eta_{i,t}, \tag{15}$$

where θ_i and γ_t are the individual's and time-based fixed effects, respectively. The instrumental variable is the interaction dummy between being in the group with higher marginal tax change, the treatment group and a dummy for all the periods after the tax reform, denoted by $\bar{t}$. However, since I'm not interested in fixed effects as they don't provide any policy-relevant information, I take first differences.

$$\ln y_{i,t+1}/y_{i,t} = \gamma_t + \epsilon \ln(1-\tau_{i,t+1})/(1-\tau_{i,t}) + e_{it} \tag{16}$$

$$\ln(1-\tau_{i,t}) = \lambda 1\{i \in T\}1\{t = \bar{t}\} + \eta_{i,t}. \tag{17}$$

The central identifying assumption is that the income variation across individuals after tax reform is driven solely by the reform. There are many reasons why these identifying assumptions would not hold in reality. For example, changes in income inequality such that the income growth rate across treatment and control groups differs violates the parallel trend hypothesis. And even in the absence of those, in panel analysis such as this, there's a issue of mean reversion as taxable income for higher brackets has a substantial transitory component causing variations on the taxable income which cannot be attributed to the marginal tax rates. These identification issues are commonly dealt with by introducing additional exogenous variables in equation (14) such as using base-year income splines. Weber (2014) breaks away significantly from this framework by proposing other kind of instruments. The Bayesian estimation I discuss here is general enough to account for both modifications of this basic model.

Bayesian 2SLS This estimation procedure is drawn from Rossi *et. al.* (2005) and Lopes and Polsen (2014). Suppose I have the simultaneous equation regression model.

$$y = \beta x + w'\gamma + e_1 \tag{18}$$

$$x = z'\delta + e_2 \tag{19}$$

$$\begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \sim \mathcal{N}(0, \Sigma) \tag{20}$$

Where y is the dependent variable, x is a single endogenous independent variable, w is a vector of exogenous independent variables and z is a vector of control variables. The parameter prior distributions:

$$\begin{pmatrix} \beta \\ \gamma \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \beta_0 \\ \gamma_0 \end{pmatrix}, V_{\beta,0} \right) \tag{21}$$

$$\delta \sim \mathcal{N}(\delta_0, V_{\delta,0}) \tag{22}$$

$$\Sigma \sim \mathcal{IW}(v_{\Sigma,0}, V_{\Sigma,0}) \tag{23}$$

These are conforming priors, standard in Bayesian IV application. The most important hyperparameter choice is done regarding β, which in our specific application is the ETI parameter. To keep our simulations more general and not subject to this specific prior choice, I draw the hyperparameter β_0 randomly from $N(0,1)$. This hyperparemeter distribution emulates the fact that the majority of ETI estimators in the literature are within this interval[7].

It's not possible to estimate this model via direct sampling as the joint posterior distribution does not have a closed-form solution. Instead I estimate using Gibb's Sampling from the conditional posteriors

$$\beta, \gamma | \delta, \Sigma, Y, X, Z, W \sim \mathcal{N}(\mu_{\beta,p}, V_{\beta,p}) \tag{24}$$

$$\delta | \beta, \gamma, \Sigma, Y, X, Z, W \sim \mathcal{N}(\mu_{\delta,p}, V_{\delta,p}) \tag{25}$$

$$\Sigma | \beta, \gamma, \delta, Y, X, Z, W \sim \mathcal{IW}(V_{\Sigma,p}), \tag{26}$$

[7]See list of literature parameters in the appendix.

where to draw for β and γ I use

$$V_{\beta,p} = (V_{\beta,0}^{-1} + \sum_{i=1}^{n} \tilde{x}_i \tilde{x}_i')^{-1} \tag{27}$$

$$\mu_{\beta,p} = V_{\beta,p}(V_{\beta,0}^{-1}\mu_{\beta,0} + \sum_{i=1}^{n} \tilde{x}_i \tilde{y}_i,') \tag{28}$$

$$\tilde{x}_i = (1, x_i)'/\sqrt{\sigma_{1|2}} \tag{29}$$

$$\tilde{y}_i = ((y_i - (x_i - z_i'\delta)\sigma_{12}/\sigma_{11})/\sqrt{\sigma_{1|2}} \tag{30}$$

$$\sigma_{1|2} = \sigma_{22}(1-\rho^2), \tag{31}$$

where σ_{ij} are the respective ij entries in the variance covariance matrix Σ. Notice that this algorithm is only valid for the jointly normal errors case and for a specific covariance structure $\sigma_{12} = \rho\sigma_{11}\sigma_{22}$ as the the conditional error $\sigma_{1|2}$ is derived from the normal projection property.

A similar strategy is used for the δ draws

$$V_{\delta,p} = (V_{\delta,0}^{-1} + \sum_{i=1}^{n} \tilde{z}_i \tilde{z}_i')^{-1} \tag{32}$$

$$\mu_{\delta,p} = V_{\delta,p}(V_{\delta,0}^{-1}\mu_{\delta,0} + \sum_{i=1}^{n} \tilde{z}_i \tilde{x}_i \tag{33}$$

$$\tilde{x}_i = (x_i - (y_i\gamma - \beta x_i)\sigma_{12}/\sigma_{1|2} \tag{34}$$

$$\tilde{z}_i = z_i/\sqrt{\sigma_{1|2}} \tag{35}$$

$$\sigma_{1|2} = \sigma_{22}(1-\rho^2). \tag{36}$$

Finally for Σ draws

$$v_{\Sigma,p} = v_{\Sigma,0} + n \tag{37}$$

$$V_{\Sigma,p} = V_{\Sigma,p} + \sum_{i=1}^{n} e_j e_j'. \tag{38}$$

Refer to Rossi Rossi et al. (2012) for more details about the derivation of the formulas. I can simulate draws from $p(\epsilon|Y, X, Z, W)$ distribution by sampling $\beta^{(s)}$ from the algorithm described above, after discarding some number of initial iterations.

4.1 Data Generating Process

I assume every individual has the quasi-linear preferences $u(c, y) = c - \epsilon/(1+\epsilon){y^0}^{\frac{1}{\epsilon}} y^{1+1/\epsilon}$ with constant ϵ. This formulation is very close to the one in Saez et al. (2012). I chose this preference because it wields a Marshallian income supply y that is increasing on the net-of-marginal-tax rate $1-\tau$ and because y does not have income effects. The first trait is consistent with the findings of empirical ETI literature. Assuming away income effects is a necessary step for identification, as discussed in the preview section. Suppose all individuals face the non-linear tax schedule:

$$T(y) = \begin{cases} \tau_l y & \text{if } y \leq \bar{y} \\ \tau_h \bar{y} + \tau_h (y - \bar{y}) & \text{if } y > \bar{y}. \end{cases}$$

Solving for the Marshallian income supply

$$y(1-\tau) = \begin{cases} y^0 (1-\tau_l)^\epsilon & \text{if } y^0 < \frac{\bar{y}}{(1-\tau_l)^\epsilon} \\ \bar{y} & \text{if } \frac{\bar{y}}{(1-\tau_l)^\epsilon} \leq y^0 \leq \frac{\bar{y}}{(1-\tau_h)^\epsilon} \\ y^0 (1-\tau_h)^\epsilon & \text{if } y^0 > \frac{\bar{y}}{(1-\tau_h)^\epsilon}, \end{cases}$$

where $y(1-\tau)_{i,t}$ is determined in all periods by the pre-reform tax schedule $\{\tau_{l0}, \tau_{h0}, \bar{y}^0\}$, where $r \sim U(r_{min}, r_{max})$, and the potential income y^0 which is randomized for every individual on the initial period according to $N(\mu_0, v_0)$ and evolves across time following an AR(1) process $y_t^0 = \alpha y_{t-1}^0 + u_t$ with $u_t \sim N(0, v_u)$. The panel data is simulated for H periods. During period T the tax schedule exogenously changes to $\{\tau_{l1}, \tau_{h1} + r, \bar{y}\}$. The simulation was repeated $S = 20,000$ times.

Under this DGP it is possible for certain individuals to switch between control and treatment groups because of income shocks, muddying the identification of the ETI. On this simulation of the dynamic panel, this threat is handled by tracking and excluding such individuals.

Bunching is a potentially bigger concern in identifying the ETI. The individuals closer to the center of the bunching set are less likely to leave it, reducing the impact of the tax reform. Individuals closer to the borders of the bunching set are likely to leave or enter it, but even in that case, their responses do not identify the ETI. To avoid both those issues I also exclude from the analysis all individuals who enter the bunching set in set in pre or post reform periods.

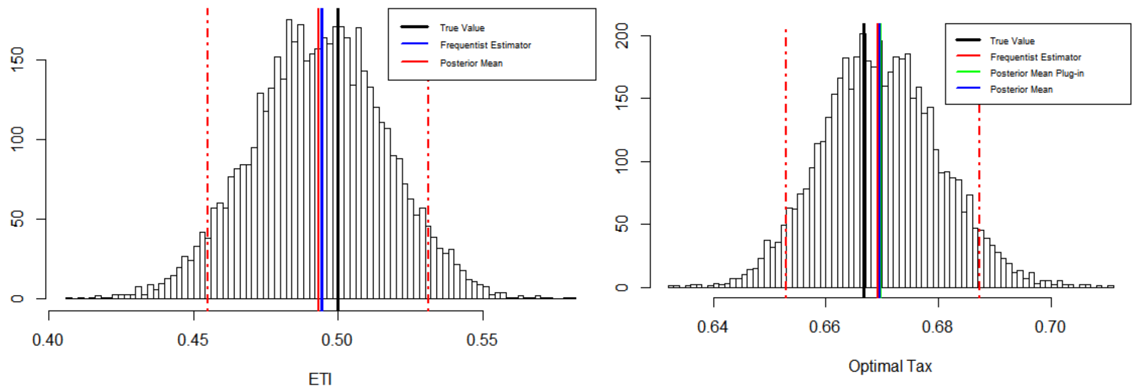
For the choice of simulation parameters, refer to Appendix A. The simulation exercise presented in the section is fairly robust to any choice of those. The optimal taxes are assumed to follow decision rules τ_b^d with $g_0 = (0, 0.2, 0.5)$.

4.2 Results

First, I will single out one simulation to illustrate the results in the previous sections. Then I will analyze the performance of point and interval estimators for the entire simulation. Figure 1 showcases a sample of the full density of posterior ETI and the implied posterior revenue-maximizing tax for the highest tax bracket. One can observe from this sample simulation that all three point-estimators proposed perform relatively similar in terms of approaching the true optimal tax, independent of decision rules, for the range of welfare functions considered in this simulation.

Notice how the non-linearity of the optimal tax formula affects the uncertainty by skewing

Figure 1: Example Histogram of Posterior ETI and posterior optimal taxes using τ_b^d decision rule.



the distribution to the left. For example, the equal tailed credible interval for the ETI with 90% credibility has the bounds 46.62 and 53.15. The respective equal tailed interval for revenue maximizing taxes are 55.64% and 58.85%. With this distribution on hand, policy makers know they shouldn't propose the top-marginal tax to be any higher then 58% as that tax is very likely to be beyond the Laffer Curve maximum. Also, by looking at ETI posterior distribution, I assure that ϵ have no significant probability of assuming values that would enter the discontinuous domain $\tau()$ and violate the convergence conditions.

Figure 2: Example Histogram of Posterior Marginal Revenues and Total Revenue Change.

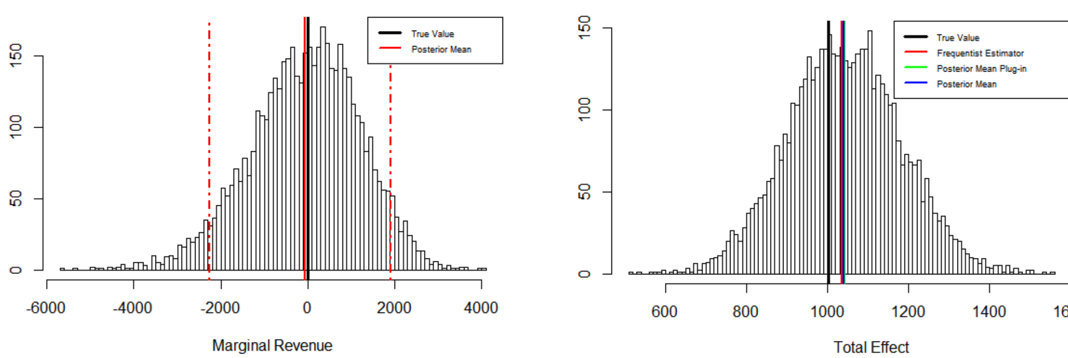


Figure 2 shows a sample histograms of the posterior marginal revenue distribution and the behavioral effects. The marginal revenue posterior gives a wide look into the Laffer

Table 1: Point Estimator Evaluation

Decision Rule	RMSE (10^{-1})
Definition 1: Classic Plug-In	1.2326
Definition 2: Bayesian Absolute Loss	2.2834
Definition 2: Bayesian Quadratic Loss	9.4438
Definition 3: Bayesian Expected Welfare	1.3664

Curve slope and the equal-tailed maximum. Both this plot and the optimal tax can tell us how likely it is set to tax policy policymaker of objective desirability but the marginal tax gives a measure, weighted, by the probabilities, of possible revenue losses due to taxation.

The second histogram of Figure 2 refers to the analysis of total revenue change given a certain tax reform that only affects the top income bracket. For this example specifically, I am increasing the top bracket marginal tax from 45% to 60%. The plot refers to the distribution of revenue variation assuming ETI is constant and ignoring possible changes in the tax base. This distribution could be very useful for the design of tax policy. As argued in the previous section, tax reforms are often designed with deficits restrictions or other fiscal policy objectives in place. The posterior of the revenue variation given a tax change can tell us the probability that alternative tax reform will achieve pre-determined fiscal objectives. For example, if this government needs a surplus of 1000 per capita for the next year, the probability that this tax reform delivers is 92.63%.

Estimator Performance Now I will compare the result of point and interval estimators using alternative decision rules. To gauge the performance of the optimal tax point estimators, I calculate the Root Mean Squared Errors (RMSE) taking the real optimal tax as the population parameter. The results are shown in Table 2. All estimators are Bayesian 2SLS, except for the the first one who's a traditional frequentist DiD plug-in decision. I also provide results using absolute difference instead of the quadratic loss for the second Bayesian decision rule.

As expected I observe a very significant loss of performance in terms of RMSE when

Table 2: Interval Estimator Evaluation

Decision Rule	Coverage	Interval Measure
Definition 4: Delta Method	.90	.47
Definition 5: Plug-In	.77	.83
Definition 6: Credible Set of Posterior Optimal Tax	.89	.11

one uses the Bayesian quadratic optimal tax decision and a relatively smaller loss when Bayesian absolute loss is used. This in accordance with my theoretical results where the rationalized optimal tax decision requires a reformulation of the decision problem using the welfare function, otherwise point estimators derived from Bayesian posteriors are biased in finite samples. My proposed Expected Welfare decision correct this distortion and performs similarly to using plug-in frequentist estimators.

Table 2 shows the results for the comparison of interval estimates. To evaluate those estimates I used Interval Measure (IM) besides just coverage. The IM is given by the average of the distance between the real optimal tax and the closest bound of the interval estimate. A lower IM is good for as long as it has correct coverage. The Bayesian Plug-In has both large coverage distortion and large interval measures. Using the Delta Method to derive intervals for the optimal tax wields intervals with correct coverage but since prior information is not used, it has excess IM. My proposed method presents a minimal size distortion of 1% but has a substantially lower IM compared to the Delta method.

5 Conclusions and Further Research

I developed a model where policy makers do not face the decision about the optimal tax policy and transfers alone but the statistical decision about estimating the necessary and sufficient parameter for the optimization, ϵ. I derived a decision rule for point estimation of optimal taxes that is both consistent and incorporate prior information. Remarkably, this derivation required a reformulation of the optimal taxation problem by merging the two decision problems, the policy design and the statistical one. In addition, when taking into account the uncertainty of the parameter, it is revealed that it is not possible to estimate the optimal tax and the optimal transfers simultaneously. I also derived a decision rule for interval estimation that has the correct coverage and is substantially smaller than simply plug-in in the bounds of confidence and credible sets in optimal tax formulas, a common practice in optimal tax literature.

While the theoretical problem I considered is highly stylized, the results can be easily applied to more sophisticated normative analysis for as long as closed-form solutions for optimal policy are readily available.

A Appendix A - Simulation Parameters

DGP Parameter		Prior Parameters	
N	1000	S	5000
H	2	S^{-}	50
T	1	β_0	$N(0,1)$
ϵ	0.5	γ_0	0
μ_0	30000	σ_{11}	10
v_0	35000	σ_{22}	0.1
α	1.05	σ_{12}	0
v_u	300	δ_0	0
τ_{l0}	0.3	$V_{\delta,0}$	0.1
τ_{h0}	0.31	$v_{\Sigma,0}$	5
τ_{l1}	0.3	$V_{\Sigma,0}$	$5I$
τ_{h1}	0.45		
$\bar{y}$	25000		

B Appendix B - List of Literature ETI Estimates

Paper	ϵ Estimate
Feldstein (1995)	0.12
Auten and Carroll (1999)	0.55
Moffitt and Wilhelm (1998)	0.35 - 0.97
Gruber and Saez (2002)	0.12
Kopczuk (2005)	0.36
Giertz (2007)	0.30
Giertz (2010)	0.19 - 0.43
Kleven and Schultz (2014)	0.186 - 0.257
Weber (2014)	0.858
Doerrenberg et al. (2017)	0.54 - 0.68

References

Auten, G. and R. Carroll (1999). The effect of income taxes on household income. *Review of economics and statistics 81*(4), 681–693.

Chetty, R. (2009). Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods. *Annu. Rev. Econ. 1*(1), 451–488.

Doerrenberg, P., A. Peichl, and S. Siegloch (2017). The elasticity of taxable income in the presence of deduction possibilities. *Journal of Public Economics 151*, 41–55.

Feldstein, M. (1995). The effect of marginal tax rates on taxable income: a panel study of the 1986 tax reform act. *Journal of Political Economy 103*(3), 551–572.

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

Giertz, S. H. (2010). The elasticity of taxable income during the 1990s: new estimates and sensitivity analyses. *Southern Economic Journal 77*(2), 406–433.

Gruber, J. and E. Saez (2002). The elasticity of taxable income: evidence and implications. *Journal of public Economics 84*(1), 1–32.

Heathcote, J., K. Storesletten, and G. L. Violante (2017). Optimal tax progressivity: An analytical framework. *The Quarterly Journal of Economics 132*(4), 1693–1754.

Kleven, H. J. and E. A. Schultz (2014). Estimating taxable income responses using danish tax reforms. *American Economic Journal: Economic Policy 6*(4), 271–301.

Kopczuk, W. (2005). Tax bases, tax rates and the elasticity of reported income. *Journal of Public Economics 89*(11-12), 2093–2119.

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *The review of economic studies 38*(2), 175–208.

Moffitt, R. A. and M. Wilhelm (1998). Taxation and the labor supply: Decisions of the affluent. Technical report, National Bureau of Economic Research.

Rossi, P. E., G. M. Allenby, and R. McCulloch (2012). *Bayesian statistics and marketing.* John Wiley & Sons.

Saez, E., J. Slemrod, and S. 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.

Sheshinski, E. (1972). The optimal linear income-tax. *The Review of Economic Studies 39*(3), 297–302.

Slemrod, J. (1996). High-income families and the tax changes of the 1980s: the anatomy of behavioral response. pp. 169–192.

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