A Structural Approach to Income Elasticity Measurement

By Thomas Aronsson, Katharina Jenderny, and Gauthier Lanot

Measuring the elasticity of taxable income (ETI) is central for tax policy design. Yet, there are few arguments which support or infirm that current methods yield measurements of the ETI that can be trusted. We have shown in a related simulation study that estimators based on indirect inference principles can be expected to produce more precise estimates of the ETI than any of the most commonly used methods. Our purpose in this paper is to apply indirect inference estimation to Swedish tax data. This is the first application of this method to empirical data in the field of the ETI measurement.

JEL: H24; H31; D60
Keywords: Elasticity of Taxable Income, Income Tax, Indirect Inference

In this paper, we apply a new, indirect-inference-based method to estimate the elasticity of the taxable income with respect to the marginal net-of-tax rate (ETI) to Swedish data. The ETI is a central statistic for tax policy design. In fact, it is often referred to as “sufficient statistic”, i.e., a parameter that provides information on the behavioral response to marginal taxation along all relevant margins (Feldstein, 1995). Yet, although measurable in principle, there is little agreement on the magnitude of the ETI. The two dominating methodological approaches, the instrumental variable (IV) regression-based approach and the bunching approach, fail to produce similar point estimates for the ETI. Point estimates based on the regression approach typically exceed

* Department of Economics, Umeå School of Business, Economics and Statistics, Umeå University, Sweden. ◦ Department of Economics, School of Business and Economics, Freie Universität Berlin, Germany.
Email: thomas.aronsson@umu.se, katharina.jenderny@umu.se, gauthier.lanot@umu.se.
Acknowledgements: Financial support from The Jan Wallander and Tom Hedelius Foundation (project P2016-0140:1), the Swedish Royal Academy of Sciences, and the Swedish Research Council is gratefully acknowledged.

¹The marginal net-of-tax rate is defined as one minus the marginal tax rate.
point estimates based on the bunching approach, sometimes by an order of magnitude, even using the same data. Despite a large body of literature, policy makers are thus still largely in the dark when trying to assess the behavioral response to marginal income taxation.

In earlier work, we have analyzed the performance of the regression-based approach and the bunching approach terms of bias and precision in a Monte-Carlo study (Aronsson, Jenderny and Lanot, 2017). We find that the IV regression estimators may suffer from considerable bias and be very imprecise, whereas the bunching estimators perform better in the model environment we set out in the paper. We also found that the indirect inference estimator is practically unbiased and much more precise than both the bunching and IV estimators under a variety of conditions. We have therefore suggested to measure the ETI with an indirect inference estimator.

In this paper, we apply the indirect inference estimator to Swedish micro data.

The paper is structured as follows: Part I contains a literature review and describes how the present paper contributes to this literature. In part II, we set up a simple behavioral model. In part III, we describe the newly suggested estimator. In part IV, we describe the data source. In part V we present and discuss the results. Concluding remarks are presented in part VI.

I. Literature

Issues

With our newly suggested estimator, we aim at overcoming problems of the hitherto prevailing estimators of the ETI, the regression-based method and the bunching method. We will refer to these two methods as the "conventional methods". We discuss the literature on both conventional methods, first in order to motivate our alternative indirect inference approach, second because we will compare the results of our newly suggested estimator to results based on
conventional methods, and third because we will use elements of both methods as naive statistics in our estimation strategy. After discussing regression-based method and the bunching method we will turn to the literature on indirect inference.

The regression approach seeks to identify the ETI by comparing relative changes in taxable income of tax units or groups of tax units between two periods to relative changes in their net-of-tax rates. In the early literature, group-based comparisons of income shares were conducted on cross-sectional data (see Feenberg and Poterba, 1993; see also the review by Saez, Slemrod, and Giertz, 2012). Feldstein (1995) was the first study to use panel data, which allowed to keep the composition of groups of tax units constant over time to avoid endogenous selection into groups. Later studies used difference-in-difference regressions including further control variables instead of simple group mean comparisons.

The regression-based approach faces two main methodological challenges. First, following tax units over time introduces the problem of mean reversion, which is akin to an initial conditions problem, and describes the correlation between the error term and the dependent variable in first differences. A high-income tax unit in the initial period is likely to have a lower income in the following period for idiosyncratic reasons, irrespective of tax rate changes. The mean reversion problem is typically addressed by controlling for the initial income level. Second, the marginal net-of-tax rate is an endogenous variable to the change of the income level in directly progressive tax systems, and

The approach faces several other obstacles to identification, such as secular trends in income, shifting between tax bases, and data-related problems such as tax reforms that change both tax rates and tax bases, for example. For a discussion see, for example, Saez, Slemrod and Giertz (2012). Yet, mean reversion and the endogeneity of the net-of-tax rate are inherent to the method itself and cannot be solved by data improvements. Our aim is to compare the methods in a framework that is reduced to the core problems of the method, which is why we focus on these two issues, while the other problems are not included in our simulated framework.

As we will argue in part 777, it is not clear that this control variable will solve the missing variables problem.
instrumental variables techniques are used to account for this endogeneity. A common instrument for the change in the net-of-tax rate is a hypothetical change in the net-of-tax rate that uses the tax schedules of the period for which the ETI is to be measured, but applies them only to the start-year income (Gruber and Saez, 2002; Carroll, 1998; Auten and Carroll, 1999). Weber (2014) argued that using the initial income level for the instrumentation does not solve the endogeneity problem, and suggests to use higher-order lags of the income instead.

Regression-based estimates of the ETI differ greatly across methods and countries (e.g. Neisser, 2017), who shows that estimates range roughly between -1.5 and 2). Some of these differences can be explained by conceptual differences, such as different tax systems and taxpayer types. For instance, the ETI is expected to be higher when the share of self-reported income is large, and if deduction possibilities are abundant (Kopczuk, 2005; Doerrenberg, Pechl and Siegloch, 2017; Kleven and Schultz, 2014). Other differences reflect methodological advances, such as the proper instrumentation of endogenous variables. The approach of Feldstein (1995) does not use any instrumentation to account for the endogeneity of the net-of-tax rate, and yields large elasticity estimates, ranging roughly between 1 and 3. Using US data and controlling for lagged income, Gruber and Saez (2002) obtain elasticities between 0.4 and 0.6, depending on the income control. Weber (2014) shows that instrumentation with longer time lags of the instruments leads to baseline estimates of the ETI that are twice as large as those found by Gruber and Saez (2002), using the same data.

The bunching approach provides an alternative method to estimate the ETI. Modern income tax and benefit systems are usually piecewise linear: the marginal tax rate is typically constant within well defined intervals of the taxable income, while it changes in a discontinuous manner between intervals. This creates either kinks or notches to the budget constraint of the tax payer.
At a kink point, the marginal tax rate changes in a discontinuous fashion (e.g., between two proportional tax brackets), while the tax due changes in a discontinuous fashion at a notch. In practice, notches occur less frequently than kink points. We thus focus on kink points when discussing the methodological approach.\(^4\)

Kink points help identifying the ETI based on the behavior of agents located at or close to the kink points. Household tax positions will typically bunch at any kink of the tax schedule. Optimizing individuals choose taxable income so as to equalize their marginal rate of substitution (MRS) between the utility cost of acquiring an additional unit of income (e.g., in terms of leisure foregone) and the disposable income to the marginal net-of-tax rate. At a kink point, economic theory predicts that all tax units whose MRS is less than or equal to the net-of-tax rate in the first, but not the second bracket should generate taxable income up to that kink point exactly. The amount of tax units at the kink point can therefore identify the average responsiveness of tax units to the marginal tax rate (i.e., the ETI), by contrasting the excess mass of bunching tax units to a counter-factual income distribution in the absence of a kink point. Saez (2010); Chetty et al. (2011); Bastani and Selin (2014); Hargaden (2015); Kleven and Waseem (2013) use the amount of bunching relative to a theoretical counter-factual without bunching, to measure the behavioral response to the tax code. Kleven (2016) provides a recent methodological review.

Similar to the regression approach, the bunching approach to measuring the ETI faces methodological challenges. The first challenge is that the construction of a counter-factual income distribution in the absence of a kink point, which is required to operationalize the measurement, is not straightforward, and its functional form matters for the result. Even though some authors claim

\(^4\)For discussions on notches, see e.g., Kleven and Waseem (2013); Hargaden (2015); Kleven (2016); Slemrod (2013)
that the bunching approach is non-parametric, the estimation of a counter-factual income distribution requires functional form assumptions, the statistical consequences of which are often opaque (see also Einav, Finkelstein and Schrimpf, 2017, Blomquist and Newey, 2017, and Bertanha, McCallum and Seegert, 2018). The second challenge is that the bunching typically occurs not only at the exact kink income, but in an interval, which has to be specified by the researcher. The third challenge is that the estimator is local by definition, and its validity is restricted to a particular income level, and possibly household type at that level. The fourth challenge is that optimization frictions are likely to bias the estimates downwards, especially in the case of wage earners (Chetty, Friedman and Saez, 2013; Bastani and Selin, 2014).

Both regression and bunching estimators only use a limited amount of the data available. Regression methods are typically based on a linearization of the budget constraint and do not take into account that tax kinks have an influence on behavior, while bunching methods use cross-sectional information only. By comparison, many earlier studies on the labor supply (see the reviews by Blundell and Macurdy (1999) and Blundell, Macurdy and Meghir (2007)) used maximum likelihood (ML) methods to estimate jointly the behavioral parameters and the parameters that describe the distribution of heterogeneity in the population. In practice, while ML methods are suitable to overcome the limitations of the regression and bunching estimators in general, they are analytically difficult to apply in contexts where the analyst wishes to account for repeated observations over time or/and when the distribution of the un-

5Furthermore, the estimations of the counter-factual income distribution typically use both observations below and above the kink to fit the same distribution. This procedure ignores the fact that the distribution of taxable income depends on the tax rate. The observed distribution above the kink should therefore follow a different functional form than the observed distribution below the kink. Chetty, Friedman and Saez (2013) evade using a functional form, and instead use regional variation in information on the EITC schedule to construct counter-factual income distributions.

6Exceptions in the context of health spendings are Einav, Finkelstein and Schrimpf (2015, 2017), who allow for inter-temporal substitution of spendings.
observed components of the model is not normal. Modern simulation based methods provide a possible alternative by combining the advantages of ML estimation with mathematical feasibility. We suggest to use indirect inference principles (see Gouriéroux, Monfort and Renault (1993) for details) to estimate the ETI.

The indirect inference approach has been applied in different contexts (see Browning, Ejrnæes and Alvarez (2010), Low and Pistaferri (2015), and Nenov (2015), for example), but not in the literature measuring the ETI. It relies on two elements; first, a model which potentially generates the data but depends on a set of unknown parameters, among those the ETI, and, second, a large enough set of auxiliary statistics which can be estimated on the sample data as well as on the simulated data from the model for any value of the parameters. Assuming the theoretical model is a good description of the process that generates the observed data, the values the auxiliary statistics take when measured from the observed data will be similar to the values of the same statistics when measured from the simulated data at the correct parameters. Gouriéroux, Monfort and Renault (1993) show that the parameter values that minimize the distance between the estimated auxiliary statistics obtained from the sample data and the ones obtained from the simulated data will have good statistical properties and in particular (asymptotic) unbiasedness.

The empirical evidence is captured by the measurement on the observed sample data of auxiliary statistics. The inference is indirect since the choice of parameter estimates for the model is guided by the ability of the simulated data to generate auxiliary statistics values that are close (or identical) to the ones the empirical data generates. The method can be applied in any context where it is (relatively) easier to simulate data from a given theoretical model than it is to calculate the moments or the likelihood the theoretical model implies. We argue that this is in general the case in the context of the estimation of the ETI.
The main contribution of our study is to apply indirect inference estimators, which are novel in the context of ETI estimation, to Swedish micro data. and compare the results to results from conventional methods of measuring the ETI (regression and bunching methods)

II. Behavioral Model

General framework

Our suggested indirect inference estimator relies on simulation methods. We will therefore first set up a behavioral model that describes the choice of taxable income in response to the net-of-tax rate. Given this model environment, we will describe the details of the estimator in part III.

The general framework we present here is designed to provide a simple model in which the main methodological questions are nevertheless relevant. We can therefore use this model environment to compare different estimators, but we avoid too much complexity that would make the comparison between estimators more difficult. For this reason, we focus on labor income and assume that individuals respond to the wage they are offered. In response to an offered wage, individuals determine their level of work effort, and together the offered wage and the labor supply determine a given individual’s gross earnings.

The model presented below assumes that labor is the only income source, and that the utility is quasi-linear in private consumption. The reason for these simplifications is that we aim to define the most stylized model framework that will still reflect the core estimation issues in the literature. Both assumptions can of course be relaxed in principle. In fact, we argue in part 4 that the estimator we propose is more suited to relax these assumptions than the methods currently used in the ETI literature. Our exact specification is the one adopted by Saez (2010), and therefore corresponds exactly to the
A structural approach to income elasticity measurement. As we show below, the model also reproduces the general specification used in the regression-based approach, where it is more common to state the reduced-form relationship between the change in income and the change in the net-of-tax rate directly than to formulate a complete structural model (see, e.g., Gruber and Saez (2002)). Consequently, our model provides a favorable framework for the estimators we analyze.

The specification rests on the preferences which yield the basic log linear specification of the labor supply function

$$u(c,h) = c - \gamma \frac{\eta}{1 + \frac{1}{\alpha} \left( \frac{h}{\eta} \right)^{1 + \frac{1}{\alpha}}}$$

where $\alpha$, $\gamma$ and $\eta$ are all positive.\(^7\) The labor supply function takes the form:

$$\ln h^*(w,\eta) = -\alpha \ln \gamma + \alpha \ln w + \ln \eta$$

Equation (2) gives a direct interpretation to the parameters of the utility function: $\alpha$ is the wage elasticity of the labor supply, and both $\eta$ and $\gamma$ describe the disutility of work. $\gamma$ determines the average disutility of work, while $\eta$ is assumed to be one on average and introduces heterogeneity between individuals. $\eta > 1$ corresponds to a below-average disutility of work, while $\eta < 1$ corresponds to an above-average disutility of work. Observe that the specification excludes income effects. Chetty (2012) suggests that $\alpha$ (the labor supply elasticity) = 0.33 is a credible central prior based on his reading of the accumulated (US) evidence on the intensive margin.

The log linear specification has another appealing feature in describing a straightforward relationship between the offered wage, $w$, and gross earn-

\(^7\)Saez (2010) does not specify $\gamma$ as he does not assume $\eta$ to be one on average.
ings/taxable income $wh$. Indeed we have,

$$(3) \quad \ln wh^* = -\alpha \ln \gamma + (\alpha + 1) \ln w + \ln \eta,$$

where $\alpha + 1$ is now the elasticity of earnings relative to the wage.

In the following, we adjust the model in two dimensions that are relevant to most empirical applications. First, we account for the fact that we typically do not observe the wage and the disutility of work, but only earned income. We show that we can treat the variability of both unknown factors as one single unknown component in this case. Second, we introduce taxation in the model, which enables us to define the elasticity of taxable income and to address the endogeneity of the net-of-tax rate to the optimal choice of earnings. The latter is crucial for the empirical estimation using the regression approach.

*The variability of the wage and the disutility of work*

Heterogeneity between agents in this model stems from the variability of the wage offer and the variability of the disutility of work. Yet, we only observe earned income which, in turn, depends on both $w$ and $\eta$. It is therefore useful to understand the possible interactions between the two. Observe that it is always possible to rewrite the preferences over a bundle $(c, h)$ given the disutility of work $\eta$ (as defined in Equation (1)), as the utility over a bundle $(c, wh)$ given the wage $w$ and the disutility of work $\eta$, i.e.,

$$(4) \quad u(c, h) = c - \gamma \frac{\eta}{1 + 1/\alpha} \left(\frac{wh}{w\eta}\right)^{1+\frac{1}{\alpha}},$$

and we deduce the preference over consumption and earnings,

$$(5) \quad v(c, wh) = c - \gamma \frac{1}{1 + 1/\alpha} (wh)^{1+\frac{1}{\alpha}} (w^{\alpha+1}\eta)^{-\frac{1}{\alpha}}.$$
When specified in this fashion the disutility of work depends on the quantity

\[ \omega \equiv w^{\alpha+1} \eta. \]

The behavioral assumption requires then that the worker determines earnings, \( wh \), so as to maximize \( v(c, wh) \) such that \( c = wh + R \), with \( R \) being unearned income. Optimal earnings are then:

\[ \ln wh^* = -\alpha \ln \gamma + \ln \omega. \quad (6) \]

This property suggests that the marginal distribution of earnings depends on the distribution of \( \omega \) only. In the absence of any information about the individual wage, the distribution of optimal earnings will only be informative about the distribution of \( \omega \) overall and not about the distribution of its individual components \( w \) and \( \eta \). In our simulations, we can thus reduce heterogeneity to the combined unknown component \( \omega \).

**Taxation**

In order to use the model to predict an individual’s change in optimal earnings in response to a tax rate change, we need to explicitly model income taxation. Assume that the individual’s optimal labor supply choice is on a regular part of the the budget constraint (i.e., not situated at a kink or a discontinuity). It must then satisfy:

\[ \ln h^* = -\alpha \ln \gamma + \alpha \ln \left( \tau_c[wh^*, x]w \right) + \ln \eta, \quad (7) \]

where the amount of tax paid, \( T(wh, x) \), depends on the level of earnings as well as on other variables observed or unobserved, \( x \). The variable \( \tau_c[wh, x] \equiv 1 - \tau[wh, x] = 1 - \frac{\partial T}{\partial wh} \) is the marginal net-of-tax rate, defined as one minus the marginal tax rate.
We can re-formulate Equation (7) to describe the optimal level of earnings, which takes the form:

\[(8) \ln \wh^* = \kappa + \alpha \ln \tau^c[\wh^*, \bx] + \ln \omega\]

with \(\kappa \equiv -\alpha \ln \gamma\). \(\alpha\) measures the response of earnings to a marginal increase in the net-of-tax rate, i.e., the ETI. The expression in Equation (8) sets a starting point for the methodological discussion concerning the estimation of \(\alpha\) where \(\ln \omega \equiv (\alpha + 1) \ln \w + \ln \eta\) plays the role of the unobserved component.

The specification developed in Equation (8) is the basis for most empirical measurements of the ETI. It is the simplest framework for understanding the effect of a tax system on the distribution of earnings. In most cases it is possible to trace back the empirical specification to our theoretical specification, and in particular to interpret the parameter of the net-of-tax rate as the ETI. When using bunching methods, information from a single cross section is in principle sufficient to obtain a measurement of the ETI within this exact modeling framework.

Equation (8) is only apparently linear in the unobserved component, since the unobserved component \(\ln \omega\) determines the level of earnings and, in turn, the marginal net-of-tax rate, \(\tau^c[\wh, \bx]\). In the presence of a complex tax system where the marginal tax rate varies with earnings, the relationship between the wage and earnings is no longer linear.\(^8\) The net-of-tax rate, apparently a regressor in Equation (8), is endogenous, i.e., optimal earnings and the net-of-tax rate are determined jointly.

We use the model specified above in order to simulate data and compare that simulated data with observed microdata in our indirect inference estimator.

\(^8\)Figure ?? in Appendix ?? illustrates this fact in a simple piecewise linear case and in a smooth case.
III. Indirect Inference Estimator

Given the specification of the earnings function (and in the absence of any information on the wage), all observations and all earnings histories are informative about the ETI. Yet, each of the conventional approaches described in section I uses only a fraction of the available information. The regression approach only focuses on differences and disregards the information on bunching, while the bunching approach only uses the cross-sectional information of observations near the kink point. Indirect inference allows us to combine the information from the earnings levels and earnings growth in order to obtain better behaved estimators of the ETI. While our particular focus here is on a simple behavioral model, the same principle can be extended to more demanding environments.

The indirect inference approach relies on an assumption concerning the data generating process. In our case, the data is generated according to the economic model structure (preferences and constraints) and the assumptions concerning the structure and distribution of the unobserved components. The maintained hypothesis throughout is that the modeling structure is the correct one, however the exact parameter values which generate the sample data at hand are unknown. Denote \( \Xi \) the vector of parameters of the model (under the restrictions we discussed in the model section)

\[
\Xi = (\alpha, \kappa, \sigma, \phi, \rho),
\]

and denote \( \Xi_0 \) the particular parameter vector which generates the data. \( \Xi_0 \) is unknown and the statistical problem is concerned with its measurement.

On the basis of the observed data we can measure the auxiliary statistics, which contain information about \( \Xi_0 \), and we denote this measurement \( \hat{s}_0 \equiv \hat{s}(\Xi_0) \). Unfortunately, we are unable in general to retrieve directly an estimate of \( \Xi_0 \) from \( \hat{s}_0 \). The relationship between the parameter vector which generate
the data and the vector of auxiliary statistics is typically too complex or even beyond our ability to characterize completely to be able to "inverse" it.

The method of indirect inference suggests that we can obtain good estimates of $\Xi_0$ by using the modeling structure to simulate synthetic data and calculate the vector of auxiliary statistics on the synthetic data given a guess, $\Xi_g$, for the value of the parameter vector that generates the data. This simulation step provides us with a simulated measurement $\hat{s}(\Xi_g)$ of the auxiliary statistics.

While it is not feasible to determine directly how far a given guess $\Xi_g$ is from the true value of the parameters $\Xi_0$, we are able to evaluate the distance between the vectors of auxiliary statistics $\hat{s}_0$ and $\hat{s}(\Xi_g)$. The intuition is therefore that if $\Xi_g$ is such that $\hat{s}(\Xi_g) \approx \hat{s}_0$, i.e., $\Xi_g$ allows the simulated auxiliary statistics to match the observed values of the auxiliary statistics, then $\Xi_g \approx \Xi_0$. $\Xi_g$ is then a good estimate of the true value of the parameter vector.

An indirect inference (I-I) estimator for $\Xi_0$ is then the vector $\hat{\Xi}_{I-I}(\hat{s}_0)$ which minimizes the distance $Q(\Xi_g, \hat{s}_0, A)$ between the observed and the simulated auxiliary statistics. A natural distance is the Euclidian distance

$$Q(\Xi_g, \hat{s}_0, A) = (\hat{s}_0 - \hat{s}(\Xi_g))'A(\hat{s}_0 - \hat{s}(\Xi_g)),$$

where $A$ is a constant positive definite matrix. In effect, we are using the minimization of the objective $Q(\Xi_g, \hat{s}_0, A)$ to solve approximately the equation $\hat{s}_0 = \hat{s}(\Xi)$ for $\Xi$.

The large sample theory in terms of the number of observations and the number of simulated observations is well developed and will apply to our context directly. In particular, the I-I estimator is asymptotically normally distributed for any positive definite $A$. Given a sensible choice for $A$, estimates of the asymptotic variance covariance matrix of the estimator exist and can be calculated from quantities that are derived from the objective function at its optimum. Gouriéroux, Monfort and Renault (1993) or Jiang and Turnbull
provide presentations of the large sample theoretical properties of the method and describe a wide array of applications.

In the current paper we focus on showing that I-I can provide a feasible solution to the issue of unbiased estimation. We will not pursue the question of producing the most efficient I-I estimator. We therefore limit ourselves to the case where A is set to the identity matrix. Hence any improvements that are obtained in terms of bias by our implementation of I-I do not preclude further potential efficiency gains that can be obtained by a more adapted (but more demanding) choice of A.

In practice we use the default GMM optimizer in Stata to compute the I-I estimates. The starting values are obtained systematically from the auxiliary statistics of the sample, such that all starting values are endogenously determined based on the data.

Choice of auxiliary statistics

We consider auxiliary statistics from both types of conventional estimators, i.e., cross sectional statistics including bunching behavior, and panel statistics such as the autocorrelation coefficient of taxable income. Furthermore, we focus on statistics which can be calculated directly from the sample data (and therefore from the simulated data). Our choice of auxiliary statistics is guided mainly by the principle of analogy. Since we estimate all the parameters of the model, i.e., the ETI as well as all other parameters which arise from our distributional assumptions (location and scale), the composition of the unobserved component as well as the dynamics of the transitory component, we require auxiliary statistics which will capture separately distinct features of the observed distribution. To achieve this, we will use five sets of auxiliary statistics: means of income levels and growth, the proportions of tax units above and below the kink point, transitions between the two tax brackets, as well as variances and covariances of income levels, income growth, and the
net-of-tax rate. We describe these statistics in detail in Appendix ??.

IV. Data

We use microdata on Swedish incomes from 2001 to 2013. The data combines register data such as taxable income from various sources with survey data that contains socio-economic characteristics. In particular, the data contains labor income on the individual level and on the household level.

V. Results

VI. Conclusion
REFERENCES


