

Measuring Tax Efficiency Under Income Uncertainty: An Empirical Assessment of the Tax Reform Act of 1986

Daniel Grodzicki and Akshaya Jha*

November 13, 2012

Abstract

In this paper, we derive and implement an empirical test for the Pareto efficiency of a nonlinear tax schedule. Building on the theoretical framework of Boadway and Sato (2011) we develop an estimable model of optimal taxation under uncertainty. We derive an inequality determining whether a given tax schedule is Pareto efficient, similar to Werning (2007). Empirically, we use our framework to evaluate the efficiency properties of the tax resulting from Tax Reform Act of 1986. We combine data on individuals' tax returns from the SOI public use files with data on income expectations from the Survey of Consumers in order to structurally estimate this model. Surprisingly, we find that the Pareto efficiency of a tax schedule is more likely to be violated at very low incomes rather than very high incomes.

JEL: H21

Keywords: Efficiency, Optimal Taxation

PRELIMINARY DRAFT PLEASE DO NOT CITE

*Department of Economics, Stanford University, 579 Serra Mall, Stanford, CA 94305, danielg3@stanford.edu and akshayaj@stanford.edu respectively. We would like to thank Tim Bresnahan, Liran Einav, Caroline Hoxby, Florian Scheuer, and Frank Wolak for invaluable advice and mentorship.

1 Introduction

The Tax Reform Act of 1986 (TRA86) signified a major overhaul of the U.S. tax system. Born out of a treasury department report from November 1984 titled "Tax Reform for Fairness, Simplicity, and Economic Growth", TRA86 was designed to drastically simplify an overly complicated existing code, promote fairness by doing away with special deductions, and encourage economic growth and efficiency by reducing tax induced behavioral distortions¹. Since that time, there has been a lot of work studying the effects of TRA86. In particular, there has been some interest in evaluating the efficiency benefits from this reform in terms of behavioral distortions induced by the tax². In this vein, Feldstein (1995) uses TRA86 in order to demonstrate that substantial labor supply effects may be at the heart of behavioral distortion brought on by taxation, and indirectly quantifies the possible revenue implications implied by the reform³. Importantly, the efficiency implications of this reform have been little studied empirically within an optimal taxation framework (Mirrlees, 1971).

The literature on optimal taxation has evolved mostly in a theoretical setting and largely separate from the empirical literature on tax responses often used to study the reform. Only recently has there been some effort to bridge this gap⁴. In this paper, we look to extend to the empirical literature

¹To this end, it drastically reduced the number of individual income tax brackets from 15 in 1986 to 2 in 1990 (and from 7 to 5 corporate income tax brackets). It lowered the top marginal individual rates from 50% to 28% and corporate rates from 45% to 34%, but it raised the bottom individual rate from 11% to 15%. It also raised the level of the standard individual deduction to \$5,000 and the personal exemption to \$2,000 as well as expanded the earned income tax credit (EITC) to \$953. Furthermore, it eliminated a number of standard deductions. For example, it repealed a deduction of 10% of a lower earning spouse's wage income (up to \$30,000), it limited deductions for employee business expenses, travel, and entertainment, and repealed deductions for state and local sales taxes. It also took away individual dividend exclusions and the 60% exclusion for long term capital gains.

²Auerbach (1989) shows that efficiency gains from TRA86 were small, as the gains from reducing non-uniformities in capital taxation were offset by an increase in inter-temporal distortions. Interestingly, Goulder and Thalmann (1993) argue that TRA86 actually worsened inter-temporal efficiency by increasing the effective tax rate on various type of capital income; nevertheless, this negative effect was more than offset by increased efficiency due to improved capital allocation and a reduction in labor market distortions. Similarly, Gravelle and Kotlikoff (1989) estimate an increased efficiency on the order of 0.85 percent of present value consumption.

³The underlying notion is that a tax schedule cannot be efficient, even in a second best sense, if lowering it would increase the amount of revenue it can raise - a reference to the renowned Laffer-curve.

⁴Saez (2001) re-derives optimal tax formulas so that they can be computed using estimated labor supply elasticities. In the paper, he also implements some of these formulas using previously estimated quantities. Moreover, Werning (2007)

on optimal taxation by deriving and implementing a test for the Pareto efficiency of a nonlinear tax schedule for the more general case in which agents face ex-ante income uncertainty and for which the tax schedule varies according to observable individual characteristics. Moreover, this paper contributes to the literature on the efficiency of TRA86 by empirically investigating this reform through the perspective of the theory of optimal taxation. In the spirit of Werning (2007), we evaluate whether the (nonlinear) income tax schedule resulting from this reform can be considered efficient. In other words, this test reveals whether there exists some underlying social welfare function that can rationalize the individual income tax schedule implied by TRA86. Importantly, a test for Pareto efficiency is inherently a one sided test. Namely, we will never evaluate a given tax schedule as having rates that are too low; the test for efficiency can only tell us whether a (local) decrease in rates can increase tax revenue due to the behavioral changes in labor supply choices induced by this decrease.

In this paper, we construct an estimable static model of optimal taxation under income uncertainty based on Boadway and Sato (2011) and develop a methodology to estimate it using data on individual tax returns. Deriving a condition for Pareto efficiency as in Werning (2007), we then test empirically whether this condition holds. In particular, we combine panel data from the Statistics of Income (SOI) public use tax files with data on income expectations from the Survey of Consumers (SC) to recover the empirical distribution of income uncertainty faced by agents. We next use this distribution to recover the underlying distribution of unobserved types (commonly denoted “ability” or “skill”) and to estimate the Pareto condition defined in our model. The model is estimated for tax year 1990, as by that time the reform’s provisions had been firmly implemented and agents had some time to become acquainted with the new tax laws⁵. Furthermore, we consider a static framework in part because it is analytically tractable, but more importantly because it lends itself better to the short term nature of our question. Since we seek to investigate the impact resulting from this particular reform, we trade off the potential benefits of considering the fuller dynamic problem for the added clarity gleaned from the static model. Moreover, a static context allows us to enrich our analysis by adding both income uncertainty and by

derives a testable condition for the Pareto optimality of an optimal tax schedule, setting in better defined and more general terms what types of (nonlinear) schedules can be considered second-best efficient. An implication of Pareto efficiency of a tax schedule as shown by Werning is that it is devoid of local Laffer effects. This is because it is not possible to locally reduce the tax liability without also reducing the revenue generated from the tax.

⁵In a robustness check, we consider how our results vary when estimated on a “pre TRA” 1986 sample.

allowing the tax schedule they face to vary by their observable characteristics⁶.

We incorporate income uncertainty into our framework firstly because it became particularly salient as a result of TRA86. This is because the reform repealed an income averaging provision that been prevalent for the past several decades. Under this provision, individuals were allowed to average income over several years in order to avoid paying excessively high income tax rates in years in which their income was high. Following the reform, this option was no longer available to the vast majority of filers. Secondly, since the 1970s, income uncertainty has become an increasingly important phenomenon⁷. Moreover, despite its simplifications, the tax schedule resulting from TRA86 still treated individuals differently based on observable characteristics. For example, married couples with children still faced a different tax schedule than childless singles, as did those whose earnings were mainly from capital income versus those whose earnings were from wages. Consequently, we allow for the tax schedule to vary based on these (and other) observable characteristics.

Using the tax schedule resulting from the reform, we find that efficiency is more sensitive to changes in the tax at the low end of the income (skill) distribution. In fact, preliminary findings suggest that there might be some inefficiencies due to the tax rates imposed on very low incomes. With our estimated model primitives, we then look deeper into what tax schedules might still be considered efficient. In particular, we consider the set of linear tax schedules for which this condition still holds. Our preliminary results suggest that there is a very wide range of flat tax rates which can be efficient (i.e. rationalized by some social welfare function). Comparing to results in Werning (2007), our model suggests that a much wider range of schedules might be considered efficient when income uncertainty is considered in an individual's decision making.

The rest of the paper is organized as follows. The next section describes the model. Section 3 introduces the data and discusses measurement and identification. Section 4 describes the estimation procedure. In section 5 we present our results, while in section 6 show our counterfactuals. Section 7 concludes.

⁶This is similar to the concept of “tagging” and is discussed further below.

⁷Dynan et al. (2008) calculate that the volatility of household income has increased by one third between 1970 and 2000. Decomposing this volatility into variability and uncertainty, Cunha and Heckman (2007) further show that, for less skilled workers, 60% of this increase in volatility is due to a rise in uncertainty; for more skilled workers only 8% is due greater uncertainty. Gottschalk and Moffitt (2009) further verify these trends by show that the transitory variance of annual earnings has increased over this period.

2 Model

We write a model of optimal income taxation with ex-ante income uncertainty that closely follows Boadway and Sato (2011). In this model, individuals of heterogeneous and unobserved earnings ability (skill) choose the amount of labor they supply under uncertainty over their eventual compensation⁸superscript,⁹. To reflect the reality of the actual tax code, we also allow agents in our model to vary by observable characteristics (e.g. marital status, business ownership status) and face different tax schedules based on these observables (e.g. married couples with children face different taxes than childless singles).

Incorporating heterogeneity based on observables directly impacts our strategy for identifying uncertainty in earnings. In particular, we assume that the distribution of wage shocks is independent of skill (a), *conditional* on agents' observable characteristics (x)¹⁰. Allowing the government to vary taxes based on these observables, we solve the model under this new independence assumption¹¹. Lastly, as we are investigating whether TRA86 implemented efficient tax rates, as determined by this model, we do not derive an optimal tax rate. Our goal is to derive a Pareto efficiency condition as in Werning

⁸Importantly, individual income uncertainty is resolved after the labor supply decision is made. Examples of this might include wage workers who receive a bonus or pay cut at the end of the year, business owners who might receive a windfall from a positive demand shock, or stock owners whose stock unexpectedly increase or decrease in value.

⁹This model differs from the standard problem of Mirrlees (1971) in that agents in our model face uncertainty in realized income, and so the quantity of labor supplied does not map directly into observed income as in the standard Mirrlees model. Therefore, the Revelation Principle argument that the government can choose a tax schedule such that agents truthfully reveal their unobserved types (denoted “ability” or “skill”) does not hold in our case. To see this, note that individuals with the same ability will choose to supply the same amount of labor, but will almost surely earn different realized incomes from this effort.

¹⁰A crucial underlying assumption made in Boadway and Sato (2011) is that the distribution of the wage shock does not depend on skill a . Although this assumption is an important simplification theoretically, it is unlikely to hold empirically. It is more probable that income uncertainty varies with other individual characteristics correlated with ability. For example, a business owner and a wage worker possibly face very different uncertainty over how much money they take home at the end of the year. Moreover, business owners are, on average, more likely of higher ability than wage workers. If business owners and wage workers are considered as one population, then it may no longer be plausible that the income shock is independent of individual ability. However, we may be able to restore independence between the income shock and ability if business owners and wage earners are separated into different populations.

¹¹This is akin to the concept of “tagging”, whereby individuals are taxed differently according to observables (Akerlof, 1978; Boadway and Pestieau, 2006; Cremer et al., 2010)

(2007) that we can estimate using our data¹².

2.1 Basic Setup

Our economy consists of a continuum of workers with observable characteristics $x \in \chi$ and heterogeneous earning abilities $a \in [\underline{a}, \bar{a}]$ that are continuously distributed according to $a|x \sim F(a|x)$ such that $F'(a|x) = f(a|x)$. In this economy, workers initially choose an effective labor supply (z). After choosing z , they receive an exogenous and unexpected income shock (ϵ), whose cumulative distribution function, $G(\epsilon|x)$, differs based on their observable type (x). We make two critical identifying assumptions. The first is that the observables in x are immutable. As will be described later in the empirical section, these observables contain elements that represent long term life choices (e.g. marriage, children, owning a business). Given that our analysis is static, we consider these to be largely unchangeable in the short term and thus effectively immutable¹³. The second is that, conditional on observables x , the distribution of income shocks is independent of individuals' ability type; $\epsilon|x \perp\!\!\!\perp a|x$.

Thus, individuals' realized income from work (y) is determined by labor supply choice (z) as well as by the realization of their income shock (ϵ). Writing $y(z, \epsilon)$, we note that $y(z, \epsilon)$ is assumed to be monotonically increasing in both z and ϵ . Following Boadway and Sato (2011), we invert $y(z, \epsilon)$ to get $\epsilon(y, z)$. Then we can write $G(\epsilon|x)$ equivalently as $G(y, z|x)$. Observe that $G_y(y, z|x) \geq 0$ is the marginal distribution of income, and $G_z(y, z|x) \leq 0$ can be interpreted as the change in number of workers earning less than y as z changes¹⁴.

The government observes only individuals' realized income (y) and characteristics (x). Consequently, it sets a non-linear income tax schedule that varies with realized income and is conditional on observables, $T(y|x)$. Moreover, workers have preferences only over consumption (c) and labor supplied (z), where consumption is defined as income net of tax, ($c(y|x) = y - T(y|x)$). Lastly, it is assumed that the distribution of shocks $G(\epsilon|x)$ faced by workers is common knowledge.

¹²We also intend to empirically investigate in future work the impact of uncertainty on optimal tax rates in terms of its redistribution, insurance, and incentive effects.

¹³This is in contrast to the intuition put forth in dynamic models of optimal taxation such as Werning (2007), where optimal tax schedules are required to be incentive compatible with individuals' choices over the life-cycle.

¹⁴These, as well as other properties of $G(\cdot)$ claimed in this section, follow from our empirical assumption that $y = z\epsilon$.

2.2 The Workers' Problem

Workers of ability type a make an effective labor supply decision $z(a)$ prior to the realization of their income shock. Assuming expected utility is Von Neumann-Morgenstern (VNM) and that bernoulli preferences are separable in income (consumption) and labor for simplicity, the consumer's problem is given by

$$\max_z U(z; a, x) = \int_{\underline{y}}^{\bar{y}} \left[u(c(y|x)) G_y(y, z|x) - \left(\frac{z}{a} \right)^\theta \right] dy \quad (1)$$

s.t.

$$c(y|x) = y - T(y|x)$$

such that workers' value function is then defined by

$$\begin{aligned} v(a|x) &= \arg \max_z \int_{\underline{y}}^{\bar{y}} \left[u(c(y|x)) G_y(y, z|x) - \left(\frac{z}{a} \right)^\theta \right] dy \\ &= \int_{\underline{y}}^{\bar{y}} \left[u(c(y|x)) G_y(y, z(a|x)|x) - \left(\frac{z(a|x)}{a} \right)^\theta \right] dy \end{aligned} \quad (2)$$

where $u' > 0$ and $u'' < 0$. Then taking a first order condition, we get that

$$\frac{\theta}{a} \cdot \left(\frac{z(a|x)}{a} \right)^{\theta-1} = \int_{\underline{y}}^{\bar{y}} u(c(y|x)) G_{yz}(y, z(a|x)|x) dy = - \int_{\underline{y}}^{\bar{y}} u'(c(y|x)) c'(y|x) G_z(y, z(a|x)|x) dy \quad (3)$$

Note that a sufficient condition for the second order condition to be satisfied is that $c'(y|x) = 1 - T'(y|x) \geq 0$ and that $G_{zz}(y, z|x) \geq 0$. Moreover, if the second order condition is satisfied, effective labor supply ($z(a|x)$) is increasing in ability¹⁵.

2.3 The Government's Program

The government maximizes a weighted social welfare function of the form

¹⁵These conditions are verified empirically below.

$$W = \int_{x \in \mathcal{X}} \int_{\underline{a}}^{\bar{a}} \beta(a|x) v(a|x) f(a|x) da dx \quad (4)$$

subject to the following constraints

$$v(a|x) = \int_{\underline{y}}^{\bar{y}} \left[u(c(y|x)) G_y(y, z(a|x)|x) - \left(\frac{z(a|x)}{a} \right)^\theta \right] dy \quad (5)$$

$$R \leq \int_{x \in \mathcal{X}} \int_{\underline{y}}^{\bar{y}} \int_{\underline{a}}^{\bar{a}} [T(y|x)] G_y(y, z(a|x)|x) f(a|x) da dy dx \quad (6)$$

$$\frac{\theta}{a} \cdot \left(\frac{z(a|x)}{a} \right)^{\theta-1} = \int_{\underline{y}}^{\bar{y}} u(c(y|x)) G_{yz}(y, z(a|x)|x) dy \quad (7)$$

The first constraint is simply the definition of consumers' value function. The second constraint is the government's revenue constraint. This expression simply ensures that the overall tax revenue collected is above the exogenous amount of revenue that the government must raise¹⁶. Note that tax revenues are summed over all x 's; there is no requirement that the government earn a certain amount of revenue from a given x . As a simplification, we assume that a law of large numbers applies such that the government does not face any aggregate uncertainty in the overall tax revenue collected. The third constraint is the consumer incentive compatibility (IC) constraint arising from the consumers' first order condition. The Lagrangian function can be written as

¹⁶We assume that the amount of revenue needed is greater than zero, $R > 0$.

$$\begin{aligned}
\max_{v(a|x), z(a|x), c(y)} \mathcal{L} = & \int_x \int_{\underline{a}}^{\bar{a}} \beta(a|x) v(a|x) f(a|x) da dx + \\
& \lambda \left[\int_x \int_y^{\bar{y}} \int_{\underline{a}}^{\bar{a}} [y - c(y|x)] G_y(y, z(a|x)|x) f(a|x) da dy dx - R \right] + \\
& \int_x \int_{\underline{a}}^{\bar{a}} \mu(a|x) \left[\int_y^{\bar{y}} \left[u(c(y|x)) G_y(y, z(a|x)|x) - \left(\frac{z(a|x)}{a} \right)^\theta \right] dy - v(a|x) \right] da dx + \\
& \int_{x \in \chi} \int_{\underline{a}}^{\bar{a}} \gamma(a|x) \left[\frac{\theta}{a} \cdot \left(\frac{z(a|x)}{a} \right)^{\theta-1} - \int_y^{\bar{y}} u(c(y|x)) G_{yz}(y, z(a|x)|x) dy \right] da dx \quad (8)
\end{aligned}$$

The government here solves its program for each attribute group $x \in \chi$ individually. Nevertheless, it is concerned with maximizing social welfare over all groups. Moreover, as mentioned previously, we do not require that a certain amount of revenue be raised from any given group. As is evident by the revenue constraint, it allows for cross-subsidies to exist (e.g. one group pays more and another pays less to make up the overall revenue requirement). Solving point-wise for $v(a|x)$, $z(a|x)$, and $c(y|x)$, respectively, and suppressing arguments for concision, we characterize the solution as

$$\beta(a|x) f(a|x) = \mu(a|x) \quad (9)$$

$$\lambda \int_y^{\bar{y}} [y - c(y|x)] G_{yz}(\cdot|x) f(a|x) dy = \gamma(a|x) \left[\left(\frac{\theta(\theta-1)}{a^2} \right) \left(\frac{z(a|x)}{a} \right)^{\theta-2} - \int_y^{\bar{y}} u(\cdot) G_{yzz}(\cdot|x) dy \right] \quad (10)$$

$$\int_{\underline{a}}^{\bar{a}} \mu(a|x) u'(\cdot) G_y(\cdot|x) da dx = \lambda \int_{\underline{a}}^{\bar{a}} G_y(\cdot|x) f(a|x) da dx - \int_{\underline{a}}^{\bar{a}} \gamma(a|x) u'(\cdot) G_{yz}(\cdot|x) da dx \quad (11)$$

Note that the FOC on $z(\cdot|x)$ uses the envelope condition from the consumer's problem¹⁷.

¹⁷From the consumer's problem, we have that $\int_y^{\bar{y}} u(c(y|x)) G_{yz}(y, z(a|x)|x) dy = \frac{\theta}{a} \cdot \left(\frac{z(a|x)}{a} \right)^{\theta-1}$.

2.4 Solving for a Pareto Efficiency Condition

Using the above characterization of the solution to the government's problem, we derive a condition for Pareto efficiency (Werning, 2007). For a tax schedule $T(y|x)$ to be Pareto efficient it must be the case that, given a schedule, we can find non-negative Pareto weights ($\beta(a|x) \geq 0 \forall a \in [\underline{a}, \bar{a}]$ and $x \in \chi$) such that the above optimality conditions are satisfied and the revenue constraint is exactly satisfied ($\lambda > 0$). Dividing equation (10) by λ , equation (11) by $u'(\cdot)$, and substituting equations (9) and (10) into equation (11), we get that

$$\int_{\underline{a}}^{\bar{a}} \frac{\beta(a|x)f(a|x)}{\lambda} \cdot G_y(\cdot|x) da = \frac{1}{u'(\cdot)} \left[\int_{\underline{a}}^{\bar{a}} G_y(\cdot|x)f(a|x) da \right] - \int_{\underline{a}}^{\bar{a}} \hat{\gamma}(a|x) G_{yz}(\cdot|x) da \quad (12)$$

where

$$\begin{aligned} \hat{\gamma}(a|x) &= \frac{\gamma(a|x)}{\lambda} = \frac{\int_{\underline{y}}^{\bar{y}} [y - c(y|x)] G_{yz}(\cdot|x) f(a|x) dy}{\left(\frac{\theta(\theta-1)}{a^2}\right) \left(\frac{z(a|x)}{a}\right)^{\theta-2} - \int_{\underline{y}}^{\bar{y}} u(\cdot) G_{yzz}(\cdot|x) dy} \\ &= \frac{\int_{\underline{y}}^{\bar{y}} T(y|x) G_{yz}(\cdot|x) f(a|x) dy}{\left(\frac{\theta(\theta-1)}{a^2}\right) \left(\frac{z(a|x)}{a}\right)^{\theta-2} - \int_{\underline{y}}^{\bar{y}} u(\cdot) G_{yzz}(\cdot|x) dy} \end{aligned} \quad (13)$$

As aforementioned, $T(y|x)$ is *PE* $\iff \beta(a|x) \geq 0$ for each $a \in [\underline{a}, \bar{a}]$ and $x \in \chi$. Since λ , $f(\cdot)$ and $G_y(\cdot)$ are non-negative (with λ strictly positive, given that the revenue constraints binds) we can write

$$T(y|x) \text{ is } PE \iff 0 \leq \frac{1}{u'(\cdot)} \left[\int_{\underline{a}}^{\bar{a}} G_y(\cdot|x) f(\cdot|x) da \right] - \int_{\underline{a}}^{\bar{a}} \hat{\gamma}(a|x) G_{yz}(\cdot|x) da \quad (14)$$

where we have assumed throughout that all disposable income is consumed, or that $c(y|x) = y - T(y|x)$. Note that this inequality must hold for every level of income (y) for each set of characteristics (x).

In the following, we estimate separately the objects in Equation (14) and test for Pareto efficiency using the effective tax rates $T(y|x)$ implied by TRA86. Namely, we define each $x \in \chi$, estimate $G(y, z|x)$ and thus recover $f(a|x)$ from the consumers' optimization condition. We can then plug these estimated magnitudes into Equation (14).

3 Data

As aforementioned, an objective of this paper is to empirically assess the efficiency of the tax schedule executed by TRA86 using our model. To this end, we combine two sources of data. The first is from the IRS Statistics of Income (SOI) Division's public use tax files. The SOI data provides information on households' tax returns for a given year. The second is from the University of Michigan Survey Research Center's Survey of Consumers (SC). This survey is intended to provide regular assessments of consumers' attitudes and expectations. More importantly, it contains information on individuals' expectation of future earnings. The matched data thus provides a measure of individuals' characteristics, the taxes they paid, their realized income, and their expected earnings. Throughout, we use adjusted gross income (AGI) as our measure of realized income (y).

3.1 Description and Match

Each year SOI produces a public use file (PUF) of individual income tax returns by sampling and compiling data from the IRS's U.S. individual income tax returns, Form 1040. Between 1979 and 1990 the SOI furnished data on a panel of individual returns. This panel is based on an unstratified random sample drawn using the last four digits of the individual's social security number and is unweighted¹⁸. The data contain essentially all information found in an individual's income tax return, subject to some censoring of sensitive information. For example, age and geography are censored at 65 and state, respectively. Moreover, geography for individuals earning more than \$200,000 annually is not reported¹⁹. As most provisions of TRA86 had not been fully phased in until 1990, and to account for the time needed for individuals to incorporate these changes into their decision making, we use the SOI panel for years 1989-1990. For this period, there were roughly 20,000 observations in each year. As the panel is unbalanced, we drop individuals not present in both years²⁰.

We use data from the SC to obtain information on individuals' expected earnings. The monthly SC is a nationally representative survey of roughly 500 telephone interviews with individuals living in the contiguous United States. The survey uses a rotating panel design, in which roughly 60% of ob-

¹⁸We obtain this data through the National Bureau of Economic Research.

¹⁹Matching to the SC is done using geography. For this purpose, these individuals are probabilistically assigned a geographic location.

²⁰Attrition accounted for less than 5% of the sample in 1989.

servations each month are new respondents and 40% are repeat respondents that had been interviewed six month prior. Each month, new respondents are drawn from a geographically stratified list-assisted sampling frame using random digit dialing (RDD) telephone sampling. To avoid repeat individuals and to obtain a workably sized sample, we use data on expectations from the first six months of 1990. From the SC, we are interested in the following questions:

- 1. Now, thinking about your total income from all sources (including your job), how much did you receive in the previous year?*
- 2. By what percent do you expect your family's income to increase/decrease during the next twelve months?*

We thus match the SOI and the SC on previous year income with observable cells and use the panel structure of the SOI data to recover expected income. Specifically, for this match, we first separate the data into cells by marital status, number of kids, and region²¹. Since high income households in the SOI data ($AGI \geq \$200,000$ in 1989/1990) are not identified by region, we split those high income households into cells by marital status and number of children. Within each cell we perform a nearest neighbor match by the log of household income (Booth, 2003)²². Note that in the SOI data, income is measured as the AGI in 1989, while in the SC income is reported as earnings received in the previous year (1989).

Table 1 shows descriptive statistics for both the SOI and SC data. Note that, despite being slightly richer, the population in the SC on average looks very similar to that in the SOI. In both cases, marital status, number of children, and geographic location match quite well.

3.2 Effective Labor Supply

Our objective is to be able to measure how realized earnings deviate from the expected income targeted via the amount of effective labor supplied. Namely, we assume that earnings expectations as measured

²¹The SOI data are geographically categorized by state, while SC's geographically categorizes data into four geographical categories (Northeast, South, North Central/Midwest, West). Consequently, for the purposes of data match, we aggregate households in the SOI into one of the four SC categories using definitions from the US Census.

²²We exclude about 221 observations ($\sim 1.1\%$ of the sample) for which the log difference for the match was larger than 1.3 in absolute value.

Table 1: Summary Statistics for SOI and SC Samples

	SOI		Survey of Consumers	
	Mean	S.D.	Mean	S.D.
Income (1990)	30,766	69,263	n.a.	n.a.
Income (1989)	28,411	34,690	36,679	33,049
Expected Pct Income Change	n.a.	n.a.	0.058	0.16
Pct. Married	0.559	0.496	0.615	0.480
Number of Kids	0.610	1.006	0.710	1.050
North East	0.203	0.401	0.214	0.410
North Central/Midwest	0.245	0.430	0.275	0.440
South	0.329	0.470	0.327	0.470
West	0.216	0.410	0.179	0.380
Income Over 200k	0.007	0.085	0.005	0.070
Pct Own Business	0.129	0.336	n.a.	n.a.
Pct Own Stock	0.206	0.404	n.a.	n.a.
Pct Claimed as Dependent	0.093	0.290	n.a.	n.a.
N	20,818		2,811	

by the SC are formed as the solution to a labor supply decision, and so they are reflective of agents' effective labor supplied. It then follows that the deviation of realized income from expected earnings, the unanticipated portion of earnings, is our measure of an individual's income shock.

Since we test for Pareto efficiency in the tax year 1990, we use expectations on income changes formed at the beginning of 1990 along with previous year's income to calculate expected earnings. We then compare these expected earnings to realized income to back out an income shock for each individual. More formally, for each individual, we calculate

$$z_{1990} = agi_{1989}^{SOI} \cdot (1 + \mathbb{E}[\% \Delta y_{1989}^{SC}]) \quad (15)$$

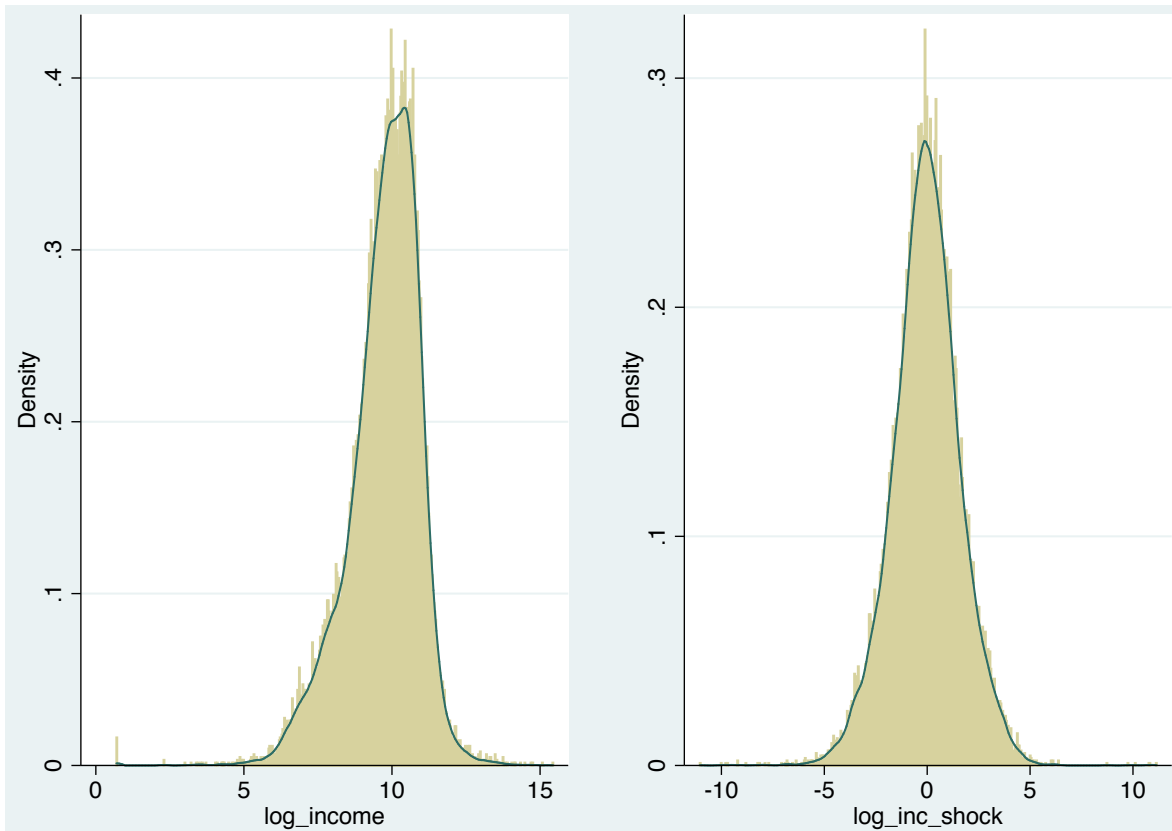
We assume that income shocks are multiplicative, or that $y = z \cdot \varepsilon \iff \varepsilon = \frac{y}{z}$. Then we have that

$$\hat{\epsilon} = \frac{agi_{1990}^{SOI}}{\hat{z}_{1990}} \quad (16)$$

$$\log[\hat{\epsilon}] = \log[agi_{1990}^{SOI}] - \log[\hat{z}_{1990}] \quad (17)$$

Having calculated each agent's effective labor supply (\hat{z}) and income shock ($\hat{\epsilon}$) we can recover the empirical distribution of shocks faced by any set of individuals. Figure 1 plots the unconditional empirical distribution of $\log[y]$ and $\log[\hat{\epsilon}]$ in 1990 on the left and right panels, respectively²³. As expected, the

Figure 1: Unconditional Empirical Distributions of Log AGI and Log Income Shocks



Notes: The figure shows the distribution of the $\ln[1 + agi]$ in 1990 on the left panel and the distribution of the log of the income shock, $\epsilon = \frac{y}{z}$. Data are from the SOI PUF tax panel in 1989 and 1990 and the Survey of Consumers for the first six months of 1990.

log income distribution is skewed to the right, containing a larger mass below the mode. Surprisingly,

²³For conditional distributions of log income and log of the income shock see appendix.

income uncertainty is almost perfectly log normally distributed with zero mean. From this we glean that, on average, individuals are correct when forming their income expectations. Also, we see that, when individuals are wrong, they are equally likely to have guessed too low as they are to have guessed too high.

3.3 Characteristic Cells

One important assumption we make in the model is that income shocks are independent of individual ability, conditional on a set of observables. This is in contrast to Boadway and Sato (2011), which assumes independence without conditioning on observables. We assert that, realistically, the uncertainty faced by different types of individuals is likely correlated with their ability to generate income. In other words, it is likely that individuals with different observable characteristics (e.g. those that are married, those with children, or those that own a business) are subject to different distributions of income uncertainty. We also showed that, because the government sets a tax conditional on observables, its problem should also be solved conditional on these observable characteristics. In this section we define the observables used in the estimation.

We divide our sample into 14 separate categories, or bins, based on our (discrete) observable characteristics. We choose these bins primarily to accord with the differential treatment afforded to individuals in these groups by the tax code. However, we also consider the importance of differential income uncertainty faced by these groups (e.g. married individuals might face less risk than single individuals, and business owners might face more risk). Lastly, due to the relatively small size of our sample, we must make these bins coarse enough so that they are of a workable size. Based on this, we choose bins according to the following characteristics:

1. Claimed as a Dependent
2. Married
3. Children
4. Owns Stock
5. Owns a Business

Dependent status, marital status, and children are directly reported on tax returns. We infer whether a filer is a business or stock owner by looking at whether they claimed some income (either positive or

negative) from that source. Ultimately, we create 14 separate bins. Table 2 outlines the categories more precisely and details what is included in each. Note that the average number of filers per bin is 1,487

Table 2: Description of Characteristic Bins

Bin ID	Filer Can Be Claimed as Dependent	Filer is Married	Filer Claims Children	Filer Owns Stock	Filer Owns a Business	Number of Observations
1	0	0	0	0	0	5460
2	0	0	0	0	1	388
3	0	0	0	1	0	1047
4	0	0	0	1	1	151
5	0	0	1	0	0	200
6	0	1	0	0	0	2332
7	0	1	0	0	1	530
8	0	1	0	1	0	1139
9	0	1	0	1	1	341
10	0	1	1	0	0	5050
11	0	1	1	0	1	885
12	0	1	1	1	0	1006
13	0	1	1	1	1	357
14	1	0	0	0	0	1932

observations. The smallest category, unmarried stock and business owners with no children, contains 151 observations. The largest category, unmarried filers with not children, stock of business, contains 5,460 observations.

3.4 Deriving an Tax Schedule

Having categorized individuals into bins based on observables, we determine the tax rates faced by the individuals in each bin, $T(y|x)$. To this end, we begin by using the Taxsim model (Feenberg and Coutts, 2007) to calculate the tax liability faced by each filer in our sample. Importantly, we allow for certain individuals to pay a negative tax (receive a subsidy), as was the case for some individuals receiving the Earned Income Credit (EIC)²⁴. Importantly, we consider redistribution as acting only through the EIC, as the EIC is the sole source of direct (read: monetary) redistribution built into the tax code.

²⁴The earned income credit was a credit given to individuals earning less than \$20,264 in 1990 and who had children living with them. Because the credit entered individuals' filing as a payment, rather than as a deduction, it acted very much like a pure transfer. As a result, it was possible for individuals to pay negative taxes.

Having calculated individual tax liabilities for our sample, we construct a separate effective tax schedule for each of our bins. We do so by estimating the expected tax for each income level within bins, allowing for non-linearities in the relationship between income and the tax paid. More formally, we estimate the following regression for each individual bin

$$T(y_i) = \delta_0 + \delta_1 \ln[1 + y_i] + \delta_2 \ln[1 + y_i]^2 + \sum_{j=1}^N \delta_{j+2} (\text{Max}[0, \ln[1 + y_i] - k_j])^3$$

where k_j is the k^{th} spline knot used²⁵. We use the estimated coefficients from this regression ($\hat{\delta}_j$) to predict the tax owed by a given filer in a given bin at each level of income, $\hat{T}(y|x)$. For very low incomes, where the data is sparse, we used an algorithm to smooth out the tax schedule and to make it monotonic²⁶. Figure 2 plots an example of a non-linear tax schedule for bin 10.

4 Estimation

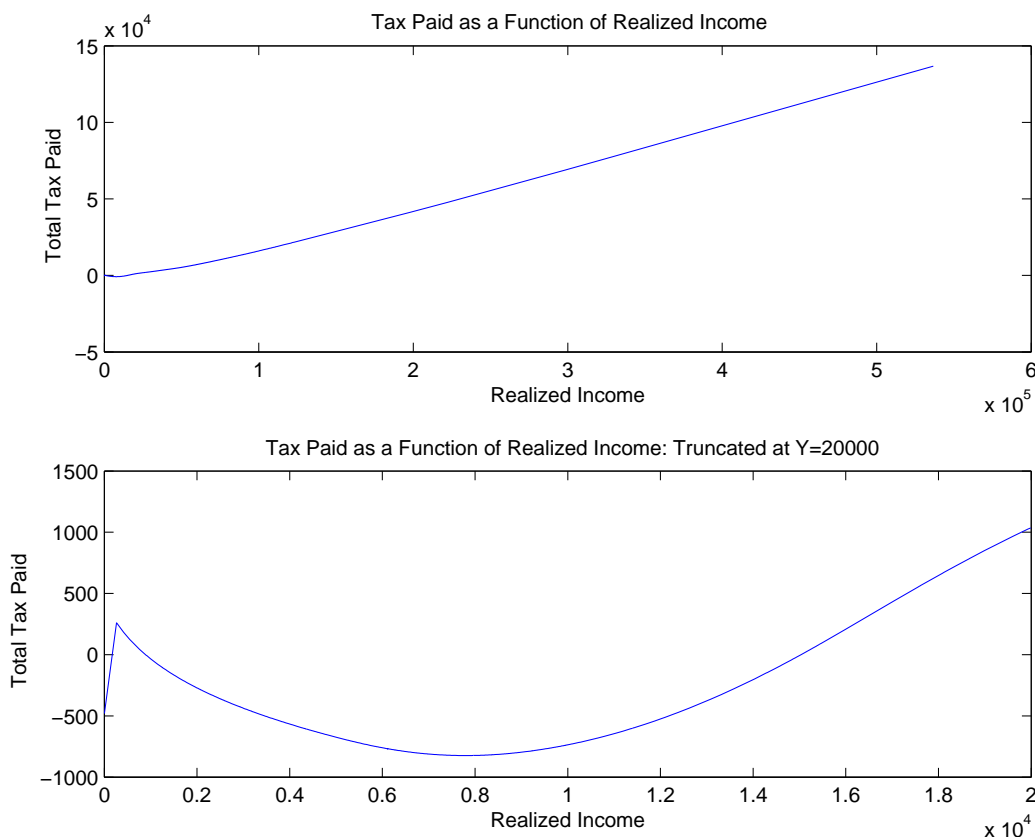
Given that we know each filer's effective labor supply (z_i), realized income (y_i), and thus income shock (ε_i), we can proceed without making any further assumptions regarding the distribution of income uncertainty. In this section, we describe the nonparametric procedures used to estimate the income uncertainty distribution ($G(\varepsilon|x)$) necessary to recover the skill distribution ($f(a|x)$) and estimate our Pareto condition²⁷. We begin by estimating the joint cumulative distribution function of realized income and effective labor $G(\varepsilon|x)$. Next, we plug this into the consumer's first order condition in order to recover the skill distribution $f(a|x)$. Lastly, we plug these into the inequality in the Pareto condition. With this last step, we can test the existing tax schedule for Pareto optimality, as well as recover the full set of efficient tax schedules.

²⁵Knots were chosen to reflect breaks in marginal rates in the tax code and to maximize the fit of the regression. See the Appendix for more information on the knots chosen and for the regression output.

²⁶Our approach is as follows. For bins in which agents did not have children, the lowest tax paid should be \$0.00, according to the written tax schedule. We found the largest income ($y^\#$) that paid no tax ($T = \$0.00$) and set tax by: $T_{y < y^\#} = 0$. For bins in which individuals have children, the EITC meant that the lowest tax paid was \$-953.00. For these bins, we found the income (y^*) which paid the lowest tax (T^*) and set a linear schedule as such: $T_{y < y^*} = \frac{T^* + 953.00}{y^* - 0} \cdot y$.

²⁷All code for this section is written in Matlab, and is available upon request.

Figure 2: Effective Tax Schedule As a Function of Realized Income



Notes: The figure shows our estimated tax schedule as a function of realized income y for covariate bin 10. That is, the vertical axis gives the total amount of tax paid for a given amount of realized income earned. The left panel is plotted over the whole income range observed in the data for this bin. The right panel is plotted from \$1 to \$20,000.

4.1 Identifying Income Uncertainty

As aforementioned, we assume that income shocks (ϵ) are multiplicative ($y = z \cdot \epsilon$). It follows that the shock distribution $G(\epsilon|x)$ can be written as $G(\frac{y}{z}|x)$, where we have data on both y_i and z_i . Moreover, all of our individual-level covariates x are discrete indicators, and so our data are segmented into mutually exhaustive and disjoint bins based on which combination of covariates is equal to 1 rather than 0. As we have a sample distribution of ϵ_i , we could first consider the following simple frequency-based estimator of $G(\epsilon|X)$

$$\hat{G}(\epsilon|x_i = x) = \frac{1}{\sum_{i=1}^N \mathbb{I}(x_i = x)} \sum_{i=1}^N G_{yz}(\epsilon_i \leq \epsilon) \mathbb{I}(x_i = x)$$

where N is the number of individuals in the sample and $\mathbb{I}(\cdot)$ denotes the indicator function. However,

because we need differentiability in ε in order to obtain the distributions necessary for our model (for example, $G_{yz}(y, z|x)$), we smooth this empirical distribution using a Normal kernel. Letting Φ and ϕ be the cumulative distribution function and probability density functions of a standard normal variable, respectively, we have that

$$\hat{G}(\varepsilon|x_i = x) = \frac{1}{\sum_{i=1}^N \mathbb{I}(x_i = x)} \sum_{i=1}^N \Phi\left(\frac{\varepsilon - \varepsilon_i}{h}\right) \mathbb{I}(x_i = x)$$

Note that here h is the smoothing parameter, such that lower values of h correspond to less smoothing (i.e: the empirical distribution is closer to the frequency estimator)²⁸. As is described in the model, we re-write the distributions in the above model in terms of y and z , so that

$$\hat{G}\left(\frac{y}{z}|x_i = x\right) = \frac{1}{\sum_{i=1}^N \mathbb{I}(x_i = x)} \sum_{i=1}^N \Phi\left(\frac{\frac{y}{z} - \frac{y_i}{z_i}}{h}\right) \mathbb{I}(x_i = x) \quad (18)$$

Taking derivatives of this expression, we obtain $\hat{G}_{yz|x}$ and $\hat{G}_{y|z|x}$ as required by the model²⁹.

4.2 Recovering the Underlying Ability Distribution

Once we have $\hat{G}_{yz|X}$, we can easily recover the skill distribution using the consumer's first order condition. Recall that this is

$$\frac{\theta}{a_i} \cdot \left(\frac{z_i}{a_i}\right)^{\theta-1} = \int_{\underline{y}}^{\bar{y}} u(c(y_i)) \hat{G}_{yz}(y_i, z_i|x_i) dy \quad (19)$$

Solving for a_i , we get that

$$a_i = \left(\frac{\theta z_i^{\theta-1}}{\int_{\underline{y}}^{\bar{y}} u(y_i - T(y_i|x_i)) G_{yz}(y_i, z_i|x_i) dy} \right)^{\frac{1}{\theta}}$$

Finally, using this equation, we have an empirical distribution for a_i ³⁰ superscript,³¹, so we can derive a similar nonparametric density

$$\hat{F}(a|x_i = x) = \frac{1}{\sum_{i=1}^N \mathbb{I}(x_i = x)} \sum_{i=1}^N \Phi\left(\frac{a - a_i}{h}\right) \mathbb{I}(x_i = x) \quad (20)$$

It follows that differentiating this function gives us $\hat{f}(a|x)$. Additionally, we now have both z_i and a_i for each individual, allowing us to recover the function $z(a)$ mapping abilities to target income.³²

²⁸Empirically, we set $h = 1.06 \text{std}(\varepsilon_i) \sum_{i=1}^N \mathbb{I}(X_i = x)$, based on ‘‘Silverman’s Rule of Thumb’’.

²⁹For full derivations see appendix.

³⁰The integral in this equation is evaluated using Adaptive Simpson’s quadrature.

³¹Note that we recover the tax schedule as described in the section above.

³²In practice, we utilize Piecewise cubic Hermite interpolation in order to recover this mapping from the data.

4.3 Recovering the set of Pareto Optimal Tax Schedules

Now, having recovered the income shock distribution, as well as the skill distribution, we can plug these estimates into the Pareto condition to test the Pareto efficiency of our estimated federal income tax schedule. Concretely, tax $T(y|x_i)$ is optimal if and only if

$$0 \leq \frac{1}{u'(y - T(y|x_i))} \left[\int_{\underline{a}}^{\bar{a}} \hat{G}_y(y, z_i|x_i) \hat{f}(a_i|x_i) da \right] - \int_{\underline{a}}^{\bar{a}} \hat{\gamma}(a_i|x_i) \hat{G}_{yz}(y, z_i|x_i) da \quad \forall y \in [\underline{y}, \bar{y}], x \in \mathcal{X} \quad (21)$$

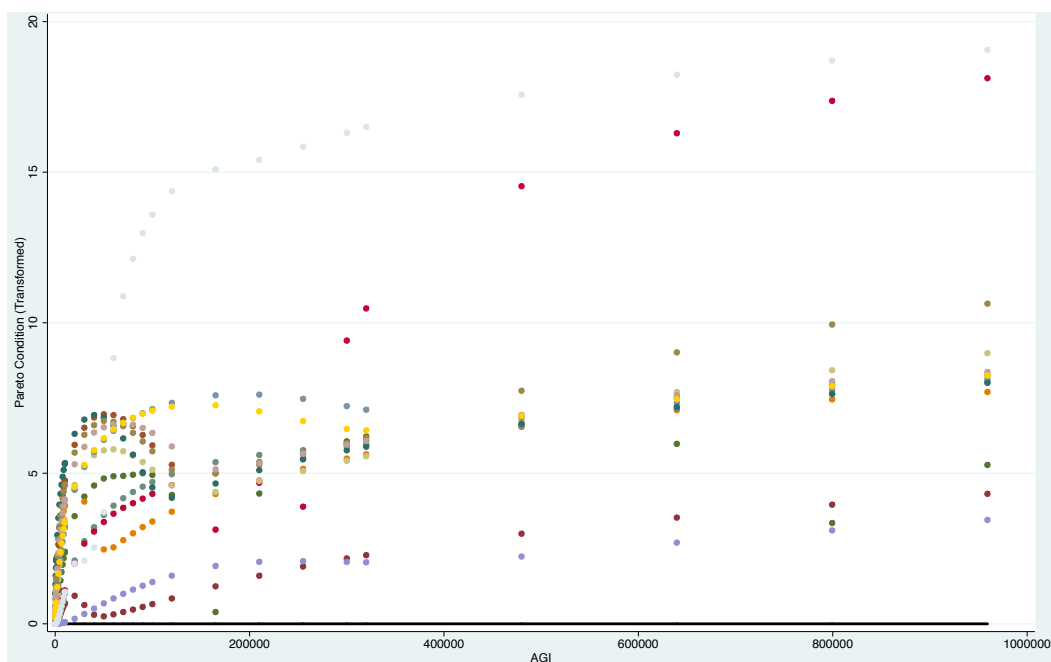
Moreover, we assume further that the skill distribution is a primitive of the model; it is invariant to changes in the tax schedule. Consequently, we can further describe the whole set of Pareto optimal tax schedules using this same equation³³.

5 Results

Our baseline specification uses 1990 data, and the actual tax schedule from 1990, in order to estimate the skill distribution (a_i) for each individual and the Pareto bound inequality for each covariate bin and a grid of realized incomes. For this baseline specification, we let the “risk aversion” parameter $\sigma = 2$ and the dis-utility from effort parameter $\theta = 1$. Our use of $\sigma = 2$ is one commonly used in the business cycle literature (Arellano, 2008). Moreover, in line with Boadway and Sato (2011), we set $\theta = 1$. This implies that agents’ utility is quasilinear in labor and thus that risk aversion enters in only through consumption. A monotonically transformed version of the Pareto bound inequality is plotted in Figure 3. In the figure, each dot represents a calculated Pareto condition for a specific x bin at a specific level of income. The black line represents the bound below which there exists inefficiency. Note that a negative value for this Pareto bound inequality for *any* realized income in our grid means that the tax schedule for that entire bin $T(\bullet|x_i)$ is not Pareto efficient. Concretely, the government can *lower* the tax rate locally and earn *more* revenue than it is currently earning. In all cases, note that the Pareto bound inequality is monotonically increasing in realized income. Contrary to the traditional intuition that Laffer effects are more likely for those earning very high incomes, we instead see potential violations of efficiency at the very *low* end of the income distribution. Table 3 shows numerically the raw calculated bounds for low incomes.

³³As $\int_{\underline{a}}^{\bar{a}} \hat{\gamma}(a_i|x_i) \hat{G}_{yz}(y, z_i|x_i) da$ was sometimes sufficiently nonsmooth to create issues with the Adaptive Simpson’s procedure, we use the simpler Trapezoidal approach with pre-defined grid points for this integral.

Figure 3: Pareto Efficiency Implied by TRA86 in 1990 ($\sigma = 2$, $\theta = 1$)



Notes: The figure shows calculated Pareto bounds for the 1990 income tax schedule. Each dot represents the bound at a particular level of income in a particular bin. The black line represents the minimum efficient bound. These bounds are transformed for the sake of visual clarity. Specifically, raw bounds are first translated upwards such that the lowest bound is normalized to 1. Then they are “compressed” using a monotonic log function. It follows that the Pareto bound (black line) is: $PB = \log(1 + |\text{lowest calculated bound}|)$. Any dots below the line represent points of inefficiency.

Note that a negative value for this Pareto bound inequality for any realized income in our grid means that the tax schedule for that bin $T(\bullet|X_i)$ is not Pareto efficient. Concretely, the government can lower the tax rate locally and increase its revenue. We see that, for the lowest income value (AGI=\$1), the Pareto bound values corresponding to covariate bin 6 is negative. In all cases, note that the Pareto bound inequality is monotonically increasing in realized income. Contrary to the traditional intuition that Laffer effects are more likely for those earning very high incomes, we instead see potential violations of efficiency at the very low end of the income distribution. We posit that this violation may be due to a lack of subsidies for very low incomes, as measured by our model; we only consider the direct subsidy, the EIC. Moreover, the baseline consumption level as defined in our model is most likely to be imprecise for low incomes. That being said, our model and estimation suggest a potential labor supply increase for some bins if the subsidies corresponding to a range of these low incomes were increased. Putting this another way, the marginal decrease in subsidy income from working another hour may be

Table 3: Calculated (Raw) Pareto Bounds at Lower Incomes

Parameters	Bin Number	AGI = 1	AGI = 250	AGI = 500	AGI = 750	AGI = 1000	AGI = 1090
$\sigma = 2$ $\theta = 1$	1	1.4E-08	9.0E-04	3.6E-03	8.1E-03	1.4E-02	1.7E-02
	2	1.1E-07	7.1E-03	2.8E-02	6.4E-02	1.1E-01	1.3E-01
	3	1.0E-06	4.3E-02	1.7E-01	3.9E-01	6.9E-01	8.2E-01
	4	2.3E-08	1.3E-03	5.3E-03	1.2E-02	2.1E-02	2.5E-02
	5	1.8E-02	2.7E-02	3.9E-02	5.3E-02	6.9E-02	7.5E-02
	6	-9.6E-11	3.4E-05	1.3E-04	3.0E-04	5.4E-04	6.4E-04
	7	3.2E-07	2.0E-02	8.0E-02	1.8E-01	3.2E-01	3.8E-01
	8	1.2E-06	7.6E-02	3.0E-01	6.8E-01	1.2E+00	1.4E+00
	9	2.5E-07	1.5E-02	6.1E-02	1.4E-01	2.4E-01	2.9E-01
	10	1.7E+00	2.7E+00	3.9E+00	5.4E+00	7.1E+00	7.7E+00
	11	8.2E-01	1.3E+00	1.9E+00	2.6E+00	3.4E+00	3.7E+00
	12	5.1E-01	8.0E-01	1.2E+00	1.6E+00	2.1E+00	2.3E+00
	13	2.3E-01	3.6E-01	5.3E-01	7.2E-01	9.4E-01	1.0E+00
	14	2.1E-08	1.3E-03	5.4E-03	1.2E-02	2.1E-02	2.5E-02
$\sigma = 2$ $\theta = 2$	1	-6.4E-08	1.3E-02	5.2E-02	1.2E-01	2.1E-01	2.5E-01
	2	3.2E-07	1.1E-02	4.5E-02	1.0E-01	1.8E-01	2.1E-01
	3	3.3E-06	7.1E-02	2.8E-01	6.4E-01	1.1E+00	1.3E+00
	4	1.1E-06	2.3E-03	9.1E-03	2.1E-02	3.6E-02	4.3E-02
	5	5.4E-02	8.3E-02	1.2E-01	1.6E-01	2.1E-01	2.3E-01
	6	-9.7E-07	2.8E-03	1.1E-02	2.5E-02	4.4E-02	5.2E-02
	7	-3.3E-06	2.8E-02	1.1E-01	2.5E-01	4.5E-01	5.3E-01
	8	-1.3E-05	5.4E-02	2.1E-01	4.8E-01	8.6E-01	1.0E+00
	9	-2.5E-06	1.3E-02	5.1E-02	1.1E-01	2.0E-01	2.4E-01
	10	1.4E+00	2.2E+00	3.2E+00	4.4E+00	5.7E+00	6.2E+00
	11	6.3E-01	1.0E+00	1.4E+00	2.0E+00	2.6E+00	2.8E+00
	12	4.6E-01	7.2E-01	1.0E+00	1.4E+00	1.8E+00	2.0E+00
	13	1.9E-01	3.0E-01	4.4E-01	6.0E-01	7.8E-01	8.5E-01
	14	4.6E-07	3.1E-02	1.2E-01	2.7E-01	4.9E-01	5.8E-01
$\sigma = 2$ $\theta = 3$	1	-5.3E-06	3.0E-02	1.2E-01	2.7E-01	4.8E-01	5.7E-01
	2	5.9E-07	1.3E-02	5.0E-02	1.1E-01	2.0E-01	2.4E-01
	3	-3.4E-04	7.9E-02	3.2E-01	7.1E-01	1.3E+00	1.5E+00
	4	3.2E-05	2.7E-03	1.1E-02	2.4E-02	4.2E-02	5.0E-02
	5	7.4E-02	1.1E-01	1.6E-01	2.2E-01	2.9E-01	3.2E-01
	6	-1.1E-05	7.2E-03	2.9E-02	6.5E-02	1.2E-01	1.4E-01
	7	-7.0E-06	3.0E-02	1.2E-01	2.7E-01	4.8E-01	5.7E-01
	8	2.0E-04	4.6E-02	1.8E-01	4.1E-01	7.2E-01	8.6E-01
	9	1.3E-04	1.2E-02	4.6E-02	1.0E-01	1.8E-01	2.2E-01
	10	1.2E+00	1.9E+00	2.8E+00	3.8E+00	5.0E+00	5.5E+00
	11	5.5E-01	8.7E-01	1.3E+00	1.7E+00	2.3E+00	2.5E+00
	12	4.2E-01	6.6E-01	9.5E-01	1.3E+00	1.7E+00	1.8E+00
	13	1.7E-01	2.7E-01	3.9E-01	5.4E-01	7.0E-01	7.7E-01
	14	-2.2E-07	4.9E-02	1.9E-01	4.3E-01	7.7E-01	9.1E-01

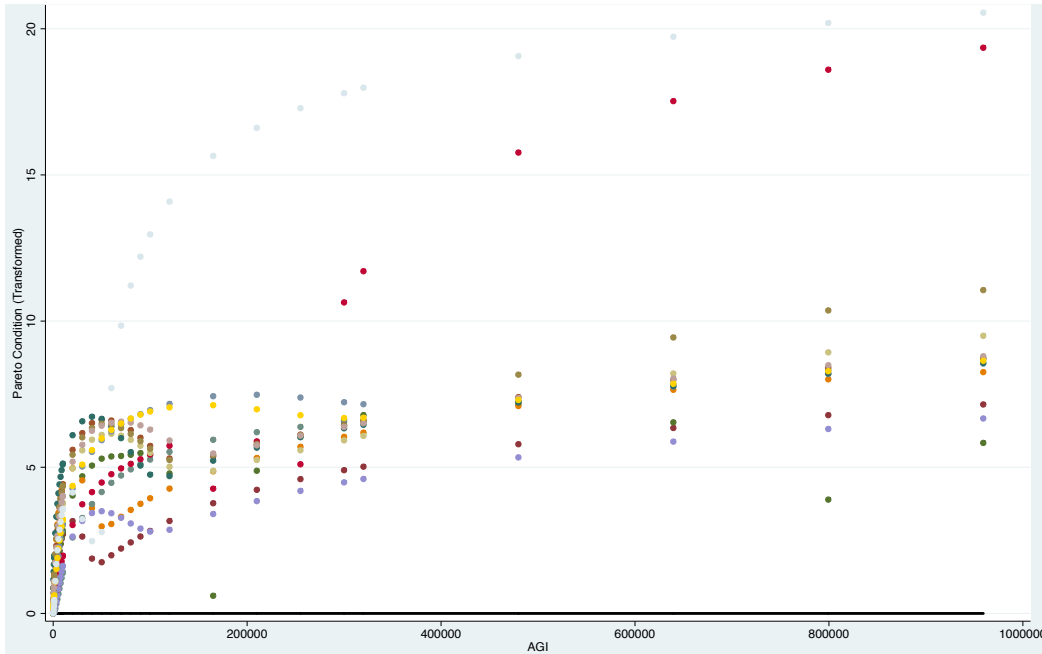
Notes: The table shows the un-transformed calculated Pareto bounds at low incomes for the various scenarios considered. Each cell is a calculated Pareto bound for a specific bin at a specific income level. Negative values (in bold) represent cells which show evidence of inefficiency.

too high for someone with very low income. More importantly, however, we see that at high incomes there is no evidence of inefficiency. This suggest that an increase in the taxes for middle and high income individuals will result in increased revenues.

Also, we consider the sensitivity of our results to the choice of parameters. Namely, we run our model with 1990 data for $\sigma = 2$ and $\theta = 2$ and for $\sigma = 2$ and $\theta = 3$, allowing for non-linearities in the dis-utility from labor. With these cases, we are looking to evaluate the sensitivity of our results to

changes in the labor elasticity of income. The results for the calculated Pareto bound inequality for these two specifications are presented in figures 4 and 5. Comparing the baseline to the $(\sigma = 2, \theta = 2)$

Figure 4: Pareto Efficiency Implied by TRA86 in 1990 ($\sigma = 2, \theta = 2$)

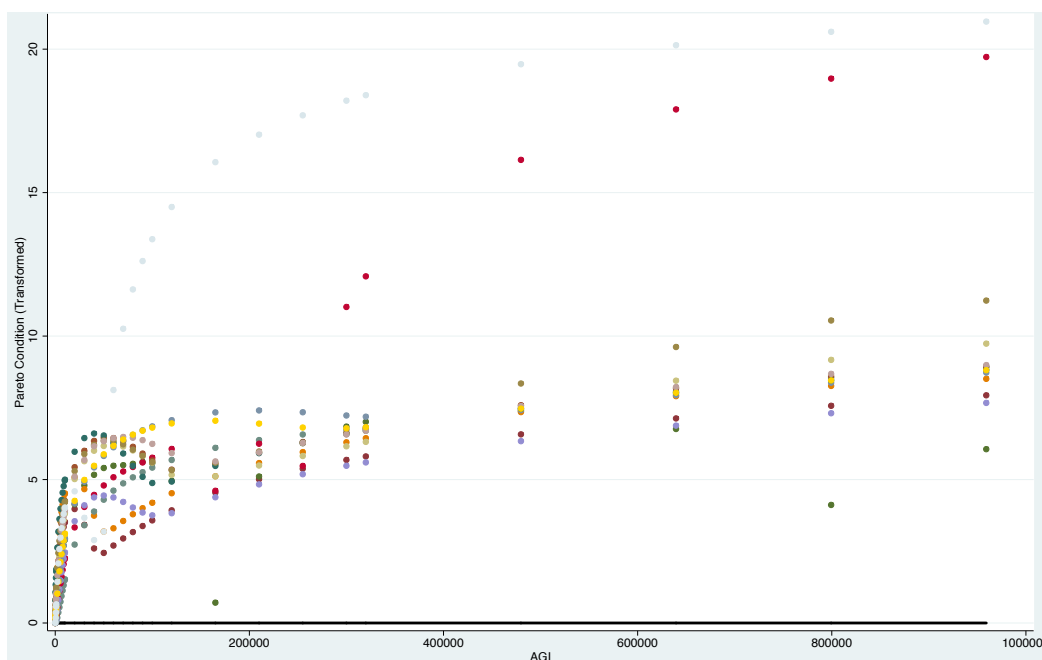


Notes: The figure shows calculated Pareto bounds for the 1990 income tax schedule. Each dot represents the bound at a particular level of income in a particular bin. The black line represents the minimum efficient bound. These bounds are transformed for the sake of visual clarity. Specifically, raw bounds are first translated upwards such that the lowest bound is normalized to 1. Then they are 'compressed' using a monotonic log function. It follows that the Pareto bound (black line) is: $PB = \log(1 + |lowest\ calculated\ bound|)$. Any dots below the line represent points of inefficiency.

specification, the Pareto bound inequality figures look very similar. In all specifications we see that Pareto inequality is monotonically increasing in realized income such that it holds well for middle and high income individuals. Consequently, we see little evidence of the traditional story that individuals with very high incomes may be unduly distorted by high marginal tax rates. On the contrary, we observe some sensitivity in the lowest income grid point (AGI = \$1). Namely, only one bin (bin 6) is not Pareto optimal in the baseline case, while the $(\sigma = 2, \theta = 2)$ specification indicates inefficient tax schedules for five bins.³⁴

³⁴The bins corresponding to inefficient tax schedules are 1, 6, 7, 8, 9, and 10.

Figure 5: Pareto Efficiency Implied by TRA86 in 1990 ($\sigma = 2, \theta = 3$)



Notes: The figure shows calculated Pareto bounds for the 1990 income tax schedule. Each dot represents the bound at a particular level of income in a particular bin. The black line represents the minimum efficient bound. These bounds are transformed for the sake of visual clarity. Specifically, raw bounds are first translated upwards such that the lowest bound is normalized to 1. Then they are 'compressed' using a monotonic log function. It follows that the Pareto bound (black line) is: $PB = \log(1 + |\text{lowest calculated bound}|)$. Any dots below the line represent points of inefficiency.

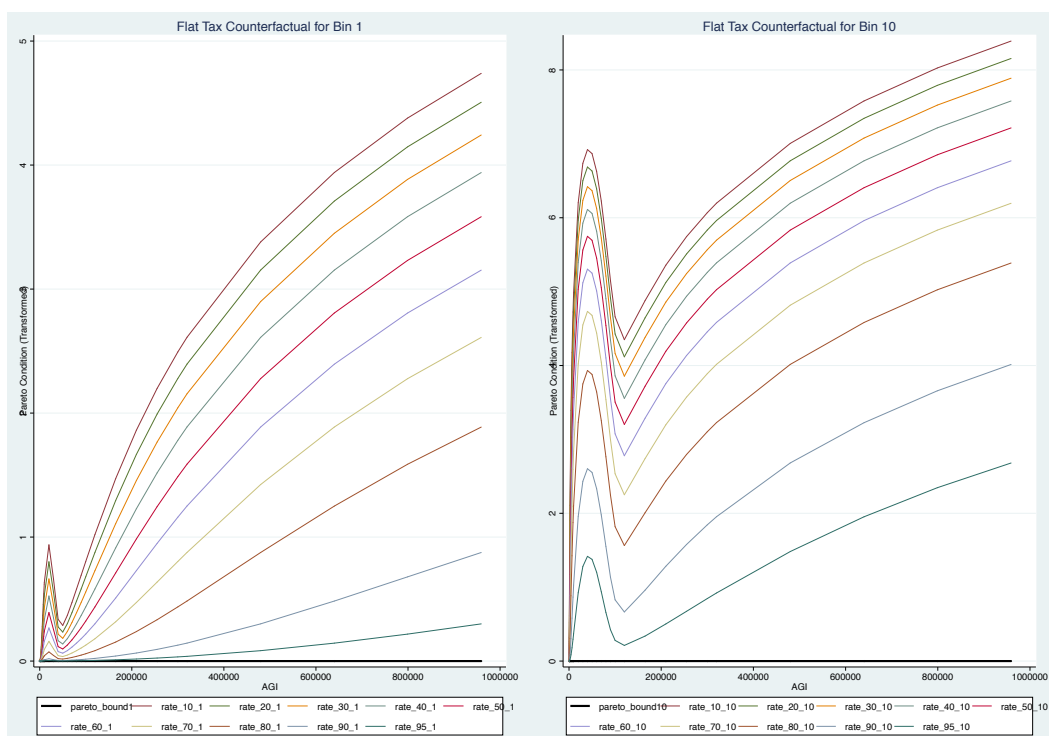
6 Flat Tax Rate Counterfactuals

Finally, as a preliminary counterfactual, we consider a range of linear tax schedules. Our goal is to trace out a version of the Laffer curve; at what marginal tax rates, for a linear tax schedule, do we see the “Laffer effect” that lowering tax rates *increases* the amount of revenue the government raises? Figure 5 shows the flat tax rate counterfactual for bins one and ten³⁵ on the left hand and right hand panels, respectively³⁶. Note that for all these cases, we estimate that the tax is Pareto efficient at middle and upper income levels. However, we do still obtain some negative values indicative of inefficiency in the lowest income level for bin 1. This is in contrast to Werning (2007), who found some evidence

³⁵From Table 2, note that bin 1 refers to all singles with no children who do not own stock or a business. Bin 10 refers to married households with children who do not own stock or a business.

³⁶These two bins are shown in the graph for purely expositional reasons. The results for these two bins are representative of the results for all 14 bins.

Figure 6: Pareto Efficiency for Counterfactual Linear Taxes in 1990 ($\sigma = 2, \theta = 1$)



Notes: The figure shows calculated Pareto bounds for a set of flat tax counterfactuals for bins 1 (right panel) and 10 (left panel). Bin 1 is comprised of agents who are unmarried, have no children, and do not own stock or a business. Bin 10 is comprised of agents who are married with children but that do not own stock of a business. Each line represents a calculated Pareto bound for a particular tax at each level of income. The tax rates considered are 10, 20, 30, 40, 50, 60, 70, 80, 90, and 95 percent, respectively. The black line represents the minimum efficient bound. As in previous figures, these bounds are transformed for the sake of visual clarity. Specifically, raw bounds are first translated upwards such that the lowest bound is normalized to 1. Then they are “compressed” using a monotonic log function. It follows that the Pareto bound (black line) is: $PB = \log(1 + |lowest\ calculated\ bound|)$. Any dots below the line represent points of inefficiency.

toward inefficiencies at rates around 80%. Nevertheless, we do observe that the bound monotonically decreases as we get closer to a 100% rate. Observe that for a 10% rate at the upper income the log of the calculated bound is nearly 5 for bin 1 and more than 8 for bin 10. However, when the tax rate is 95%, the log of the calculated bound drops to around 0.25 for bin 1 and about 2.5 for bin 2. As we use similar data to Werning (2007), as well as a similar functional form and parameter values, we posit that uncertainty works to greatly enlarge the set of tax schedules that might be considered efficient. Nevertheless, we would like to stress that these results are extremely preliminary.

7 Conclusions and Future Work

In this paper, we have derived and implemented an empirical test of the Pareto efficiency of a tax schedule. We apply this methodology to the widely studied Tax Reform Act of 1986, with the goal of studying the efficiency properties of the tax imposed by this Act. We find that, for some sets of individuals, the tax in place is not efficient. Surprisingly, this violation of efficiency arises from high tax rates at the bottom of the income distribution; we posit that subsidies (as we measure them) should be phased out at a slower rate. Our framework allows for the analysis of counterfactual taxes, which we demonstrate via our analysis of a series of linear taxes. We find that, for the case of a flat tax rate, a much broader range of taxes is considered efficient when uncertainty is included in the model.

In future work, our primary goal is to recover the entire set of Pareto efficient tax schedules for 1990. This task is difficult, as evaluating the Pareto bound inequality pointwise for each realized income y may not give us the correct set. Concretely, the evaluation of the Pareto bound inequality for any covariate bin b and realized income y requires specifying the entire tax schedule $T(y|b)$. We are still formulating a method by which we can feasibly recover the full set of Pareto efficient tax schedules.

References

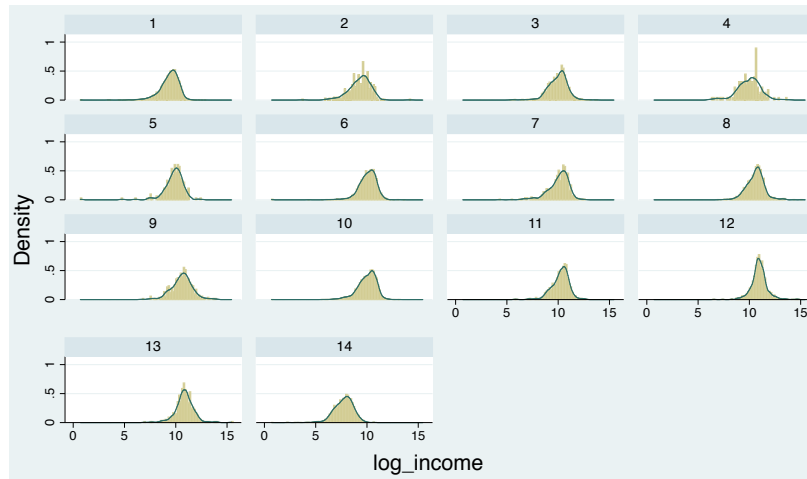
- Akerlof, George A. “The Economics of ”Tagging” as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning.” *The American Economic Review* 68, 1: (1978) pp. 8–19. <http://www.jstor.org/stable/1809683>.
- Arellano, C. “Default risk and income fluctuations in emerging economies.” *The American Economic Review* 98, 3: (2008) 690–712.
- Auerbach, Alan J. “The deadweight loss from non-neutral capital income taxation.” *Journal of Public Economics* 40, 1: (1989) 1 – 36. <http://www.sciencedirect.com/science/article/pii/0047272789900169>.
- Boadway, Robin, and Pierre Pestieau. “Tagging and Redistributive Taxation.” *Annals of Economics and Statistics / Annales d’Économie et de Statistique* , 83/84: (2006) pp. 123–147. <http://www.jstor.org/stable/20079165>.
- Boadway, Robin, and Motohiro Sato. “Optimal Income Taxation with Uncertain Earnings: A Synthesis.” CESifo Working Paper Series 3654, CESifo Group Munich, 2011. http://ideas.repec.org/p/ces/ceswps/_3654.html.
- Booth, Eric. “NEARMRG: Stata module to provide nearest-match merging of datasets.” Statistical Software Components, Boston College Department of Economics, 2003. <http://ideas.repec.org/c/boc/bocode/s434901.html>.
- Cremer, Helmuth, Firouz Gahvari, and Jean-Marie Lozachmeur. “Tagging and Income Taxation: Theory and an Application.” *American Economic Journal: Economic Policy* 2, 1: (2010) 31–50. <http://www.ingentaconnect.com/content/aea/aejep/2010/00000002/00000001/art00002>.
- Cunha, Flavio, and James J. Heckman. “The Evolution of Inequality, Heterogeneity and Uncertainty in Labor Earnings in the U.S. Economy.” Working Paper 13526, National Bureau of Economic Research, 2007. <http://www.nber.org/papers/w13526>.
- Dynan, K.E., D.W. Elmendorf, and D.E. Sichel. “The Evolution of Household Income Volatility.” Technical report, Brookings Institution, 2008.

- Feenberg, D., and E. Coutts. "An introduction to the TAXSIM model." *Journal of Policy Analysis and Management* 12, 1: (2007) 189–194.
- Feldstein, M. "The effect of marginal tax rates on taxable income: a panel study of the 1986 Tax Reform Act." *Journal of Political Economy* 551–572.
- Gottschalk, P., and R. Moffitt. "The rising instability of US earnings." *The Journal of Economic Perspectives* 23, 4: (2009) 3–24.
- Goulder, Lawrence H., and Philippe Thalmann. "Approaches to efficient capital taxation: Leveling the playing field vs. living by the golden rule." *Journal of Public Economics* 50, 2: (1993) 169 – 196.
<http://www.sciencedirect.com/science/article/pii/004727279390048X>.
- Gravelle, Jane G., and Laurence J. Kotlikoff. "Corporate Taxation and the Efficiency Gains of the 1986 Tax Reform Act." Working Paper 3142, National Bureau of Economic Research, 1989.
<http://www.nber.org/papers/w3142>.
- Mirrlees, J. A. "An Exploration in the Theory of Optimum Income Taxation." *The Review of Economic Studies* 38, 2: (1971) pp. 175–208. <http://www.jstor.org/stable/2296779>.
- Saez, Emmanuel. "Using Elasticities to Derive Optimal Income Tax Rates." *Review of Economic Studies* 68, 1: (2001) 205–29.
<http://ideas.repec.org/a/bla/restud/v68y2001i1p205-29.html>.
- Werning, I. "Pareto efficient income taxation." *Massachusetts Institute of Technology, Manuscript* 44: (2007) 45.

A Conditional Plots of Income and Shock Distributions

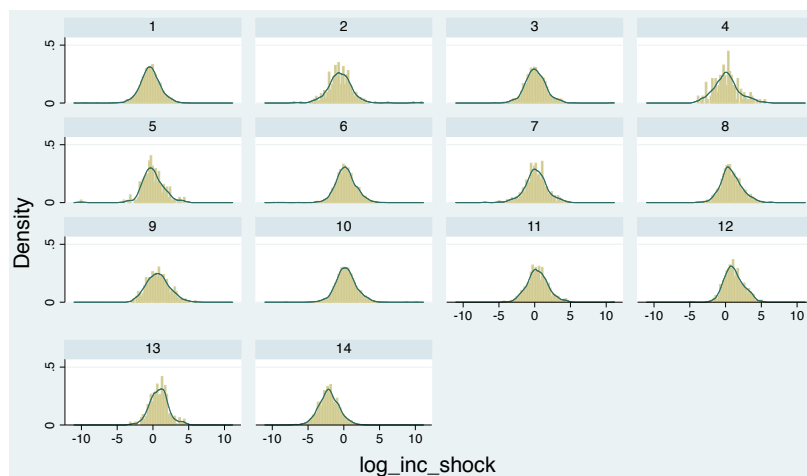
In this section, we plot the conditional income shock distributions $\hat{G}(\epsilon|x)$. Note that there is substantial variation in both the income distributions and the shock distributions for various bins.

Figure 7: Empirical Distributions of Log AGI conditional on x



Notes: The figure shows the distribution of the $\ln[1 + agi]$ in 1990 conditional on observables x . Each plot represents the income distribution for a given bin. Data are from the SOI PUF tax panel.

Figure 8: Empirical Distributions of Log AGI conditional on x

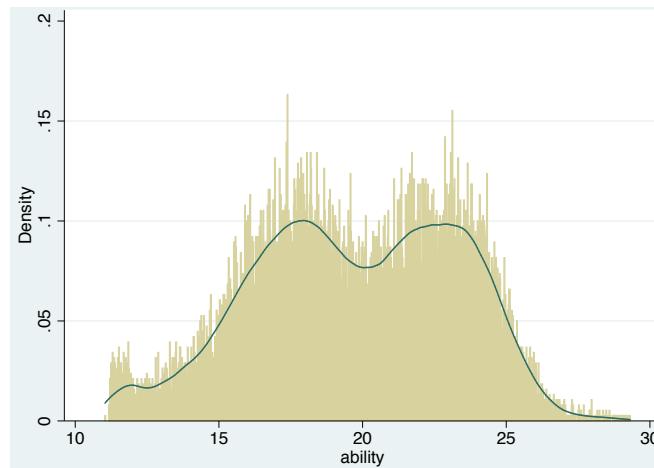


Notes: The figure shows the distribution of the log of the income shock $\epsilon = \frac{y}{z}$ in 1990 conditional on observables x . Each plot represents the distribution of income shocks for a given bin. Data are from the SOI PUF tax panel in 1989 and 1990 and the Survey of Consumers for the first six months of 1990.

B Plots of Estimated Skill Distributions

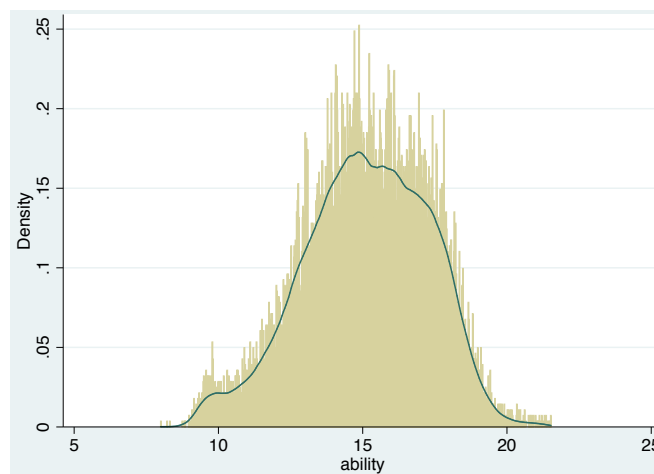
In this section, we plot the estimated skill distributions for the three baseline cases:

Figure 9: Unconditional Skill Distribution $f(a)$ ($\sigma = 2$, $\theta = 1$)



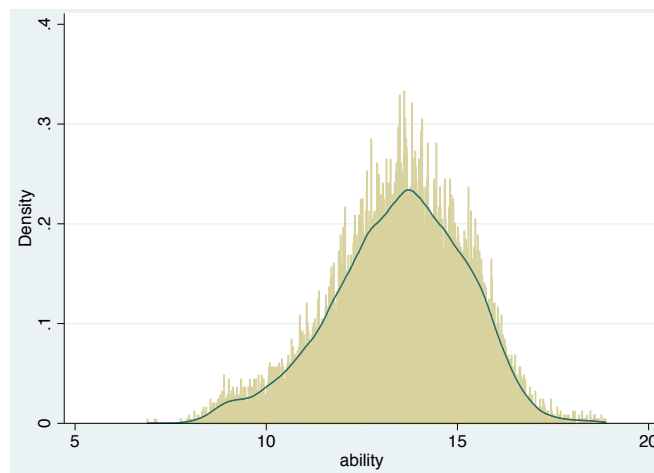
Notes: The figure shows the unconditional recovered skill distribution $f(a)$ for the base case with $\sigma = 2$ and $\theta = 1$

Figure 10: Unconditional Skill Distribution $f(a)$ ($\sigma = 2$, $\theta = 2$)



Notes: The figure shows the unconditional recovered skill distribution $f(a)$ for the base case with $\sigma = 2$ and $\theta = 1$

Figure 11: Unconditional Skill Distribution $f(a)$ ($\sigma = 2$, $\theta = 3$)



Notes: The figure shows the unconditional recovered skill distribution $f(a)$ for the base case with $\sigma = 2$ and $\theta = 1$