

Bounds on Elasticities with Optimization Frictions: An Application to Taxation and Labor Supply*

Raj Chetty

UC Berkeley and NBER

March 2009

Abstract

This paper studies the identification of price elasticities in an environment where agents face optimization frictions such as adjustment costs or inattention. I show that even when one is uncertain about how such frictions enter the agent's decision problem, one can obtain simple bounds on the price elasticity of interest. Agents are permitted to deviate arbitrarily from the optimal choice as long as the utility cost of doing so lies below an exogenously specified threshold. I derive analytical bounds on price elasticities that are a function of the observed response to a price change, the degree of optimization frictions, and the size of the price change. I apply these bounds to the literature on labor supply and taxation, allowing for optimization frictions of up to 1% of consumption. Permitting such small frictions can reconcile several disparate findings in this literature. Pooling elasticity estimates from several studies yields approximate bounds on the taxable income elasticity of (0.47, 0.54).

*E-mail: chetty@econ.berkeley.edu. Thanks to David Card, John Friedman, Patrick Kline, James Hines, Adam Looney, Day Manoli, Emmanuel Saez, and numerous seminar participants for very helpful comments. Gregory Bruich, Tord Krogh, and Zachary Liscow provided outstanding research assistance. Financial support from the National Science Foundation and Sloan Foundation is gratefully acknowledged.

1 Introduction

The standard approach to estimating structural parameters of economic models is to use revealed preference under the assumption that agents follow a known optimizing model. For example, in the literature on labor supply, a common strategy to estimate the wage elasticity of labor supply is to examine the effect of a tax change on earnings. In a labor-leisure choice model where agents optimize perfectly, the difference in labor supply under the two tax regimes identifies the parameter of the utility function that controls the wage elasticity of labor supply. In practice, however, agents are likely to deviate systematically from the optimal choices predicted by a neoclassical model because of various “optimization frictions.” For instance, adjustment costs in switching jobs or inattention to tax reforms could dampen behavioral responses to tax changes. In the presence of such frictions, the labor supply response to a tax change confounds structural preference parameters with the effects of the frictions. A small observed labor supply response to a tax change could be consistent with a large wage elasticity and large adjustment costs or a small wage elasticity and small adjustment costs. Estimating the fundamental wage elasticity rather than just the observed response is important both for welfare analysis and for counterfactual predictions, e.g. of long-run responses where frictions may become less relevant.

How can structural parameters be identified when agents face optimization frictions? One natural approach is to explicitly model the deviations from the standard model and estimate the structural parameters that govern behavior in that more refined model. In the case of adjustment costs, one can try to identify both the preference parameters and the distribution of adjustment costs using additional moments. The limitation of this approach is that it is difficult to model all the factors that affect choices. For example, the effects of tax and transfer policies on behavior differ substantially depending on their salience (see e.g., Duflo et al. 2006, Chetty et al. 2008). Given the lack of a widely accepted theory of salience, it is difficult to credibly estimate preference parameters in a model that incorporates such effects. Even better understood frictions such as adjustment costs can enter models in a variety of ways, creating model uncertainty.

In this paper, I show that even when one is uncertain about how frictions enter the agent’s decision problem, one can obtain informative bounds on the structural parameters of interest.

The key assumption I make is that optimization frictions lead to deviations whose utility cost falls below an exogenously specified threshold δ . For instance, in the case of attention, a plausible restriction is that agents pay attention to “important” features of their budgets – that is, they have misperceptions about their budget sets only to the extent that they do not generate large utility costs. The parameter δ can be interpreted the degree of model uncertainty due to optimization frictions that is permitted. Given a value of δ , the support of the optimization errors that agents make is bounded. This bounded support condition in turn produces bounds on the range of structural parameters consistent with observed behavior.

To formalize this idea, I analyze a simple neoclassical demand model in which heterogeneous agents have isoelastic demand functions with different intercepts but a common price elasticity. The price elasticity is determined by a parameter of the utility function ε . I introduce optimization frictions into this nominal model through an additional error term in the demand equation. The conditional expectation function of this optimization error is unknown. In particular, the optimization error may not be orthogonal to the price. This permits systematic differences between observed responses to a price change and the responses predicted by the nominal model, such as under-reaction to a price increase because of adjustment costs.

I restrict the range of the optimization error by requiring that agents choose points “near” the optimum as predicted under the nominal model. Specifically, I permit any error such that the agent’s choice yields a payoff within δ percent of the maximal payoff attainable under the nominal model.¹ The restriction of the error based on a welfare metric is useful for two reasons. First, there are a wide variety of optimization frictions that may affect behavior in different ways. By restricting the choice set for labor supply based on the utility loss, one can remain agnostic about the particular model through which frictions affect behavior. Second, even for a particular class of deviations, it is difficult to impose a priori restrictions directly on the optimization errors.

I derive bounds on the price elasticity ε that are accurate for small δ in two steps. First, I characterize the set of choices that give the agent a payoff within δ units of the maximum using a quadratic approximation to the objective function. The width of the choice set

¹This should be viewed as a weak requirement: it is unlikely that any economic model predicts (average) behavior perfectly. In practice, there is likely to be additional model uncertainty beyond that due to optimization frictions. I do not consider such additional sources of model uncertainty here, effectively assuming that the econometrician has refined the model to the point that it accurately describes behavior up to optimization frictions.

depends inversely on the curvature of the objective function around the optimum and falls at a square-root rate with δ . The root- δ convergence implies that even very small amounts of model uncertainty can generate substantial uncertainty about agents' choices. Critically, the curvature of the objective function is pinned down by the price elasticity ε . This permits characterization of the choice set without having to identify other parameters beyond the elasticity of interest.

Next, I derive bounds on the set of fundamental price elasticities ε that are consistent with an observed response to a price change. The bounds are derived by calculating the smallest and largest possible shifts in the choice sets that could generate an observed response to a price change. Because of the connection between the curvature of the payoff function and the price elasticity, the bounds on the price elasticity are invariant to parametric modelling assumptions for small δ . The bounds are a function of the observed elasticity, the degree of optimization frictions (δ), and the size of the price change used to identify the treatment effect.

The bounds shed light on the empirical strategies that are most informative in the presence of model uncertainty due to optimization frictions. First, the bounds shrink at a quadratic rate with the size of the price change used to identify the elasticity. As a result, pooling several small price changes – although useful in terms of improving the statistical precision of estimates – yields much less informative bounds than examining a single large price change. Second, the bounds are asymmetric: for typical parameter values, the lower bound is generally much closer to the observed elasticity than the upper bound. If a positive elasticity is observed, the lower bound on the nominal elasticity is strictly positive. Hence, there is considerable value in observing a positive elasticity in order to rule out the null of no response. Third, when a zero elasticity is observed, the upper bound on ε can be conveniently expressed in terms of the utility cost of ignoring the price change. This permits straightforward calculations of the range of elasticities consistent with zero behavioral response in a δ class of models prior to an empirical analysis, analogous to power calculations used to evaluate statistical precision.

I apply these methods to investigate the implications of optimization frictions for the literature on taxation and labor supply. I first show that permitting optimization frictions of less than 1% of consumption can reconcile several stylized facts in this literature. Using simulations of the optimal labor supply choice coupled with the NBER TAXSIM calculator, I

calibrate the utility costs of ignoring each of the tax reforms used in existing studies of labor supply and taxation. The vast majority of the utility costs fall below 1%, indicating that the bounds implied by most existing studies are very wide. In many cases, zero observed responses to a tax change would be consistent with underlying elasticities greater than 1. A few studies of labor supply generate tighter bounds. These include studies of the Tax Reform Act of 1986 (TRA86), where utility costs of ignoring the tax change are close to 10% for high incomes. There are also very large utility costs of ignoring tax reforms on the extensive margin, because these costs are a first-order function of the change in the tax rate. In general, the studies which generate tight bounds on the fundamental elasticity ε are those which find non-zero behavioral responses. Optimization frictions can explain why we observe larger behavioral responses among the rich, larger responses on the extensive margin, limited evidence of bunching at kink points in non-linear budget set models, and differences between microeconomic and macroeconomic elasticity estimates.

I then calculate bounds on the taxable income elasticity using estimates from fifteen recent empirical studies. Even though there is substantial dispersion in the observed elasticity estimates, all fifteen estimates are consistent with one fundamental elasticity ε given 1% optimization frictions. That is, the lower bound of every study falls below the upper bound of every other study with $\delta = 1\%$. Few of the studies yield very informative bounds by themselves; many cannot rule out elasticities well above $\varepsilon = 1$. However, when the studies are pooled, I obtain bounds on the taxable income elasticity of $\varepsilon \in (0.47, 0.52)$. The tightness of the bounds is partly driven by macroeconomic studies and certain studies of high-income earners around the TRA86, whose estimates have been questioned on statistical grounds. Even when one excludes these studies, the bounds remain fairly informative: $\varepsilon \in (0.3, 0.54)$. Although the numbers from this exercise are purely illustrative, they show that informative bounds can be obtained while remaining fairly agnostic about the structure of optimization frictions.

This paper draws upon tools from the partial identification, near rationality, and robust control literatures. The recent literature on partial identification considers problems such as identification in environments with missing data, where point identification is infeasible unless one makes strong assumptions about the missing data (see e.g., Manski 2007, Imbens 2007). In the present paper, uncertainty about the structural model due to optimization frictions creates nuisance parameters that make point identification infeasible. Papers in the partial identifi-

cation literature typically derive bounds by making assumptions such as stochastic dominance of wage distributions for labor force participants relative to non-participants (Blundell et al. 2007). Here, I derive bounds by assuming that agents are “near rational,” as in the menu cost literature in macroeconomics (Akerlof and Yellen 1985, Mankiw 1985). I use results from the menu cost literature on the second-order costs of deviating from an optimum to bound the range of the nuisance parameter. The focus on bounds in a class of models around a pre-specified nominal model parallels the robust control literature (Hansen and Sargent 2007). The robust control literature analyzes optimal policy with a minimax criterion and model uncertainty, whereas the present paper considers identification of the nominal model’s parameters in the same setting. Like the robust control results, the methods proposed here do not provide an excuse for failing to build an accurate model. The bounds are valid only if the nominal model is correct up to optimization frictions. In this sense, the bounds here should only be used as a method of evaluating the robustness of structural parameter estimates after developing an appropriate model of economic behavior.

The remainder of the paper is organized as follows. The next section sets up a simple framework to analyze the effects of optimization errors on treatment effect estimates of elasticities. The results on price elasticity bounds are given in Section 3. Section 4 presents the application to labor supply and taxation. Section 5 concludes.

2 A Simple Framework

Consider a static demand model with N individuals who have heterogeneous tastes over two goods, x and y . Individual i has wealth Z_i and quasilinear utility

$$u_i(x, y) = y + \alpha_i \frac{x^{1-1/\varepsilon}}{1-1/\varepsilon} \tag{1}$$

This utility function has two convenient properties: (1) it is a money metric and (2) it generates heterogeneity in the levels of demand across agents but a common price elasticity of demand. I focus on this particular utility primarily for analytical convenience; a more general utility specification can be permitted by using equivalent-variation measures and numerical computation of bounds following the method described below. Moreover, the standard specification in many literatures – such as the literature on labor supply discussed below – assumes

a constant price elasticity of demand, making this a natural specification to start with.²

The agent solves the following utility maximization problem:

$$\max_x U_i(x) = \alpha_i \frac{x^{1-1/\varepsilon}}{1-1/\varepsilon} - px$$

The optimal demand for good x satisfies the first order condition $U_i'(x^*) = 0$, implying

$$x_i^*(p) = \left(\frac{\alpha_i}{p}\right)^\varepsilon \tag{2}$$

$$\log x_i^*(p) = \alpha_i - \varepsilon \log p \tag{3}$$

Let $\alpha = \sum_i \log x_i^*(p = 1)/N$ denote the mean (log) value of x in the population at a price of 1 and $\nu_i = \log x_i^*(p = 1) - \alpha$ denote the deviation of individual i from the mean. Then we can write the individual i 's demand function as

$$\log x_i^*(p) = \alpha + \varepsilon \log p + \nu_i$$

Our objective is to identify ε , the structural preference parameter that controls the price elasticity of demand absent frictions. We are interested in identifying ε either to make counterfactual predictions – such as how different (unobserved) prices would affect demand – or for welfare analysis, such as the deadweight cost of a tax or regulation.

In this paper, I focus on estimates of the demand elasticity ε using a treatment effect estimator that compares demand under two prices, p_0 and p_1 . Under an orthogonality condition on the error term due to preference heterogeneity ($E\nu_i|p = 0$), the elasticity ε of interest is identified by the difference in mean log demand across the treatments:

$$\varepsilon = \frac{E \log x_1^* - E \log x_0^*}{\log p_1 - \log p_0} \tag{4}$$

Following the robust control literature, I refer to the initial model specified in (2) as the “nominal” model. Equation (4) shows that it is adequate to estimate the observed demand response to a price change to point identify ε . Figure 1 illustrates how measuring the demand response to a price increase from p_0 to p_1 identifies ε . I now explore how this treatment effect estimator is affected by the introduction of optimization frictions.

²Although the objective in (1) reflects a consumption choice problem, it can be adapted to other problems with a simple change of notation. For instance, standard models of the choice of labor supply for an individual or the choice of quantity for a profit-maximizing firm take the same form as (1).

Optimization Frictions: Two Examples. Consider two examples of optimization frictions that may lead agents to deviate from the optimal choice under the nominal model specified in (2). First, suppose each agent pays an adjustment cost k_i to change his level of consumption. Suppose that the initial price is p_0 and that initial consumption $x_i^0 = x_i^*(p_0)$ is set according to (2). Agent i chooses x_i to maximize

$$\alpha_i \frac{x^{1-1/\varepsilon}}{1-1/\varepsilon} - px - k_i \cdot (x \neq x_i^0). \quad (5)$$

Let $\Delta u_i(p) = u_i(x_i^*(p)) - u_i(x_i^0)$ denote the utility gain from reoptimizing demand. Then the resulting demand function $x_i(p)$ follows a threshold rule:

$$\begin{aligned} x_i(p) &= x_i^*(p) \text{ if } \Delta u_i(p) > k_i \\ &= x_i^0 \text{ else} \end{aligned}$$

Let the observed elasticity estimated using the price change from p_0 to p_1 be denoted by

$$\hat{\varepsilon} = \frac{E \log x_1 - E \log x_0}{\log p_1 - \log p_0} \quad (6)$$

In this model, it is straightforward to establish that

$$\hat{\varepsilon} = P(\Delta u_i(p) > k_i) \varepsilon$$

where $P(\Delta u_i(p) > k_i)$ denotes the fraction of agents who adjust their consumption. The observed elasticity $\hat{\varepsilon}$ no longer identifies the “fundamental” elasticity ε because the observed response is attenuated by adjustment costs.

As a second example, suppose agents misperceive the tax-inclusive price of a good p when buying a good x . Evidence indicates that individuals are inattentive to tax rates and confuse average with marginal tax rates (Fujii and Hawley 1988; de Bartolome 1995; Chetty, Looney, and Kroft 2008; Finkelstein 2008; Chetty and Saez 2009). Let $\tilde{p}(p)$ denote the agent’s perceived price as a function of the true price. After choosing x , agents spend the money they have left on the other (numeraire) good. In this case, agent i chooses x to maximize

$$\alpha_i \frac{x^{1-1/\varepsilon}}{1-1/\varepsilon} - \tilde{p}_i(p) \cdot x \quad (7)$$

The resulting demand function $x_i(p)$ is

$$x_i(p) = \left(\frac{\alpha_i}{\tilde{p}(p)} \right)^\varepsilon$$

and the observed elasticity is

$$\widehat{\varepsilon} = \varepsilon \frac{E \log \tilde{p}(p_1) - E \log \tilde{p}(p_0)}{\log p_1 - \log p_0}$$

Again, the observed elasticity $\widehat{\varepsilon}$ confounds the fundamental elasticity of interest ε with other parameters, in this case the effect of the price change on mean perceived prices.

How can one identify the fundamental elasticity ε in the presence of optimization frictions? One strategy is to derive estimators for ε in the more general models described in (5) or (7). For example, by estimating the distribution of adjustment costs, one may be able to infer the fraction of agents who respond to the price change and thereby estimate $\varepsilon = \widehat{\varepsilon}/P(\Delta u_i(p) > k_i)$.³

While model refinement should be the main strategy used to estimate ε , there are two problems with relying solely on this approach. First, it is hard to know precisely how to model and estimate the impact of frictions on $\widehat{\varepsilon}$. Adjustment costs could arise from time costs of job search, the costs of switching consumption plans, or cognitive costs. Moreover, they are likely to vary across occupations, income levels, and institutional settings. In the inattention example, the modelling problem is more complex: lacking a theory of perceptions, it is unclear how one would estimate the $\tilde{p}(p_1)$ function.

A second problem is that it is difficult to model all optimization frictions. The examples above are merely two of numerous frictions that may arise. For instance, non-standard preferences such as reference dependence and biases such as inertia and bounded rationality will generate additional deviations in behavior. Hence, the mapping between observed behavioral responses and structural parameters is likely to be imperfect even in rich structural models that attempt to account for many frictions.

Bounds with Optimization Frictions. Even if the exact structure of optimization frictions is unknown, it is possible to obtain bounds on the parameter of interest ε . Define agent i 's "optimization error" as the difference between his optimal demand as predicted under the nominal model in (2) and his observed demand:

$$\phi_i = x_i(p) - x_i^*(p)$$

³Chetty et al. (2009) implement this approach to estimate the taxable income elasticity in a model with adjustment costs.

Then the observed demand function can be written as

$$\log x_i(p) = \alpha - \varepsilon \log p + \nu_i + \phi_i \quad (8)$$

The ϕ_i term is a nuisance parameter that enters agents' decision problems but cannot be identified by the econometrician. The key difference between the optimization error ϕ_i and the error due to preference heterogeneity ν_i is that we do not know the properties of the distribution of ϕ_i . In the examples above, optimization frictions induce correlations between ϕ_i and p . Hence, we wish to remain agnostic about the conditional expectation function $E\phi_i|p$, allowing agents' deviations from x_i^* to be endogenous to the price. Without placing restrictions on the added ϕ_i error term, the standard treatment effect comparison gives

$$E \log x_i^*|p_1 - E \log x_i^*|p_0 = \varepsilon[\log p_1 - \log p_0] + [E\phi_i|p_1 - E\phi_i|p_0] \quad (9)$$

Without assumptions on ϕ_i , ε is unidentified by the treatment effect because the second term in this expression is unknown. Intuitively, if we place no restrictions on perceptions or adjustment costs, an observed response to a price change can be reconciled with any fundamental price elasticity.

Equation (9) shows that by bounding the support of ϕ_i , one can obtain bounds on ε without making additional assumptions about $E\phi_i|p$. The conventional orthogonality condition on the error term can be dropped in exchange for a bounded support condition if one is willing to settle for set identification instead of point identification. This approach is useful in situations when the central tendency of the nominal model may differ from observed behavior in a manner that changes with p , as is the case with un-modeled optimization frictions.

I restrict the support of ϕ_i by imposing the requirement that agents make choices “near” the optimal choice under the nominal model. That is, restrict ϕ_i to lie within the set

$$\Phi_i(\Delta) = \{\phi_i : U(x_i^*) - U(x_i^* + \phi_i) < \Delta\} \quad (10)$$

Equation (10) permits the agent to deviate from the model in (2) only to the extent that the utility cost of doing so – as calculated under the nominal model – falls below an exogenously specified threshold Δ . This restriction captures the simple intuition that agents may deviate from their unconstrained optimum if the welfare costs of doing so are not too large. The parameter Δ measures the degree of optimization frictions that one permits, and can be

loosely interpreted as a measure of “economic robustness” – the larger the value of Δ one permits when estimating ε , the more robust the estimate is to optimization frictions. I refer to the set of models that generate optimization errors $\phi_i \in \Phi_\Delta$ as a “ Δ class of models” around the nominal model. For example, the adjustment cost model specified above with $k_i \leq \Delta$ for all i lies in the Δ class of models around (2). Similarly, the misperceptions model lies in the Δ class of models around (2) if the price misperception $|\tilde{p}_i(p) - p|$ never generates a utility loss of more than Δ .

A Δ class of models maps a price p to a choice set instead of a singleton. Let

$$X_i(p, \Delta) = \{x_i^*(p) + \phi_i : \phi_i \in \Phi_i(\Delta)\}$$

denote the choice set predicted by a Δ class of models. Figure 2 illustrates the choice set $X_i(p, \Delta)$. In this example, the optimal choice is $x_i^* = 62$ and the set of choices that yield utility within Δ units of the optimum range from $x = 56$ to $x = 68$, as shown by the red interval on the x axis.

To relate the choice sets to a bound on ε , consider an experiment that raises the price from p_0 to p_1 . Figure 3 illustrates the choice sets at the two prices, $X(p_0, \Delta)$ and $X(p_1, \Delta)$. The fundamental elasticity ε controls the movement of the choice sets with the price p , as illustrated by the dashed blue line in the figure. The black lines illustrate that various responses $[x(p_1) - x(p_0)]$ may be observed for a given value of ε , including large reductions, zero response, or even small increases. Which response is observed depends on the nature of the optimization frictions. Large adjustment costs could generate a zero response. If the price increase reflects a change in tax policy that raises tax rates but makes taxes less salient, one might observe an increase in demand if consumers are inattentive. If consumers face adjustment costs and are in an environment where their optimal consumption level is trending over time, they may follow an (S,s) adjustment policy and “overreact,” cutting demand by more than would be predicted by their fundamental elasticity.

These examples illustrate that optimization frictions destroy the one-to-one mapping from the observed response to the fundamental elasticity in (4). Hence, there is a range of fundamental elasticities consistent with a given observed elasticity $\hat{\varepsilon}$ in a Δ class of models. Let this range be denoted by

$$r(\hat{\varepsilon}, \Delta) = (\varepsilon_L(\hat{\varepsilon}, \Delta), \varepsilon_U(\hat{\varepsilon}, \Delta)).$$

The objective of this paper is to characterize $r(\widehat{\varepsilon}, \Delta)$. The range $r(\widehat{\varepsilon}, \Delta)$ measures the uncertainty in the fundamental elasticity due to mis-specification of the behavioral model, much as a statistical confidence interval measures the uncertainty in the parameter estimate due to sampling error. I focus on the range of the estimates of ε rather than other measures of their dispersion (such as the variance) because we are uncertain about the prior distribution over the models within the Δ class. A standard approach in such cases is to adopt a minimax criterion, focusing on worst-case scenarios (Hansen and Sargent 2007).

3 Bounds on Price Elasticities

The characterization of $r(\widehat{\varepsilon}, \Delta)$ can be broken into two steps that are conceptually analogous to standard method-of-moments or maximum likelihood identification procedures. I first characterize $X_i(p, \Delta)$, the choice set at price p for a given value of the fundamental elasticity ε . I then identify the set of fundamental elasticities ε such that the movement in $X_i(p, \Delta)$ with p is consistent with an observed treatment effect $\widehat{\varepsilon}$.

Note that the difference in conditional expectations of the error due to preference heterogeneity ($[E\nu_i|p_1 - E\nu_i|p_0]$) drops out of the treatment effect estimator in (9) under the orthogonality condition $E\nu_i|p = 0$. Hence, preference heterogeneity that leads to shifts in demand for x has no impact on $r(\widehat{\varepsilon}, \Delta)$. I therefore drop the i subscript from this point onward to simplify notation.

Bounds on Choice Set. The choice set can be computed numerically given a value of p and ε by identifying the set of points which yield utility within Δ units of the maximum under the specification in (2), as shown in Figure 2. The following lemma presents a simple analytical characterization of the choice set for small Δ using a quadratic approximation to the utility function $u(x, y)$ in the nominal model. This approximate expression simplifies implementation of the bounds and offers intuition about their key determinants.

Lemma 1. For small Δ , the agent's choice set is approximately

$$X(p, \Delta) = \{x : |x - x^*| < [-2\frac{\Delta}{U_{xx}(x^*)}]^{1/2}\} \quad (11)$$

$$= \{x : \frac{|x - x^*|}{x^*} < [2\varepsilon\frac{\Delta}{px^*}]^{1/2}\} \quad (12)$$

Proof. By definition,

$$X_i(p, \Delta) = \{x_i(p) : U(x_i^*(p)) - U(x_i(p)) < \Delta\}$$

Taking a quadratic approximation to U yields

$$\begin{aligned} U(x^*(p)) - U(x(p)) &= -U_x(x^*)(x - x^*) - \frac{1}{2}U_{xx}(x^*)(x - x^*)^2 \\ &= -\frac{1}{2}U_{xx}(x^*)(x - x^*)^2 \end{aligned} \quad (13)$$

where the second equality follows from the first-order condition under the nominal model, $U'(x^*) = 0$. It follows that

$$X(p, \Delta) = \{x : |x - x^*| < [-2\Delta/U_{xx}(x^*)]^{1/2}\}$$

To pin down the second derivative, use the comparative statics of the first order condition in the nominal model ($u_x(x^*(p)) = p$), which implies:

$$u_{xx}(x^*) \frac{dx^*}{dp} = 1 \quad (14)$$

Note that the second order condition for the optimization problem, $u_{xx}(x^*) < 0$ implies that $\frac{dx^*}{dp} < 0$ and hence $\varepsilon > 0$. Recognizing that $U_{xx}(x) = u_{xx}(x)$ and substituting this equation into (11) yields (12). The approximation error in both of these equations vanishes as $\Delta \rightarrow 0$ because the remainder of the Taylor approximation in (13) involves higher-order terms.

Lemma 1 has two useful implications for the analysis that follows. First, the width of the choice set shrinks as a square-root rate as Δ goes to zero:

$$\mu(X(p, \Delta)) = \max(X(p, \Delta)) - \min(X(p, \Delta)) = 2[-2\Delta \frac{dx^*}{dp}]^{1/2} \propto \Delta^{1/2}$$

Thus, even small amounts of model uncertainty generate a non-negligible choice set:

$$\lim_{\Delta \rightarrow 0} \frac{\mu(X(p, \Delta))}{\Delta} = \infty$$

The root- Δ shrinkage is driven by the second-order losses from moving away from the maximum of a smooth function. As illustrated in Figure 2, the utility function is necessarily flat around interior extrema, and thus a small Δ leads to a relatively wide choice set.

Second, (11) shows that the choice set is inversely related to the curvature of the objective function at the optimum, $U''(x^*)$. Curved utilities generate a narrow interval around the

optimum for a given Δ , because utility falls off sharply as one deviates from the optimum. Measuring the width of the choice set thus requires measurement of $U''(x^*)$. Here, a very useful property of the model is that $U''(x^*)$ is related to ε – the structural parameter of interest. Highly curved utilities generate small elasticities because the agent has a strong preference to locate near x^* . For example, suppose the demand for an essential medicine is perfectly price inelastic at a level x^* . The price elasticity of demand approaches zero as the curvature of the utility function approaches infinity – agents demand the medicine at any price only if they lose infinite utility by not having it. Because the utility costs of deviating from x^* are infinitely large, the choice set $X(p, \Delta)$ collapses to the singleton x^* for any Δ when $\varepsilon = 0$, as illustrated in Figure 4a. More generally, more elastic demand functions imply less curved utility functions and wider choice sets. This is illustrated in Figure 4b, which plots the choice sets when $\varepsilon = 1$. This connection between ε and the curvature of utility is extremely helpful in bounding ε because it eliminates the need to estimate a separate parameter to identify the width of the choice sets.

Bounds on the Elasticity. To derive easily interpretable bounds on ε , it is convenient to introduce a percentage measure of the degree of optimization frictions: $\delta = \frac{\Delta}{px^*(p)}$. The parameter δ measures the degree of optimization errors permitted relative to the total expenditure on x . By fixing δ instead of Δ , we permit larger errors in absolute terms for choices that are larger in magnitude. This proportional scaling simplifies the comparison of choice sets at different prices. I use the term “ δ class of models” to refer to models that lie in the class with $\Delta = \delta px^*(p)$.

Consider again the treatment effect experiment that raises the price from p_0 to p_1 , as shown in Figure 3. The upper bound on the nominal elasticity ε that could have generated an observed treatment effect $\widehat{\varepsilon}$ is that which generates the maximum shift in the choice sets consistent with $\widehat{\varepsilon}$, as shown in Figure 5a. In a δ class of models, the upper bound elasticity ε_U satisfies the condition

$$\begin{aligned} [\widehat{x}(p_1) - \widehat{x}(p_0)] &= \min(X_\Delta(p_1, \varepsilon_U)) - \max(X_\Delta(p_0, \varepsilon_U)) \\ &= x_1^*[1 - (2\delta\varepsilon_U)^{1/2}] - x_0^*[1 + (2\delta\varepsilon_U)^{1/2}] \end{aligned} \tag{15}$$

Similarly, the lower bound elasticity ε_L is that which generates the minimum shift in the choice

sets consistent with $\widehat{\varepsilon}$, as illustrated in Figure 5b:

$$\begin{aligned} [\widehat{x}(p_1) - \widehat{x}(p_0)] &= \max(X_\Delta(p_1, \varepsilon_L)) - \min(X_\Delta(p_0, \varepsilon_L)) \\ &= x_1^*[1 + (2\delta\varepsilon_L)^{1/2}] - x_0^*[1 - (2\delta\varepsilon_L)^{1/2}] \end{aligned} \quad (16)$$

The general approach to bounding elasticities with optimization frictions that is illustrated in Figure 5 does not rely on the parametric model used above. One could in principle calculate the bounds ε_U and ε_L numerically for any given utility specification. The bounds can also be calculated numerically with other types of scaling of the degree of model uncertainty, such as concave relationships between Δ and x^* . In addition, the same logic can be used to derive bounds in multi-parameter nominal models that already incorporate some aspects of adjustment costs, inattention, or other frictions. The logic for deriving the bounds is the same in all of these cases. Given a fully specified model, one first computes the set of choices that yield utility within Δ units of the maximum for a given set of parameters. One can then calculate the boundaries of the set of structural parameters that are consistent with an observed behavioral response as in Figure 5. The assumptions made above facilitate an analytical characterization of the bounds, as shown in the following proposition, which provides a formula for the bounds $(\varepsilon_L(\widehat{\varepsilon}, \delta), \varepsilon_U(\widehat{\varepsilon}, \delta))$ that is accurate for small δ .

Proposition 1. For small δ , the range of fundamental elasticities consistent with an observed elasticity $\widehat{\varepsilon}$ in a δ class of models is approximately

$$r(\widehat{\varepsilon}, \delta) = [\widehat{\varepsilon} + 4\delta(1 - \rho)(\frac{p}{\Delta p} - \widehat{\varepsilon})^2, \widehat{\varepsilon} + 4\delta(1 + \rho)(\frac{p}{\Delta p} - \widehat{\varepsilon})^2] \quad (17)$$

$$\text{where } \rho = (1 + \frac{1}{2}\frac{\widehat{\varepsilon}}{\delta}/(\frac{p}{\Delta p} - \widehat{\varepsilon})^2)^{1/2} \quad (18)$$

Proof. Using the elasticity definitions in (6) and the approximation $\varepsilon \simeq -\frac{(x_1^* - x_0^*)/x_0^*}{(p_1 - p_0)/p_0}$, we can rewrite the definition of ε_U in (15) as

$$\begin{aligned} [1 + (2\delta\varepsilon_U)^{1/2}] + \frac{\Delta x}{x_0^*} &= (1 - \frac{\Delta p}{p}\varepsilon)[1 - (2\delta\varepsilon_U)^{1/2}] \\ \Rightarrow \frac{1 + [2\delta\varepsilon_U]^{1/2} + \Delta x/x_0^*}{1 - [2\delta\varepsilon_U]^{1/2}} &= 1 - \frac{\Delta p}{p}\varepsilon \end{aligned}$$

For small δ , $\frac{1}{1-[2\delta\varepsilon_U]^{1/2}} = 1 + [2\delta\varepsilon_U]^{1/2}$ and hence this equation is approximately

$$\begin{aligned} (1 + [2\delta\varepsilon_U]^{1/2})^2 + (1 + [2\delta\varepsilon_U]^{1/2})\Delta x/x_0^* &= 1 - \frac{\Delta p}{p}\varepsilon_U \\ (1 + [2\delta\varepsilon_U]^{1/2})^2(1 - \frac{\Delta p}{p}\widehat{\varepsilon}) &= 1 - \frac{\Delta p}{p}\varepsilon_U \\ (1 + 2\delta\varepsilon + 2[2\delta\varepsilon]^{1/2})(1 - \frac{\Delta p}{p}\widehat{\varepsilon}) &= 1 - \frac{\Delta p}{p}\varepsilon_U \end{aligned}$$

where the second line follows from the fact that $x(p_0) = x_0^*[1 + (2\delta\varepsilon_U)^{1/2}]$ at the upper bound. For small δ , the $2\delta\varepsilon$ term in the third line vanishes relative to $2[2\delta\varepsilon]^{1/2}$. Hence, this equation is approximately

$$\begin{aligned} (1 + 2[2\delta\varepsilon_U]^{1/2})(1 - \frac{\Delta p}{p}\widehat{\varepsilon}) &= 1 - \frac{\Delta p}{p}\varepsilon_U \\ \Rightarrow \varepsilon_U^2 - (2\widehat{\varepsilon} + 8\delta\varepsilon_U(1 - \frac{\Delta p}{p}\widehat{\varepsilon})^2 / (\frac{\Delta p}{p})^2 + \widehat{\varepsilon}^2 &= 0 \end{aligned}$$

A parallel derivation for ε_L using the definition in (16) yields the same quadratic equation. Solving this quadratic equation and taking its upper and lower roots yields the values of ε_L and ε_U in (17). Using the remainder terms in the approximation, it is straightforward to establish that the difference between the exact bounds and the approximate bounds in (17) vanishes as $\delta \rightarrow 0$.

Equation (17) maps the price change used for identification ($\frac{\Delta p}{p}$), the elasticity estimate $\widehat{\varepsilon}$, and the degree of model uncertainty (δ) to bounds on the fundamental elasticity. Figure 6 plots the bounds $(\varepsilon_L, \varepsilon_U)$ vs. $\widehat{\varepsilon}$ for four combinations of δ and $\frac{\Delta p}{p}$. The top two panels consider model uncertainty of $\delta = 1\%$, while the lower two panels consider $\delta = 0.1\%$. The left panels have a price change of $\frac{\Delta p}{p} = 40\%$, while the right panels have $\frac{\Delta p}{p} = 20\%$. With a price change of 20%, the bounds are very wide. For instance, an observed elasticity of $\widehat{\varepsilon} = 0.2$ is consistent with fundamental elasticities between $\varepsilon_L = 0.01$ and $\varepsilon_U = 2.3$. The bounds become considerably tighter, particularly for low values of $\widehat{\varepsilon}$, when the price change used to estimate $\widehat{\varepsilon}$ is larger. With $\frac{\Delta p}{p} = 40\%$ and $\widehat{\varepsilon} = 0.2$, $\varepsilon_L = 0.05$ and $\varepsilon_U = 0.85$. The reason is that the movement in the choice sets for a given value of ε is larger when $\frac{\Delta p}{p}$ is larger, and thus there are a narrower set of observed responses $\widehat{\varepsilon}$ consistent with any given ε . Large price variation is essential to obtain informative bounds when one is concerned about optimization frictions. This is because optimization frictions can lead to small responses to small price changes even if the fundamental elasticity is large.

Comparing the top and bottom panels of Figure 6, we see that the bounds also becomes more informative as δ is reduced, as one would expect. This confirms the intuition that refining the model so that there are fewer frictions omitted from the nominal model yields more precise estimates. This is why the bounds approach proposed here is a complement to rather than a substitute for the primary approach of refining economic models.

Although the numerical calculations in Figure 6 are based on a formula that was derived from the particular utility specification in (2), these numbers are actually valid more generally. For a general utility $U(x)$, the elasticity of demand $\varepsilon(x)$ varies with the level of x . I show in Appendix 1 that (17) approximately bounds a weighted average of the elasticities $\varepsilon(x)$ between x_0 and x_1 irrespective of the functional form of $U(x)$. Again, the approximation error vanishes as $\delta \rightarrow 0$. The source of this invariance result is that the curvature of utility $U''(x)$ is pinned down by the elasticity $\varepsilon(x)$ for any utility according to (14). Coupled with the fact that we only made use of a quadratic approximation to $U(x)$ rather than its specific form to derive the results in Lemma 1 and Proposition 1, one can see why (17) would be approximately valid for arbitrary $U(x)$. In this sense, (17) and the calculations in Figure 6 can be interpreted as general elasticity bounds in frictionless neoclassical models rather than as a numerical example for a specific parameterization of utility. Note, however, that these bounds ignore statistical uncertainty and thus cannot be directly applied to finite sample estimates of $\hat{\varepsilon}$.⁴

Properties of the Bounds. The bounds offer some insight into what can be learned from treatment effect estimators about fundamental elasticities. First, the bounds are asymmetric around the observed elasticity: $\varepsilon_U - \hat{\varepsilon} > \hat{\varepsilon} - \varepsilon_L$. This asymmetry arises from the proportional relationship between the width of the choice sets and ε , as shown in Lemma 1. Small fundamental elasticities are inconsistent with large observed values of $\hat{\varepsilon}$, making the lower bound relatively tight. In contrast, high fundamental elasticities generate very wide choice sets – because they imply a relatively flat utility function around the optimum – and thus are consistent with many values of $\hat{\varepsilon}$, making ε_U large.

Second, Figure 6 shows that the lower bound $\varepsilon_L > 0$ whenever $\hat{\varepsilon} > 0$. If $\varepsilon = 0$, the choice sets collapse to a single point $x^*(p_0) = x^*(p_1)$ as shown in Lemma 1, and one will therefore never observe positive values of $\hat{\varepsilon}$. Agents who are intent on maintaining a fixed value of

⁴One can presumably develop a confidence interval for $r(\hat{\varepsilon}, \delta)$ using methods similar to those proposed by Imbens and Manski (2004).

x must face very large costs of deviating from the optimum and therefore will never do so. Hence, any study that detects a positive treatment effect is informative about the elasticity ε irrespective of the degree of optimization frictions. The following corollary of Proposition 1 establishes this result formally.⁵

Corollary 1. If $\hat{\varepsilon} > 0$, the hypothesis that $\varepsilon = 0$ is rejected for any δ : $\hat{\varepsilon} > 0 \Rightarrow \varepsilon_L(\hat{\varepsilon}, \delta) > 0$.

Proof. Follows directly from the expression for ε_L in (17), where the second term can be shown to be strictly less than $\hat{\varepsilon}$ for $\hat{\varepsilon} > 0$.

Third, it is interesting to consider what can be learned from the converse case of a study that detects zero observed behavioral response ($\hat{\varepsilon} = 0$).⁶ When $\hat{\varepsilon} = 0$, the bounds take a particularly simple form. The lower bound is $\varepsilon_L = 0$. The equation for the upper bound can be expressed in terms of the utility cost of ignoring the price change. To calculate this utility cost, suppose the agent is initially at the nominal optimum $x^*(p_0)$ and ignores the price increase from p_0 to p_1 . His utility loss from failing to reoptimize is

$$\Delta U \equiv U(x_1^*) - U(x_0^*)$$

Using a quadratic approximation analogous to that in Lemma 1 yields

$$\begin{aligned} \Delta U &= -\frac{1}{2}u_{xx}(x_1^*)(x_1^* - x_0^*)^2 \\ &= -\frac{1}{2}\frac{1}{dx^*/dp}\left(\frac{dx^*}{dp}\Delta p\right)^2 \\ &= \frac{1}{2}px_0^*\varepsilon\left(\frac{\Delta p}{p}\right)^2 \end{aligned}$$

The utility loss from failing to reoptimize as a percentage of the original expenditure level is

$$\Delta u_{\%}(\varepsilon) = \frac{\Delta U}{px_0^*} = \frac{1}{2}\varepsilon\left(\frac{\Delta p}{p}\right)^2 \quad (19)$$

We can relate this result to the equation in Proposition 1 to obtain the following representation for $\varepsilon_U(\hat{\varepsilon} = 0, \delta)$, the upper bound on the elasticity that is consistent with zero behavioral response.

⁵See Honore and Tamer (2006) for another example of bounds varying with the true parameter. They show that in panel data models where one is agnostic about the distribution of initial conditions, the bounds on the estimated degree of persistence collapse to zero when the true degree of persistence is zero.

⁶Among the feasible responses in a δ class of models, a zero response is perhaps the most likely outcome, as it requires no adjustments or attention.

Corollary 2. An elasticity ε is inconsistent with $\widehat{\varepsilon} = 0$ if the utility cost of ignoring the price change under that elasticity exceeds 4δ . That is, $\varepsilon_U(\widehat{\varepsilon} = 0, \delta)$ satisfies

$$\begin{aligned}\Delta u_{\%}(\varepsilon_U) &= 4\delta \\ \Rightarrow \varepsilon_U &= 8\delta / \left(\frac{\Delta p}{p}\right)^2\end{aligned}\tag{20}$$

Proof. When $\widehat{\varepsilon} = 0$, (17) implies $\varepsilon_U = 8\delta / \left(\frac{\Delta p}{p}\right)^2$. Combining this expression with (1) and solving yields the result.

Corollary 2 provides a simple method of determining the range of elasticities for which one can be sure to detect a behavioral response, analogous to a statistical power calculation. The utility cost of ignoring the tax change given an elasticity of ε must exceed 4δ in order for ε to be inconsistent with an observed response of $\widehat{\varepsilon} = 0$. The intuition for the 4δ condition is illustrated in Figure 7. Let $d = x^*(p_0) - \min(X(p_0, \delta))$ denote the difference between the optimal demand and the lowest demand in the initial choice set. Panel A of Figure 7 shows that at the upper bound ε_U , the difference between the optimal demands at the two prices is $x^*(p_0) - x^*(p_1) = 2d$. By definition, the percentage utility cost of choosing $\min(X(p_0, \delta))$ instead of $x^*(p_0)$ at a price of p_0 is δ . Since the utility cost of deviating by d units is δ , the utility cost of deviating by $2d$ units is 4δ , as illustrated in Panel B of Figure 7. The 4δ condition is obtained because the costs of deviating from the optimum rise at a quadratic rate. Although the 4δ threshold relies on the quadratic approximation to utility, the result in Corollary 2 points to a more general lesson. A simple way to gauge whether an experiment is informative about ε is to compute the utility costs of ignoring the price change for an agent who is initially at the optimum. Under any utility specification, a minimal requirement to detect a non-zero response given an elasticity of ε is that the utility cost of ignoring the price change exceed the optimization friction threshold δ .

Equation (20) shows that the upper bound ε_U shrinks with the square of the price change when $\widehat{\varepsilon} = 0$. Accordingly, Figure 6 shows that at an observed elasticity of $\widehat{\varepsilon} = 0$, one can only rule out elasticities $\varepsilon > 2$ when $\frac{\Delta p}{p} = 20\%$, whereas one can rule out $\varepsilon > 0.5$ with $\frac{\Delta p}{p} = 40\%$. Intuitively, the utility costs of failing to react to a price change rise with the square of the price change because of the second-order costs of deviating from interior optima. As a result, large price changes are more likely to induce a reaction for any given elasticity ε . Conversely, the set of ε consistent with zero response falls rapidly as the size of the price change rises.

Corollary 2 also shows that ε_U falls at a linear rate with δ . Although refining the model is always valuable, examining large price changes should take priority over seeking environments with fewer frictions if one wishes to obtain a tight upper bound on ε . More informative bounds are obtained when one studies a price change that is twice as large even if the environment has twice as much model uncertainty due to optimization frictions, as seen by comparing the bounds in Panels B and C of Figure 6.

4 Application: Labor Supply and Taxation

The elasticity of labor supply (or taxable income) with respect to the net-of-tax rate is a parameter of central interest for tax policy analysis and macroeconomic models. A large empirical literature in labor economics, macroeconomics, and public finance estimates this elasticity using historical variation in tax rates in the United States and other developed countries. This section evaluates the bounds on the taxable income elasticity implied by these studies when one permits optimization frictions.

The analysis has two parts. First, I calibrate the utility costs of ignoring the various tax changes in the U.S. used for identification in the existing literature. Using Corollary 2, I evaluate the range of elasticities that could be rejected by each study if a zero behavioral response were observed. These calibrations indicate that a broad range of seemingly disparate findings in the literature can be reconciled with optimization frictions of $\delta = 1\%$, even if the elasticity ε is restricted to be constant across all taxpayers. Second, I calculate the bounds implied by several studies using Proposition 1 and pool them obtain informative bounds on the taxable income elasticity. I then discuss the types of studies that would be most useful to make the bounds tighter.

4.1 Nominal Labor Supply Model

I begin by adapting the analysis above to a labor-leisure choice problem to derive bounds on the taxable income elasticity. Consider an economy of agents choosing consumption (c) and labor supply (l) who have quasilinear utility functions of the following form:

$$u_i(c, l) = c - \alpha_i \frac{l^{1+1/\varepsilon}}{1 + 1/\varepsilon} \tag{21}$$

Agent i 's budget constraint is $c = (1 - t)w_i l$, where t denotes the income tax rate. Following Feldstein (1995), I refer to $w_i l$ as agent i 's "taxable income." The optimal labor supply is $l^*(t) = \left(\frac{(1-\tau)w_i}{\alpha_i}\right)^\varepsilon$, or equivalently,

$$\log l^*(t) = \alpha_i + \varepsilon \log(1 - t)w. \quad (22)$$

In the nominal model, the structural preference parameter ε is point identified by the labor supply (or taxable income) response to a tax change from t_0 to t_1 :

$$\varepsilon = \frac{\log wl^*(t_1) - \log wl^*(t_0)}{\log(1 - t_1) - \log(1 - t_0)}$$

That is, ε equals the elasticity of taxable income with respect to the net of tax rate.

Existing studies have documented a large set of frictions that may make actual responses to tax changes differ from the pure effect of the fundamental elasticity, including costs of switching jobs (e.g. Altonji and Paxson 1992), costs of switching consumption plans (Del Boca and Lusardi 2003), inertia (Jones 2008), and inattention (Chetty and Saez 2009). However, virtually none of the existing studies that estimate ε in the taxable income literature explicitly account for frictions. It is therefore natural to evaluate the bounds on the elasticities implied by these studies using the methods proposed above.

Applying Lemma 1, the labor supply choice set in a Δ class of models around this nominal model is approximately

$$L(t, \Delta) = \left\{ l : \frac{l - l^*(t)}{l^*(t)} < \left[2\varepsilon \frac{\Delta}{c^*(t)} \right]^{1/2} \right\}$$

Let $\delta = \Delta/c^*(t)$ denote the utility loss that is permitted as a percentage of the optimal level of consumption. The labor supply choice set in a δ class of models is

$$L(t, \delta) = \left\{ l : \frac{l - l^*(t)}{l^*(t)} < [2\varepsilon\delta]^{1/2} \right\}$$

With $\varepsilon_{l^*,w} = 0.5$, the width of this choice is $2\delta^{1/2}$. If we consider a class of models that generate choices within 1% of utility-maximizing level ($\delta = 1\%$), the width of the choice set is 20%. The set of labor supply choices that generate utility within 1% of the maximum extends +/-10% around optimum predicted under nominal model, illustrating the substantial uncertainty about behavior generated by small optimization frictions.

To characterize the bounds on the labor supply elasticity, let the observed response to the tax change be denoted by

$$\widehat{\varepsilon} = \frac{\log wl(t_1) - \log wl(t_0)}{\log(1 - t_1) - \log(1 - t_0)}.$$

Proposition 1 can be directly applied to obtain bounds on the taxable income elasticity ε , replacing $\frac{\Delta p}{p} = \frac{\Delta t}{1-t_0}$, which is the relevant measure of the price change in this application. Corollary 2 can be used to determine the range of elasticities inconsistent with zero behavioral response by computing the utility costs of ignoring tax changes. With a linear tax system, one can approximate the utility cost of ignoring a tax change as

$$\Delta u_{\%}(\varepsilon) = \frac{1}{2} \left(\frac{\Delta t}{1-t} \right)^2 \varepsilon. \quad (23)$$

However, in a non-linear tax system, the utility costs of ignoring changes in the tax structure cannot be computed analytically. I therefore numerically calculate $\Delta u_{\%}(\varepsilon)$ and evaluate the range of elasticities for which (20) holds for existing studies in the literature.

4.2 Calibration Methodology

I calculate the utility costs of ignoring changes in the tax code using the utility specification in (21) with a fundamental elasticity of $\varepsilon = 1$. Since $\varepsilon = 1$ is widely viewed as an upper bound on the plausible range of elasticities (see e.g., Goolsbee 1999), a study that is unable to reject $\varepsilon = 1$ when $\widehat{\varepsilon} = 0$ is effectively uninformative about the structural elasticity.

Let $T_y(wl)$ denote the agent's tax liability as a function of his taxable income in year y . The function T will in general be non-linear and non-differentiable given the progressive bracket structure of the tax system in the U.S. In the nominal model, the agent chooses l to maximize

$$U_i(l; T_y) = w_i l - T_y(w_i l) - w_i \frac{l^{1+1/\varepsilon}}{1+1/\varepsilon} \quad (24)$$

Let $l_{i,y}^*$ denote the optimal labor supply for individual i given the tax system in year y . Now suppose there is a change in the tax system in year $y + 1$. The goal of the calibration is to compute the utility loss from ignoring this tax reform. Following the convention in the literature (e.g., Gruber and Saez 2002), I consider a three-year interval between the pre-reform and post-reform years, and define the utility loss from ignoring the tax changes that occurred between years y and $y + 3$ as

$$\Delta U = U_i(l_{i,y+3}^*; T_{y+3}) - U_i(l_{i,y}^*; T_{y+3}) \quad (25)$$

This expression can be interpreted in dollar terms because the utility function is quasilinear. Throughout, I consider a single tax filer with two children who has only labor income and no deductions other than those for children. I include both employer and employee payroll taxes and ignore state taxes. I calculate ΔU for various wage rates w_i and various years y using the following steps: (1) Calculate the income tax schedule in a “base year” y using the NBER TAXSIM calculator. Using a vector of wage rates, solve for the optimal labor supply $l_{i,y}^*$ and compute taxable income in the base year. Note that some taxable incomes are optimal for no one due to non-convexities in the tax schedule. (2) For each wage rate w_i , inflate by the CPI to obtain a comparable wage rate in year $y + 3$. Calculate the optimal labor supply $l_{i,y+3}^*$ using the tax schedule in year $y + 3$. (3) Calculate the dollar gain from reoptimizing for each wage rate w_i according to (25), using the inflated ($y + 3$) wage rate to compute taxable income in both cases.

The stata program TAXCOST.ado that calculates the utility cost of ignoring a tax reform has been made available on the NBER server. TAXCOST takes exactly the same inputs as TAXSIM. By running TAXCOST instead of TAXSIM on their datasets, researchers can calculate the utility costs of ignoring the tax changes that they use for identification in empirical studies. The program can be used to calculate utility costs for any given choice of elasticities, tax filer characteristics, and years.

4.3 Calibration Results: Synthesis of Existing Evidence

Tax Reform Act of 1986: Low vs. High Incomes. Figure 8 considers the Tax Reform Act of 1986, one of the largest reforms in the tax code in the U.S. and the focus of many empirical studies in the taxable income literature. Panel A shows the marginal tax rate schedules in 1985 (red, solid thick line) and 1988 (grey, solid blue line) for a single tax filer with two children. The tax reform cut marginal tax rates significantly across the board, especially for high income individuals. The dashed blue line in Figure 8a shows the change in the marginal net-of-tax rate $\frac{\Delta t}{1-t}$, which is the variable that corresponds to the “price change” in our analysis. The percentage change in the NTR is approximately 15-20% for those with incomes below \$100,000 and approaches 40% for those with incomes close to \$200,000. This percentage change in NTR schedule is replicated in light blue in Panels B-D for reference.

Panel B plots the dollar gain from reoptimizing, calculated according to the procedure

described above, vs. base year taxable income. Each point on the red curve is for an individual with a wage rate w_i such that his base-year taxable income (expressed in real 1988 dollars) equals the x axis value.⁷ For instance, an individual whose wage rate placed him at an optimal taxable income of \$100,000 prior to TRA86 would gain \$2,000 by reoptimizing in response to the change in the tax code if his fundamental elasticity of taxable income were $\varepsilon = 1$. To help in evaluating magnitudes, Panel C plots the percentage benefit of reoptimization, defined as the dollar gain divided by consumption in year $y+3$ (post-tax earnings) if one does not reoptimize. This percentage benefit measure corresponds to the $\Delta u_{\%}$ measure used in Corollary 2. Most taxpayers earning less than \$100,000 gain less than 2% of consumption by reoptimizing labor supply in response to TRA86.

Finally, Panel D plots the change in taxable income ($w_i l_{i,y+3}^* - w_i l_{i,y}^*$) required to fully reoptimize relative to TRA86 by income level with. A taxpayer earning \$100,000 prior to the reform would have to increase his pre-tax earnings by \$30,000 in order to reach his new optimum with an elasticity of $\varepsilon = 1$. This substantial change would still give him a utility gain (net of the disutility of added labor) of only \$2,000, because of the local flatness of utility functions that have $\varepsilon = 1$. Given that the search costs of finding a new job or additional work that pays an extra \$30,000 could well exceed \$2,000, it is plausible that this individual would not respond to TRA86 given plausible frictions.

The main lesson to be taken from Figure 8 is that the gains from reoptimization in response to TRA86 are very small for taxpayers with incomes below \$100,000 even for a fundamental elasticity as large as $\varepsilon = 1$. The gain are, however, considerably larger for high income earners. An individual with base year income of \$200,000 would gain nearly \$10,000 (8.5% of consumption) from reoptimizing in response to the tax reform. The dollar gains from reoptimization rise rapidly with income for two reasons: (1) the absolute level of the dollars at stake rise with income and (2) the change in the tax rates is larger for high incomes, as shown by the dashed blue line. The percentage gain also rises with income, showing that higher income individuals have greater reason to reoptimize even if frictions scale up proportionally with income.

The small gains from reoptimization for lower earners and large gains from reoptimization

⁷We fill in values at non-convex kinks in the base year by interpolating in order to obtain a smooth curve. Since no individual would optimally locate at these points in the base year, the value of reoptimization would be undefined.

for high earners are consistent with the findings of empirical studies of TRA86. These studies generally find substantial responses for high incomes but little or no response for lower incomes (see e.g., Feldstein 1995, Saez 2004). For individuals with incomes below \$100,000, Corollary 2 implies that one would not be able to rule out fundamental elasticities in excess of $\varepsilon = 1$ with optimization frictions of $\delta = 1\%$ even if one were to observe zero response to the tax reform. In contrast, high income earners would be leaving nearly 10% of consumption on the table by failing to reoptimize if $\varepsilon = 1$. This would make them more likely to respond even if they have the same fundamental elasticity ε as lower income households.

Note that high income individuals may also be more responsive because of differences in their ability to adjust income (lower optimization frictions) or because of heterogeneity in fundamental elasticities. The analysis here does not rule out these other channels. It merely shows that the simple model in (21) with a constant elasticity ε across individuals can explain the available data when one permits small optimization frictions. Hence, one does not necessarily need sophisticated alternative models to explain the data, a finding that recurs throughout the analysis below.

EITC Expansions: Intensive vs. Extensive Margin. Figure 9 considers another important episode in U.S. tax policy – the expansion of the Earned Income Tax Credit under the Clinton administration. Panel A shows that between 1993 and 1996, net-of-tax wage rates effectively rose by 20% for single tax filers with two or more children earning below \$10,000. This is because of the increase in the phase-in subsidy rate of the EITC to 40% during this period. For households with incomes between roughly \$15,000 and \$30,000, net-of-tax wages fell by roughly 15% because of the increase in the phase-out tax of the EITC.

Panel B plots the percentage gains of reoptimization, computed in the same way as in Figure 8c. Again, most of the percentage gains from reoptimizing in response to the EITC expansion are less than 2%. There are a small set of tax filers for whom the gains from reoptimization are around 4%; these are individuals who fall into the phase-out range only after the EITC expansion and thus experience a large increase in marginal tax rates. Given the substantial volatility of income across years for low income households (Chetty and Saez 2009), the expected gains from reoptimization are likely to be close to 2% for most households affected by this expansion. Again, the low gains from reoptimization can explain the empirical

findings. Most studies find virtually no changes in earnings in response to EITC expansions for individuals on the intensive margin (see Hotz and Scholz 2003, Eissa and Hoynes 2006). Using the result in Corollary 2, optimization frictions in excess of 1% would be consistent with these findings even if the fundamental elasticity were $\varepsilon = 1$.

Empirical studies of the EITC do find a substantial response on the extensive margin in response to the Clinton EITC expansion: labor force participation rates for single women with children surged during this period (Meyer and Rosenbaum 2002). The analysis thus far in this paper has focused solely on intensive margin reoptimization, where agents choose smoothly over earnings. To model extensive margin responses, suppose that individual i must pay a fixed cost k_i to enter the labor force (e.g. child care costs) and again consider the linear tax model in (21). Letting l_i^* denote the optimal labor supply choice conditional on working, individuals with fixed costs

$$k_i > (1 - t)w_i l_i^*(t) - \alpha_i \frac{(l_i^*(t))^{1+1/\varepsilon}}{1 + 1/\varepsilon}$$

will choose not to work. Consider an agent for whom this condition holds with equality at tax rate t_0 , and suppose this marginal individual is initially not working. Now suppose that the tax rate is reduced to $t_1 < t_0$ (e.g. through an EITC expansion). The agent's optimal response to this tax cut is to start working, as the utility from doing so is now strictly positive. The utility loss from failing to enter the labor force and work $l_i^*(t_0)$ hours is

$$\Delta u_{\text{ext}} = (1 - t_1)w_i l_i^*(t_0) - \alpha_i \frac{(l_i^*(t_0))^{1+1/\varepsilon}}{1 + 1/\varepsilon} - k_i = -\Delta t w_i l_i^*(t_0). \quad (26)$$

Finally, to convert this measure into percentage units, it is convenient to normalize Δu_{ext} by the marginal agent's consumption level if he were to work, which is $(1 - t_0)w_i l_i^*(t_0)$:

$$\Delta u_{\text{ext},\%} \equiv \frac{\Delta u_{\text{ext}}}{(1 - t_0)w_i l_i^*(t_0)} = -\frac{\Delta t}{1 - t_0}. \quad (27)$$

The critical difference between (26) and its intensive-margin counterpart in (23) is that the extensive margin utility loss is linear in Δt whereas the intensive-margin utility loss is a quadratic function of Δt .⁸ There is a *first-order* gain from reoptimizing for agents at the extensive margin. The percentage utility cost of ignoring a 10% increase in the net-of-tax

⁸Equation (26) understates the utility gain from reoptimization because it does not account for the fact that the optimal level of labor supply conditional on working will now be higher as well, given a reduction in the tax rate. This additional effect adds a second-order term analogous to that in (25).

wage is thus 10% on the extensive margin, compared with 0.5% on the intensive margin. Mathematically, the reason is that agents on the extensive margin are not near their post-reform optimum to begin with, unlike agents on the intensive margin. Intuitively, the first-order gains from a tax cut (higher consumption) are obtained on the intensive margin even if one does not change his behavior. On the extensive margin, non-workers get the first-order benefits of the tax cut (e.g. a larger EITC refund) only by changing their behavior and starting to work.

Figure 10 plots the extensive margin gains from reoptimizing relative to the Clinton EITC expansion shown in Figure 9. The x axis of these figures is the taxable income that the individual would optimally earn ($w_i l_{i,1993}^*$) were he to pay his fixed cost and work prior to the EITC expansion. On the extensive margin, the relevant tax rates are average rather than marginal tax rates. Panel A therefore plots the average tax rate vs. income prior to the EITC expansion and after the EITC expansion. The blue curve shows that individuals earning less than \$10,000 experienced a 20% increase in their net-of-tax earnings as a result of this reform.

Panel B plots the percentage utility gain from reoptimizing labor supply in response to this reform for individuals who are on the margin of entering the labor force. The denominator used to calculate the percentage benefit is the post-year consumption level, as in (27). The calibration follows the same procedure as that described in the previous section, except that at each income level, I assign the agent a fixed cost so that he would be just indifferent between working and not under the base year tax regime. I then compute the difference in utility between choosing labor supply optimally under the new tax regime in year $y + 3$ and not working in year $y + 3$ (and obtaining utility of 0).

As expected, the utility gains from reoptimization on the extensive margin are an order of magnitude larger than the corresponding values for the intensive margin. For a marginal individual who would earn \$10,000 when working in 1993, the gain from entering the labor force in response to the Clinton EITC expansion exceeds 20% of consumption (\$2,000). The intuition is straightforward: this individual would not have gotten the additional \$2,000 in the EITC refund if he had stayed out of the labor force. The utility gain exceeds the change in the average tax rate everywhere because it incorporates both the extensive margin gain in (27) and the gains from choosing a better level of hours on the intensive margin.

Because the utility gains of reoptimizing on the extensive margin are so large, even small

fundamental elasticities on the extensive margin will induce behavioral responses. In this sense, it is not surprising that empirical studies have found much clearer effects of the EITC expansion on the extensive rather than the intensive margin.

Tax Reforms from 1970-2006. The analysis above has focused on two tax major policy experiments that have received considerable attention in the empirical literature. In Figure 11, I extend this analysis to cover all tax changes from 1970-2006. I compute the percentage gain ($\Delta u\%$) from reoptimizing relative to tax reforms on the intensive and extensive margin at the 20th, 50th, and 99.5th percentile of the income distribution. The value that is plotted for year y is the percentage utility cost of choosing the same level of labor supply as the optimal choice in year $y - 3$.⁹ The data plotted in this figure, along with the corresponding percentage changes in the net-of-tax rate in each year for each of the three percentiles, are listed in Table 1.

Figure 11a shows that on the intensive margin, there is no tax change since 1970 for which the utility loss from a failure to reoptimize exceeds 1.5% of consumption for the median individual in the U.S. The utility costs of ignoring tax reforms are substantial only for the top 1% of income earners. The costs of ignoring TRA86 are an outlier for top income earners, and correspondingly TRA86 is the reform that generates the largest behavioral responses (Saez 2004). Figure 11b shows that in contrast, there are several tax changes that would generate substantial utility losses if ignored on the extensive margin even at the lower percentiles of the income distribution.¹⁰ The utility costs are particularly large for low income individuals, as changes in tax and transfer policies have induced large changes in average tax rates for low income earners. These results are consistent with findings of substantial extensive margin responses in the labor supply literature (even excluding the EITC reforms), particularly for individuals who earn low incomes when working.

The examination of the history of the tax code suggests that tax reforms in the U.S. are not a very powerful source of identification for learning about intensive margin labor supply behavior in the presence of small optimization frictions (except for very high incomes). Hence,

⁹For the extensive margin calculations, I assume that the marginal worker is in the labor force in cases where the average tax rates rises over the next three years and out of the labor force for cases where it falls. This is the relevant calculation to determine the utility gains of reoptimization for marginal agents who were not already making the optimal extensive margin choice.

¹⁰I exclude the 99.5 percentile from this figure because it is unlikely that a worker on the extensive margin would come in at the 99.5 percentile of the income distribution.

it is not surprising that studies of tax reforms have failed to detect intensive margin behavioral responses for most individuals (see e.g., Saez 2004). With 1% optimization frictions, one cannot rule out the possibility that $\varepsilon > 1$ based on such evidence.

Bunching at Kinks and Non-Linear Budget Set Models. Another common approach to estimating labor supply elasticities is to exploit the variation in tax rates across brackets within a cross section rather than focusing on reforms across years. This is the approach taken by the non-linear budget set (NLBS) literature (see e.g., Hausman 1983). A central challenge in fitting non-linear budget set models to the data is that optimizing models predict substantial bunching around kink points, but empirical income distributions for wage earners exhibit no such bunching (Saez 2002). The lack of bunching generally leads to rejection of NLBS models, or estimates inconsistent with economic theory (such as negative compensated wage elasticities). This problem has led to various ad hoc approaches to “removing” the variation at the kinks, such as smoothing the budget set (MaCurdy et al. 1990) or restricting the compensated elasticity to be positive.

Allowing for optimization frictions may provide a simple and disciplined method of improving the fit of NLBS models by providing an explanation for why individuals do not bunch at most kinks in the tax code. In particular, the utility losses from failing to locate at the kink are very small for most taxpayers. This is illustrated in Figure 12, which plots the income tax schedule in 2006 in the U.S. for a single filer with two children. The number next to each convex kink in the schedule shows the percentage utility gain ($\Delta u_{\%}(\varepsilon = 1)$) from locating at that kink point relative to optimizing under (incorrect) assumption that the rate in the previous bracket continues into the next bracket. There are many wage rates that can induce individuals to locate at each kink. The numbers listed on the figure are the (unweighted) mean percentage gain across all individuals who would optimally locate at the kink. The calculations in Figure 12 show that the utility losses from ignoring the kinks are generally quite small in magnitude. Coupled with volatility in income and the movement of the kinks over time, it is not surprising that individuals who face small optimization frictions would not bother to target the kinks.

Saez (2002) documents that the one kink at which there is some evidence of bunching is the first kink in the tax schedule, generated by the end of the phase-in region of the EITC. The

bunching at this kink is driven entirely by individuals who report self-employment income, which audit studies indicate is frequently misreported on income tax returns because of the lack of double reporting. Unlike “real” labor supply responses, misreporting can generate a first-order gain in welfare because it creates a pure transfer of resources from the government to the taxpayer. Hence, the gains from paying attention to the tax schedule are large for taxpayers who are able to manipulate their reported income, potentially explaining why one observes bunching precisely for this group. Interestingly, Saez (2002) finds that there is no bunching at the second kink of the EITC schedule (where the phase-out region begins) even for the self-employed. This is consistent with the view that the bunching at the first kink is driven by the first-order gains from misreporting. If one were able to pick reported taxable income without constraints, the best point to pick is the first kink in the EITC schedule, because this point maximizes the size of the EITC refund while minimizing payroll tax liabilities. There is therefore no reason to locate at the second kink if bunching is driven by income manipulation. In contrast, in a frictionless model where individuals change their behavior to seek second-order gains from locating at kinks, one would predict bunching at *both* the first and second EITC kinks.

This analysis suggests that introducing optimization errors as in (8) could provide a disciplined method of estimating non-linear budget set models. When fitting such models with maximum likelihood, one could effectively permit agents to choose any point that yields utility within δ units of the maximal utility they can attain. This would reduce the predicted excess mass at kink points and potentially permit identification of the model using other moments of the data. Exploiting cross-sectional variation in tax rates in this manner may yield tighter bounds on intensive margin elasticities than natural experiments given the lack of large reforms in the U.S.

Optimizing Relative to Tax System. One potential reaction to the analysis above is that individuals might never optimize relative to the tax system, because the utility gains from doing so are always negligible. In this case, the “fundamental elasticity” ε would never be relevant for actually predicting behavioral responses to a tax system. Figure 13 shows that this is actually not the case: the utility costs of completely ignoring taxes in steady state are substantial. The figure plots the percentage utility gain ($\Delta u\%$) from optimizing on the

intensive margin relative to 2006 tax code relative to optimizing under assumption that there are no taxes at all. Even for middle income taxpayers, the value of optimizing relative to the tax system is on the order of 10% of consumption per year. Especially when the benefits are aggregated over several years, it is unlikely that optimization frictions would be large enough to prevent individuals from responding to the tax code in steady state. Hence, behavioral responses to the tax system are likely to be close to what one would predict based on the fundamental elasticity ε , even though observed behavioral responses to tax reforms may be quite different because of optimization frictions.

A Synthesis of Stylized Facts. In addition to the results described above – (1) larger responses for the rich, (2) larger extensive margin responses, (3) rejection of NLBS models and no bunching at kinks for wage earners in the U.S. tax code, and (4) bunching at kinks for the self-employed – allowing for small optimization frictions may also explain the following findings:

(5) Studies identified using macroeconomic (cross-country/long-run) variation find larger elasticities than microeconomic studies that focus on short run changes (Prescott 2004, Davis and Henrekson 2005). The variation in net-of-tax rates used in these macroeconomic studies is much larger than in most microeconomic studies, and accordingly these studies generate much tighter bounds, as shown in the next subsection.

(6) Eligibility cutoffs for transfer programs such as Medicaid generate significant behavioral responses – individuals appear to actively keep their incomes below such cutoffs (e.g., Yelowitz 1995). The utility costs of ignoring such eligibility cutoffs are first-order, making it optimal to respond to them even in the presence of substantial optimization frictions.

(7) Elasticities are historically larger for secondary earners than primary earners, but have converged over time (Eissa and Hoynes 2004, Blau and Kahn 2005). The difference in observed elasticities need not reflect fundamental variation in preferences. It could instead be explained by differences in optimization frictions – for instance, primary earners may be more likely to hold manufacturing jobs with rigidities while secondary earners may be more likely to hold flexible jobs such as teaching or nursing. As secondary earners' jobs become more likely primary earners, the observed elasticities converge, even though the fundamental elasticity ε may have been constant throughout.

(8) Responses to information: Chetty, Looney, and Kroft (2008) and Chetty and Saez (2009) conduct field experiments which show that providing information about the tax code significantly affects behavioral responses to taxation. If information and salience reduce optimization frictions, they may lead to larger observed behavioral responses given a fixed fundamental elasticity ε .

(9) Small vs. large tax changes: Chetty et al. (2009) document that larger tax changes induce much larger behavioral responses using Danish data. The analysis here shows that larger tax therefore yield much sharper bounds on ε than small tax changes, and accordingly should be more likely to generate behavioral responses.

4.4 Bounds on the Taxable Income Elasticity

In this section, I calculate the bounds on the intensive-margin taxable income elasticity implied by fifteen studies in the existing literature using the formula in (17). For each study, I use the elasticity estimate reported from the authors' preferred specification. To calculate $\frac{\Delta t}{1-t}$, I either use the reported percentage change in the net of tax rate used for identification (taking means in cases with several reforms), or calculate the value of $\frac{\Delta t}{1-t}$ using data on marginal tax rates over the authors' sample period.

Table 2 shows the list of studies used for the analysis. The studies include evaluations of the major tax reforms in the U.S., cross-country comparisons from the macroeconomics literature, and studies of reforms and bunching in European countries. The table lists the point estimate of the elasticity in each study, the change in the net of tax rate used for identification, and the implied values of the bounds on the taxable income elasticity ($\varepsilon_L, \varepsilon_U$) for each study with $\delta = 1\%$ optimization frictions. For example, Feldstein's (1995) pioneering analysis of TRA86 finds an elasticity estimate of 1.04 and implies bounds of (0.37, 2.89).

The width of the bounds varies tremendously with the size of the variation used for identification. The two studies that pool tax reforms over the 1980s for the full population – Gruber and Saez (2002) and Kopczuk (2005) – use variation on average of roughly 6% in the net-of-tax-wage to identify the elasticities. These two studies imply an upper bound of $\varepsilon_U > 24$. With 1% optimization frictions, the responses detected by these studies would be consistent with enormous elasticities of taxable income, simply because the price variation used is not big enough to overcome small frictions. Chetty et al.'s (2009) study of bunching

at small kinks in Denmark also has essentially uninformative bounds, with $\varepsilon_L = 0$ and $\varepsilon_U = 8$. These results show that pooling several small reforms or using very large datasets is useful for achieving statistical precision, but is not informative about magnitudes in an environment where one is uncertain about the optimization frictions that agents face.

Figure 14 gives a visual representation of the bounds implied by each of the studies. For scaling purposes, I exclude studies that use variation in net-of-tax rates below 20% for identification, eliminating the three studies just discussed from the figure. The figure has two lessons. First, none of the intervals plotted in the figure are disjoint – that is, the lower bound of every study falls below the upper bound of every other study. Even though there is substantial dispersion in the estimates of the observed elasticities, all the estimates in the literature can be reconciled with a single fundamental elasticity ε in the simplest labor-leisure choice model given 1% optimization frictions. One perspective on this result is that it just indicates how uninformative existing studies are in the presence of optimization frictions. This view has some validity, in that some of the studies with very small or large elasticity estimates are consistent with the rest largely because they imply very wide bounds.

However, when pooled together, the fifteen studies in Table 2 actually yield very informative bounds on the taxable income elasticity. The unified bounds can be computed by taking the largest of the lower bounds and the smallest of the upper bounds. The unified lower bound across the fifteen studies is $\varepsilon_L = 0.47$, obtained from Goolsbee’s (1999) analysis of TRA86. The unified upper bound is $\varepsilon_U = 0.54$, obtained from Gelber’s (2008) analysis of the Swedish tax reform of 1991. This points to the second broad lesson from Figure 14: combining several studies can yield informative bounds even though any one study considered in isolation implies fairly wide bounds.

The tight bounds of $\varepsilon \in (0.47, 0.54)$ are achieved because of the substantial variation in the point estimates across studies. Some of this variation presumably arises from differences in optimization frictions and benefits of reoptimization in the environments studied. But part of the variation in the point estimates is known to be driven by flaws in some of the statistical estimators. For instance, the highest elasticity estimates from TRA86 have been shown to be biased upward by violations of the common trends assumption in difference-in-difference estimators and because of intertemporal shifting (Saez et al. 2009). In addition, the macroeconomic studies may be biased upward because they do not fully control for variations

in labor market institutions across country.

In view of these statistical issues, I construct a second set of unified bounds that exclude studies of top income earners around TRA86 (studies 1, 2, and 3) and the macro studies that make cross-country comparisons (studies 7 and 11). The remaining studies imply a lower bound of $\varepsilon_L = 0.30$, obtained from Goolsbee's study of a large tax reform in the 1920s and $\varepsilon_U = 0.54$, again from Gelber's (2008) study. Hence, even without the macro and TRA86 studies, one still obtains reasonably informative bounds on the taxable income elasticity. If one further excludes the Goolsbee study of the 1920s on the grounds that the fundamental parameter it estimates may differ given the very different structure of the tax system in the modern era (Kopczuk 2005), then the unified bounds are $(0.12, 0.54)$. In this case, the lower bound of 0.12 is driven by Saez's (2004) estimates for top income earner's reactions to tax changes using historical time series variation.

In addition to the statistical identification concerns, one must be very cautious in interpreting the bounds obtained from this simple exercise for a number of other reasons. First, I have assumed a common fundamental elasticity across all the studies. This strong assumption ignores variation in preferences across income levels or countries. Second, the static nominal model in (21) abstracts from many considerations in labor supply decisions, the most important of which are perhaps lifecycle effects (MaCurdy 1981). Finally, I have ignored the statistical imprecision of the point estimates, which is substantial in many of the fifteen studies. Incorporating these standard errors would yield wider bounds on ε .

What types of studies would be most useful in achieving more informative bounds on the taxable income elasticity? Ideally, one should study a tax reform that is both large and involves minimal optimization frictions – e.g. a large, salient tax change that applies to a group of workers who can adjust labor supply easily such as cab drivers. Lacking this ideal, the upper bound ε_U can be credibly tightened by microeconomic studies of very large tax reforms or cross-sectional variation, even in environments where optimization frictions are not minimal. Such studies would be of interest even if they yield zero or small elasticity estimates. The lower bound ε_L can be increased by obtaining positive elasticity estimates from moderate-sized experiments in low friction environments – e.g. well publicized experiments on subgroups that can adjust labor supply easily.

5 Conclusion

There are many frictions which prevent agents from making the optimal choices predicted by standard economic models. This paper has shown that even if a researcher is uncertain about how these frictions affect behavior, he can obtain bounds on price elasticities. The bounds are derived by requiring that agents' choices yield utility near the unconstrained maximum that can be achieved under a given nominal model. Abstractly, I exchange the standard orthogonality condition on the error term for a bounded support condition based on the utility costs of errors. The bounded support on the error term yields bounds on the price elasticity. The bounds have an approximate analytical representation that is robust to functional form specifications of utility for small degrees of optimization frictions.

Applying this method to the existing literature on taxation and labor supply offers a critique and synthesis of this literature. The critique is that many existing studies of labor supply are essentially uninformative about the taxable income elasticity because they cannot reject very large values of the elasticity if one permits optimization frictions equal to 1% of consumption. The synthesis is that several patterns in this literature can be reconciled if one permits a small degree of optimization frictions. By pooling estimates from several studies, one can achieve fairly informative bounds on the taxable income elasticity.

This paper has focused on a very simple static neoclassical model. From a methodological perspective, it would be interesting to build on the approach proposed here to derive bounds with optimization frictions in richer environments, including dynamic models and discrete choice settings. From an applied perspective, it could be useful to apply the bounds derived here to other literatures where the values of key parameters are debated, such as the elasticity of intertemporal substitution in macroeconomics or the effects of the minimum wage on employment in labor economics. Such analyses would shed light on which disagreements require fundamentally different models and which can be reconciled simply by allowing for small frictions.

References

- Akerlof, George A. and Janet L. Yellen (1985). "Can Small Deviations from Rationality Make Significant Differences to Economic Equilibria?" *American Economic Review*, 75(4): 708-20
- Altonji, Joseph G. and Christina H. Paxson (1992). "Labor Supply, Hours Constraints, and Job Mobility," *Journal of Human Resources*, 27(2): 256-278.
- Auten, Gerald and Robert Carroll (1999). "The Effect of Income Taxes on Household Behavior," *Review of Economics and Statistics*, 81: 681-693.
- Blundell, R., A. Gosling, H. Ichimura, and C. Meghir, (2007). "Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds," *Econometrica* 75(2): 323-363.
- Blau, Francine and Lawrence Kahn (2005). "Changes in the Labor Supply Behavior of Married Women: 1980-2000," NBER Working Paper 11230, 2005.
- Chetty, Raj, Adam Looney and Kory Kroft (2008). "Salience and Taxation: Theory and Evidence," *American Economic Review*, forthcoming.
- Chetty, Raj and Emmanuel Saez (2009). "Teaching the Tax Code: Earnings Responses to a Field Experiment with EITC Recipients," UC-Berkeley mimeo.
- Chetty, Raj, John Friedman, Tore Olsen, Luigi Pistaferri, and Anders Fredriksen (2009). "Adjustment Costs and Earnings Responses to Income Taxation: Evidence from Denmark," UC-Berkeley mimeo.
- Davis, Steven and Magnus Henrekson (2005). "Tax Effects on Work Activity, Industry Mix and Shadow Economy Size: Evidence from Rich-Country Comparisons," In *Labour Supply and Incentives to Work in Europe*, eds. R. Gomez Salvador, A. Lamo, B. Petrongolo, M. Ward, and E. Wasmer, Edward Elgar Press.
- de Bartolome, Charles A. M. (1995). "Which tax rate do people use: Average or marginal?" *Journal of Public Economics*, 56(1): 79-96.
- Del Boca, Daniela, and Annamaria Lusardi (2003). "Credit Market Constraints and Labor Market Decisions," *Labour Economics*, 10(6): 681-703.
- Duflo, Esther, William Gale, Jeffrey Liebman, Peter Orszag, and Emmanuel Saez (2006). "Saving Incentives for Low- and Middle-Income Families: Evidence from a Field Experiment with H&R Block," *Quarterly Journal of Economics*, 121(4): 1311-1346.
- Eissa, Nada and Hilary Hoynes (2004). "Taxes and the Labor Market Participation of Married Couples: The Earned Income Tax Credit," *Journal of Public Economics*, 88(9-10), 1931-1958.
- Eissa, Nada and Hilary Hoynes (2006). "Behavioral Responses to Taxes: Lessons from the EITC and Labor Supply," *Tax Policy and the Economy*, 20: 74-110.
- Eissa, Nada and Jeffrey Liebman (1996). "Labor Supply Response to the Earned Income Tax Credit," *Quarterly Journal of Economics*, CXI, 605-637.
- Feldstein, Martin (1995). "The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1986 Tax Reform Act," *Journal of Political Economy*, 103: 551-57.

- Finkelstein, Amy (2008). "EZ-Tax: Tax Saliency and Tax Rates," *Quarterly Journal of Economics*, forthcoming.
- Fujii, Edwin T. and Clifford B. Hawley (1988). "On the Accuracy of Tax Perceptions," *Review of Economics and Statistics*, 344-347.
- Gelber, Alexander M. (2009). "Taxation and Family Labor Supply," NBER Working Paper .
- Goolsbee, Austan (1999). "Evidence on the high-income Laffer curve from six decades of tax reform," Brookings Papers on Economic Activity.
- Gruber, Jonathan and Emmanuel Saez (2002). "The elasticity of taxable income: evidence and implications," *Journal of Public Economics*, 84 (2002) 1-32.
- Hansen, Lars Peter and Thomas J. Sargent (2007). *Robustness*, Princeton University Press: Princeton, NJ.
- Hausman, Jerry A. (1983). "Tax and Labor Supply." In *Handbook of Public Economics*, eds, Alan Auerbach and Martin Feldstein.
- Honore, Bo and Eli Tamer (2006). "Bounds on Parameters in Dynamic Discrete Choice Models," *Econometrica* 74(3): 611-629.
- Hotz, V. Joseph and John Karl Scholz (2003). "The Earned Income Tax Credit" in Robert Moffitt, ed., *Means-Tested Transfer Programs in the United States*. Chicago: University of Chicago Press.
- Imbens, Guido W. (2007). "Whats New in Econometrics? Lecture 9," NBER Summer Institute Course.
- Imbens, Guido W., and Charles F. Manski (2004). "Confidence Intervals for Partially Identified Parameters," *Econometrica* 72: 1845-1857.
- Jones, Arthur-Damon (2008). "Information, Inertia, and Public Benefit Participation: Experimental Evidence from the Advance EITC", UC Berkeley working paper.
- Kopczuk, Wojciech (2005). "Tax Bases, Tax Rates and the Elasticity of Reported Income," *Journal of Public Economics*, 89(11-12): 2093-2119.
- Liebman, Jeffrey and Richard Zeckhauser (2004) "Schmeduling", Harvard University mimeo.
- MaCurdy, Thomas. "An Empirical Model of Labor Supply in a Life-Cycle Setting." *Journal of Political Economy*, 1981, 89, pp. 1059-1085.
- MaCurdy, Thomas, David Green and Harry Paarsche (1990). "Assessing Empirical Approaches for Analyzing Taxes and Labor Supply," *Journal of Human Resources*, 25: 415-490.
- Manski, Charles F. (2007). *Identification for Prediction and Decision*, Harvard University Press: Cambridge, MA.
- Meyer, Bruce and Dan Rosenbaum (2001). "Welfare, the Earned Income Tax Credit, and the Labor Supply of Single Mothers." *Quarterly Journal of Economics*, Vol. 116, Issue 3 (August).
- Prescott, Edward (2004). "Why Do Americans Work So Much More Than Europeans?" *Federal Reserve Bank of Minneapolis Quarterly Review*, 28(1).
- Rogerson, Richard and Johanna Wallenius (2008). "Micro and Macro Elasticities in a Life Cycle Model with Taxes," forthcoming, *Journal of Economic Theory*.

- Saez, Emmanuel (2002). "Do Taxpayers Bunch at Kink Points?" UC Berkeley Working paper.
- Saez, Emmanuel (2004). "Reported Incomes and Marginal Tax Rates, 1960-2000: Evidence and Policy Implications," NBER Working Papers 10273.
- Saez, Emmanuel, Joel Slemrod and Seth Giertz (2009). "The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review" Working Paper.
- Yelowitz, Aaron S. (1995). "The Medicaid Notch, Labor Supply, and Welfare Participation: Evidence from Eligibility Expansions," *Quarterly Journal of Economics*, 110(4): 909-39.

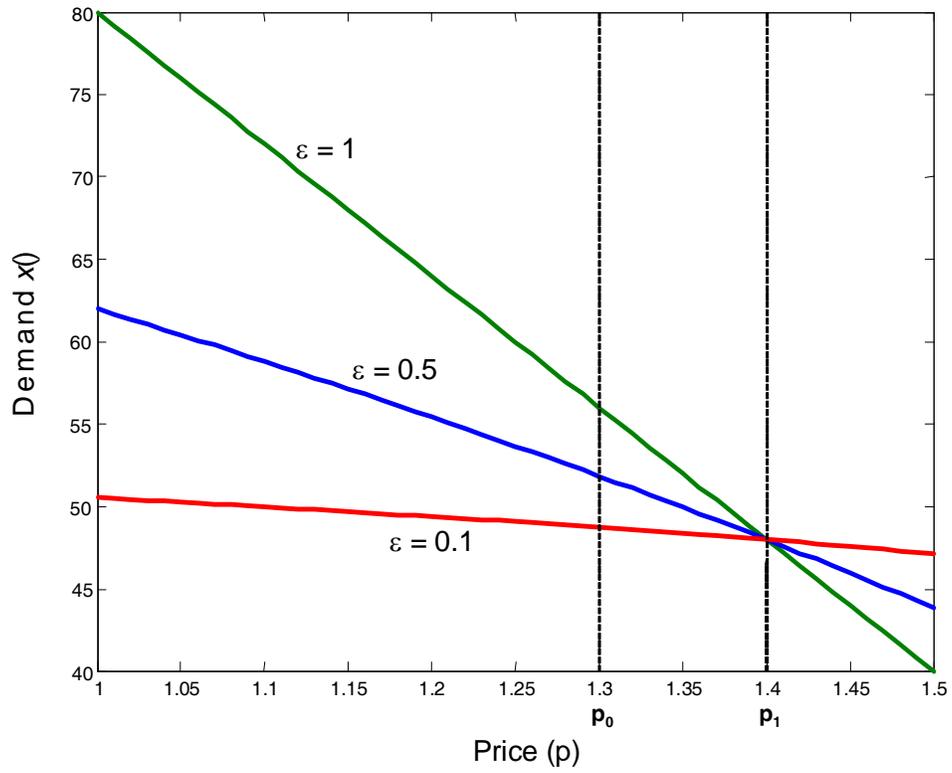
TABLE 1

Percentage Benefit of Re-Optimizing Relative to Tax Changes by Year By Income Percentile

Year	Intensive Margin						Extensive Margin					
	20		50		99.5		20		50		99.5	
	%ΔU	%ΔNTR	%ΔU	%ΔNTR	%ΔU	%ΔNTR	%ΔU	%ΔNTR	%ΔU	%ΔNTR	%ΔU	%ΔNTR
1970	0.01	-1.49	1.2	-16.61	1.2	-16.61	2.52	-2.49	2.59	-2.56	2.59	-2.56
1971	0	-0.72	1.1	12.73	0.02	2.76	0.3	0.29	1.1	-0.47	4.09	4.46
1972	0	0.8	0.04	4.55	2.14	26.58	3.57	3.71	2.24	2.25	11.1	10.17
1973	0.02	-2.3	0.79	-14.65	0.71	14.6	1.26	1.28	2.14	-2.12	4.06	3.57
1974	0.01	-1.72	0.05	-3.21	0.44	11.11	2.22	-2.23	2.47	-2.47	3.84	-3.84
1975	0.36	-14.95	1.51	-17.73	0	0	6.1	6.31	3.68	-3.65	5.88	-5.8
1976	0.53	-16.15	0	-0.9	0	0	6.9	7.18	0.25	-0.26	4.72	-4.7
1977	0.29	-13.46	0.1	-4.74	0	0	7.58	8.24	0.87	0.79	3.66	-3.65
1978	0.02	-3.73	0.19	-7.05	0	0	4.31	-4.28	1.21	-1.23	4.18	-4.13
1979	0.05	-4.76	0.29	-8.06	0	0	0.99	-1.04	1.99	-2.02	4.49	-4.45
1980	0.07	-4.76	0.06	-3.86	0	0	4.39	-4.57	4.13	-4.14	5.08	-5.02
1981	0.11	-5.65	0.15	-5.96	0	0	5.74	-6	5.12	-5.11	5.47	-5.36
1982	0.01	-1.86	0	-0.22	0	0	9.03	-8.95	3.63	-3.64	3	-3
1983	0	-0.23	0.04	2.92	0	0	5.74	-5.73	0.45	-0.46	1.38	1.43
1984	0.1	5.31	0.08	8.14	0	0	3.34	-3.34	1.73	1.7	4.78	5.13
1985	0.02	2.63	0.04	3.62	0	0	1.26	-1.28	0.86	0.83	3.73	3.94
1986	0	0.62	0	0.15	0	0	1.13	-1.13	0.58	-0.58	1.36	1.4
1987	0.09	5.2	0.79	14.03	2.03	23	7.59	8.35	3.99	3.33	13.69	13.27
1988	0.01	-2.76	0.69	13.05	7.04	44	10.08	11.43	3.94	3.39	24.39	20.96
1989	0.02	-2.44	0.72	13.42	6.99	44	9.86	11.27	4.06	3.49	27.67	25.92
1990	0.21	-7.73	0.01	-1.41	1.31	17.07	2.44	2.38	0.86	-0.87	7.77	6.86
1991	0.06	-4.4	0	-0.4	0.1	-4.17	0.9	0.87	0.6	-0.6	2.45	-2.39
1992	0.1	-5.7	0	-0.4	0.1	-4.17	2.7	2.77	0.23	-0.24	2.29	-2.24
1993	0.13	-6.58	0	0	0.91	-16.11	4.06	4.24	0.28	0.28	6.21	-5.88
1994	0.12	-9.28	0	0	0.76	-11.94	7.21	7.91	0.3	0.3	5.88	-5.58
1995	0.07	-12.52	0	0	1.13	-16.67	8.74	9.83	0.03	0.03	6.74	-6.36
1996	0.13	-12.78	0.11	0	0.08	-3.77	10.06	11.47	0.11	-0.01	2.44	-2.4
1997	0.06	-6.5	0	0	0	0	5.33	5.79	0.01	-0.01	0.02	0.02
1998	1.88	28.62	0	0	0	0	5.57	3.96	2.91	3.02	0.02	0.02
1999	2.15	30.84	0	0	0	0	3.66	1.57	3.49	3.64	0.19	-0.2
2000	2.17	30.84	0	0	0	0	3.09	0.95	3.14	3.26	0.39	-0.39
2001	0.48	15.71	0	0	0	0.87	3.76	3.49	2.35	2.44	0.46	0.46
2002	0.27	15.71	0	0	0.01	1.74	4.51	4.53	1.83	1.88	1.41	1.42
2003	0.73	15.71	0	0	0.28	8	5.35	4.95	4.03	4.24	5.52	5.55
2004	0.15	6.79	0	0	0.22	7.07	2.31	2.24	1.93	1.98	4.34	4.31
2005	0.15	6.79	0	0	0.17	6.15	1.48	1.39	1.66	1.71	3.34	3.29
2006	0.15	6.79	0	0	0	0	1.84	1.76	0.49	-0.49	0.31	0.32

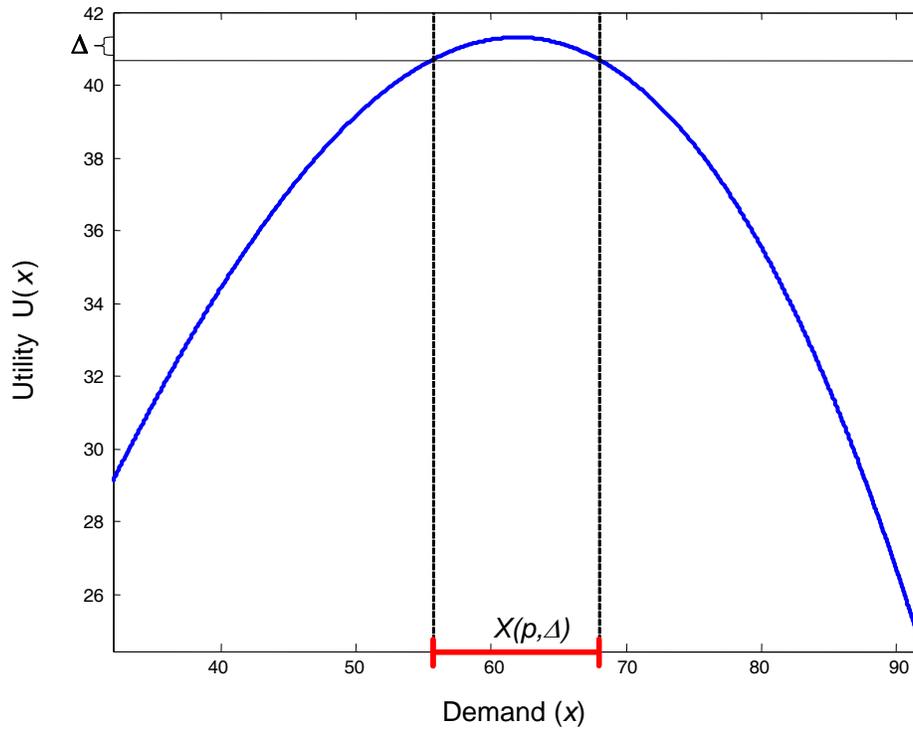
Note: Assumes that taxpayer last optimized three years before year listed in each row, and that income has changed at the rate of inflation. For single head of household with two children and no state taxes.

FIGURE 1
Identification of Elasticity in Nominal Model



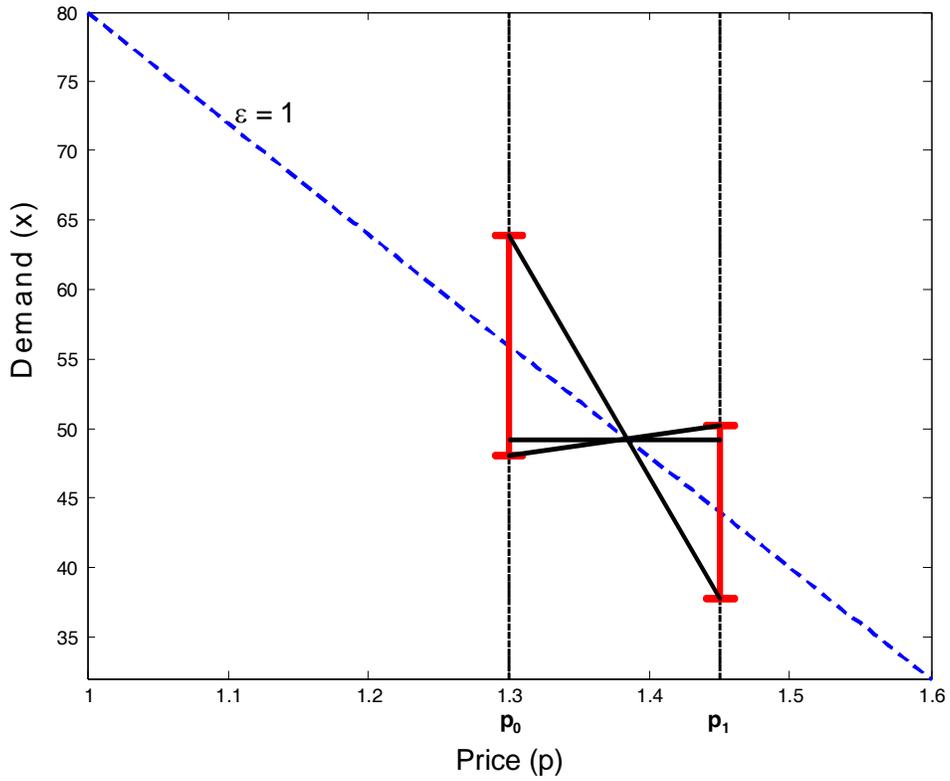
NOTE—This figure illustrates how measuring the demand response to a price increase from p_0 to p_1 identifies the structural parameter ϵ in the nominal model. Each curve depicts a demand function with a different price elasticity.

FIGURE 2
Construction of Choice Set



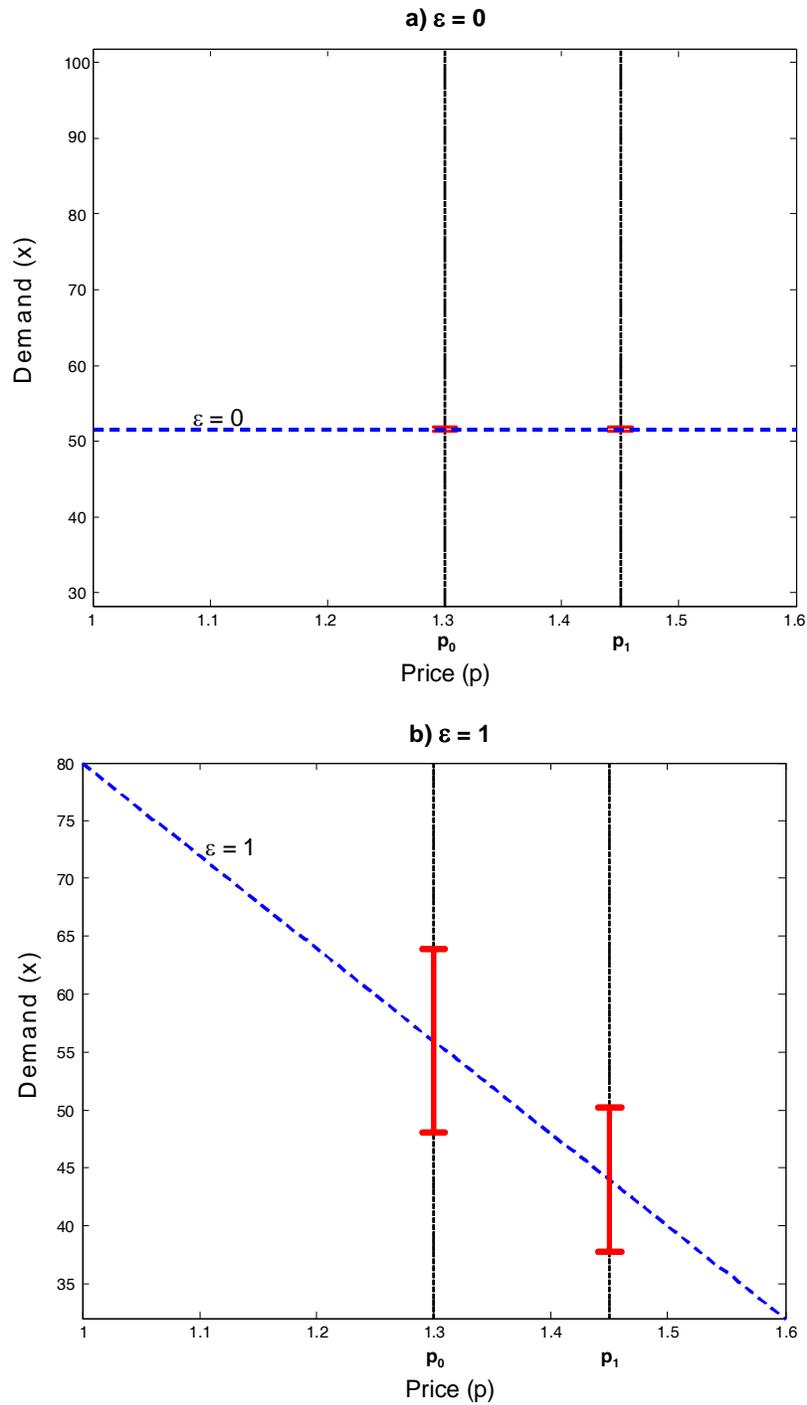
NOTE—This figure illustrates the construction of the choice set, $X_i(p, \Delta)$, predicted by a Δ class of models. The blue curve plots the utility $U(x)$. The agent is permitted to deviate from the nominal model only to the extent that the utility loss from doing so – as calculated under the nominal model – falls below the exogenously specified threshold Δ . The set of demand levels that yield utility within Δ units of the maximum is shown by the red interval on the x axis.

FIGURE 3
 Identification with Optimization Frictions



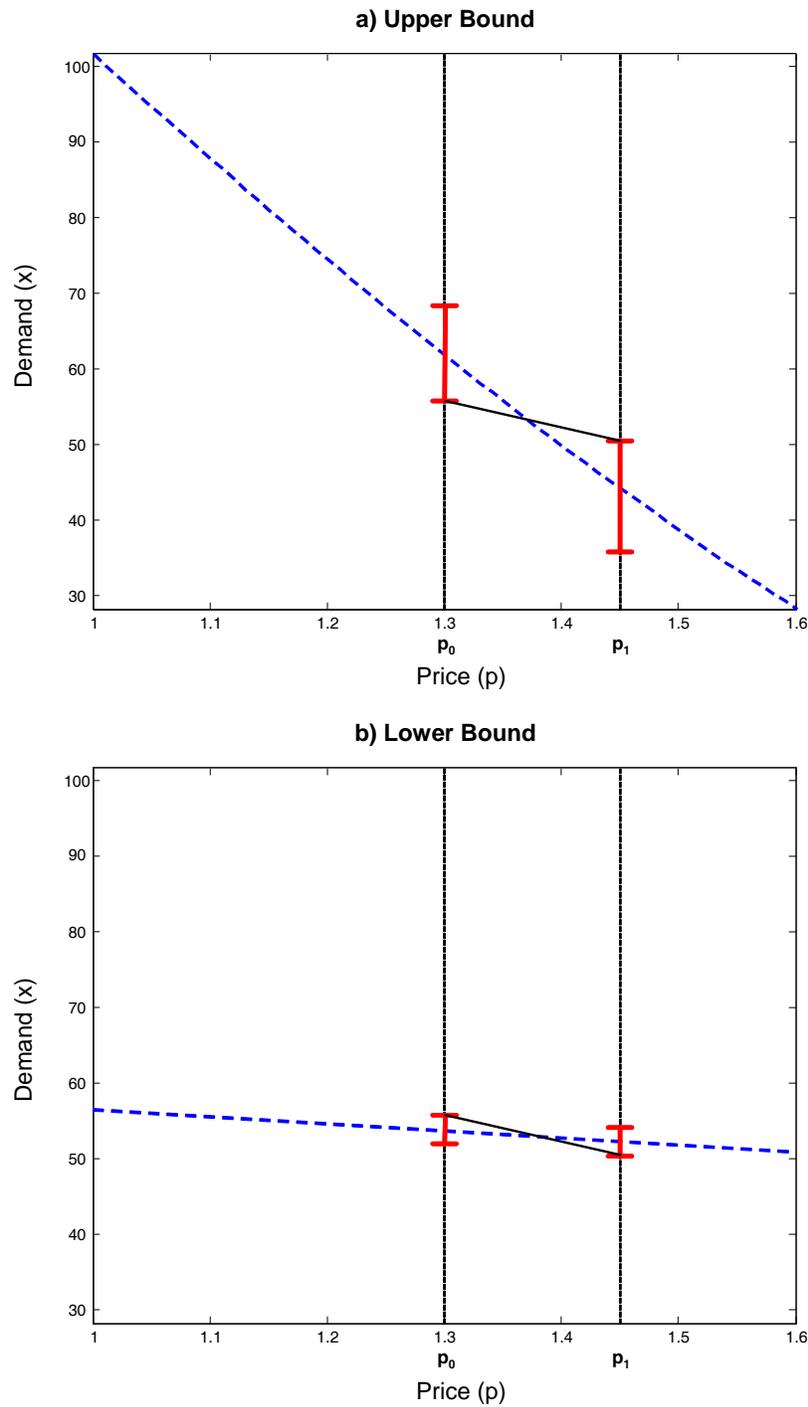
NOTE—This figure illustrates the choice sets at two price levels, $X(p_0, \Delta)$ and $X(p_1, \Delta)$. The fundamental elasticity $\varepsilon = 1$ controls the movement of the choice sets with the price p , as illustrated by the dashed blue line in the figure. The black lines illustrate that various responses $[x(p_1) - x(p_0)]$ may be observed when $\varepsilon = 1$, including large reductions, zero response, or even small increases.

FIGURE 4
Effect of Fundamental Elasticities on Choice Sets



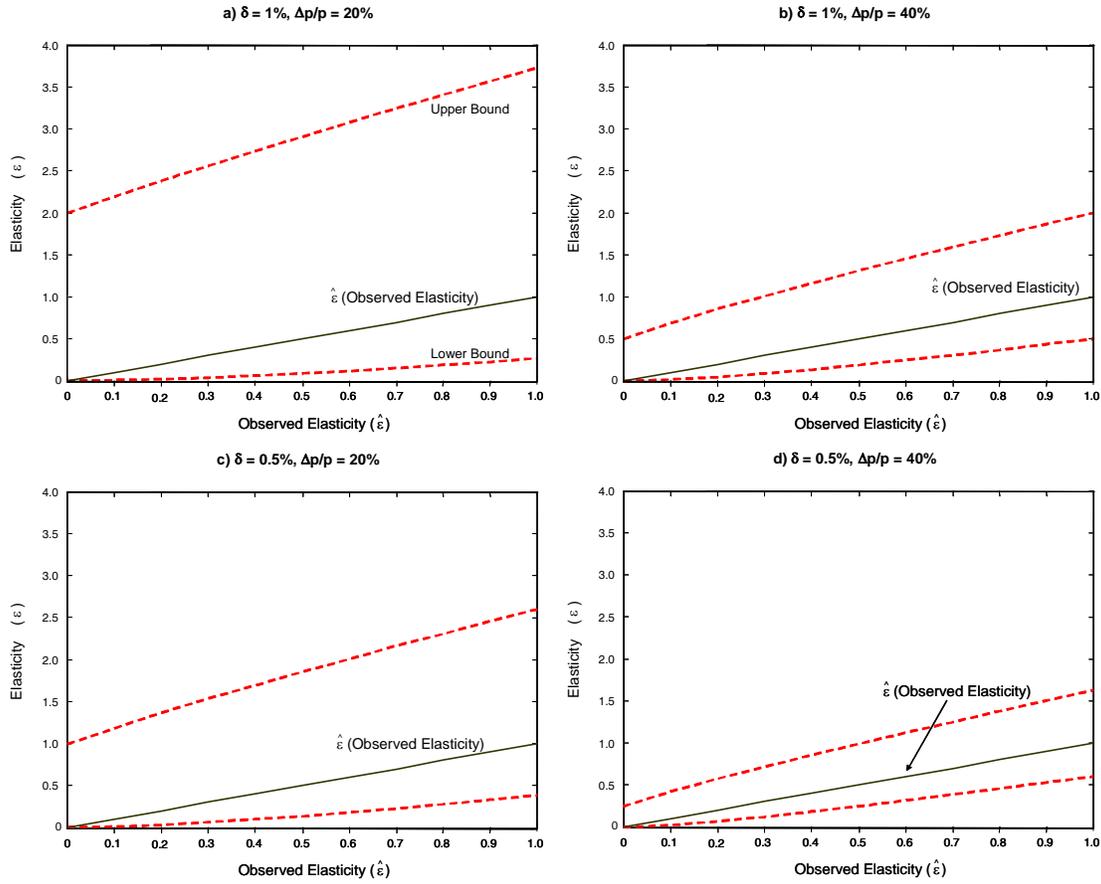
NOTE—This figure illustrates the result in Lemma 1, showing the effect of fundamental elasticities on choice sets at two price levels. When the elasticity is 0 (panel A), the choice set collapses to a single point. The choice sets are wider when $\varepsilon = 1$.

FIGURE 5
Bounding the Fundamental Elasticity with Optimization Frictions



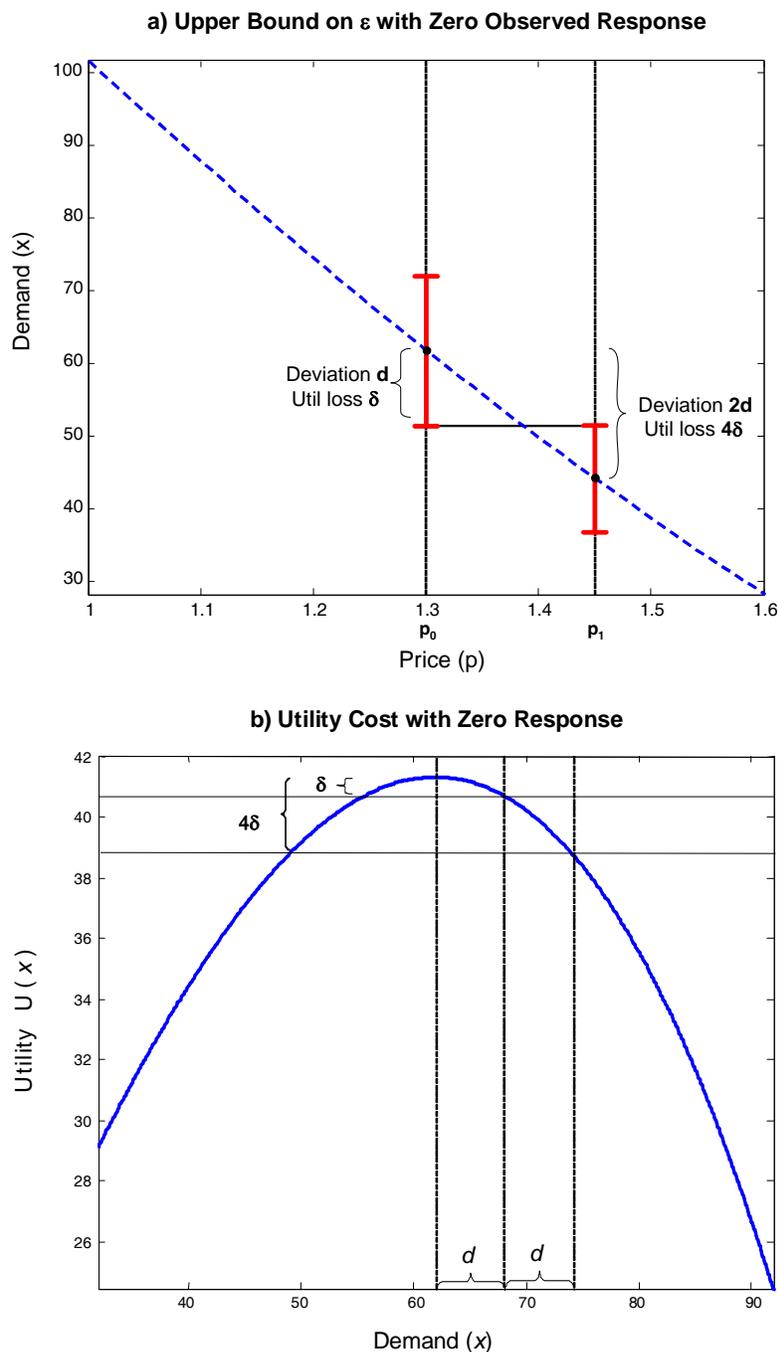
NOTE—This figure illustrates the general approach to bounding elasticities with optimization frictions. The solid black line in each panel depicts the observed treatment effect ($\hat{\epsilon}$) for a price increase from p_0 to p_1 . Panel A shows the highest fundamental elasticity ϵ (blue dashed line) that could have generated this observed treatment effect. Panel B shows the lowest fundamental elasticity ϵ (blue dashed line) that could have generated the same observed treatment effect.

FIGURE 6
Bounds on Fundamental Elasticities as a Function of Observed Elasticities



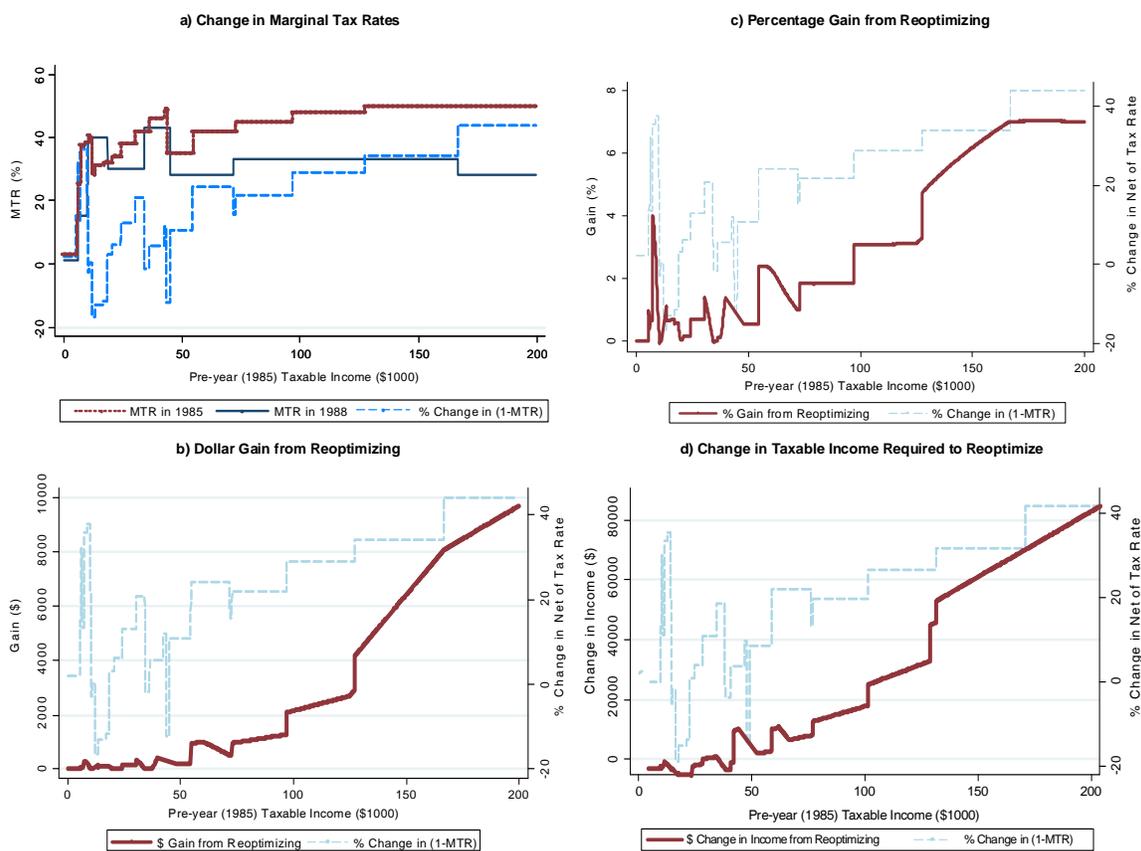
NOTE—This figure plots the bounds (ϵ_L, ϵ_U) vs. $\hat{\epsilon}$ for four combinations of δ and $\frac{\Delta p}{p}$. The bounds are computed using the formula in Proposition 1. In the top two panels, the degree of model uncertainty due to optimization frictions is $\delta = 1\%$. The lower two panels consider $\delta = 0.5\%$. The left panels have a price change of $\frac{\Delta p}{p} = 40\%$, while the right panels have $\frac{\Delta p}{p} = 20\%$.

FIGURE 7
Upper Bound on Fundamental Elasticity with Zero Observed Response



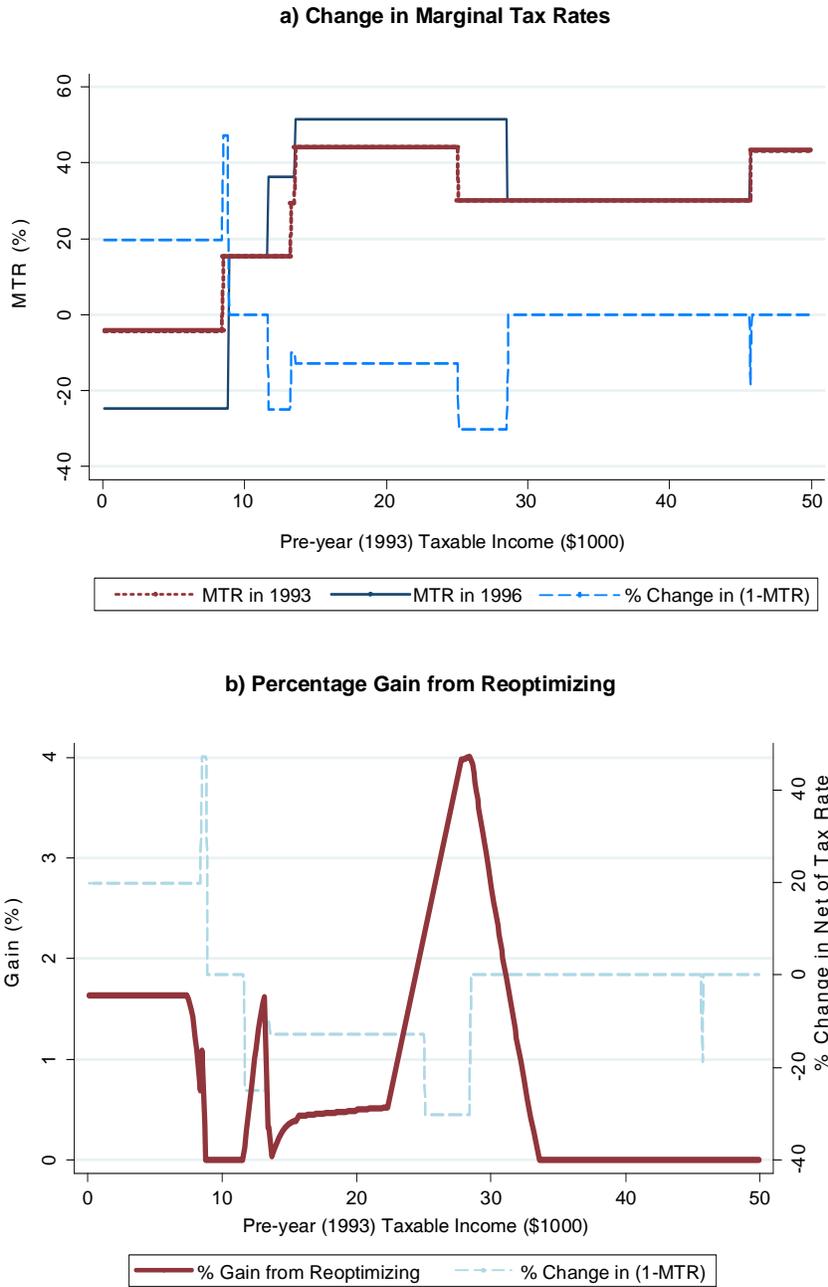
NOTE—This figure illustrates the result in Corollary 2. Panel A shows the largest fundamental elasticity (dashed blue line) consistent with zero observed response (solid black line) to a price increase from p_0 to p_1 . Under a quadratic approximation to $U(x)$, the difference between the optimal choices under the old and new prices must equal $2d$, where d is the distance from the end of the choice set to the optimal choice. Panel B shows that the utility loss from being $2d$ units away from the optimum equals 4δ . This is why the utility cost of ignoring the price change equals 4δ at the upper bound $\varepsilon_U(\hat{\varepsilon} = 0)$.

FIGURE 8
Tax Reform Act of 1986



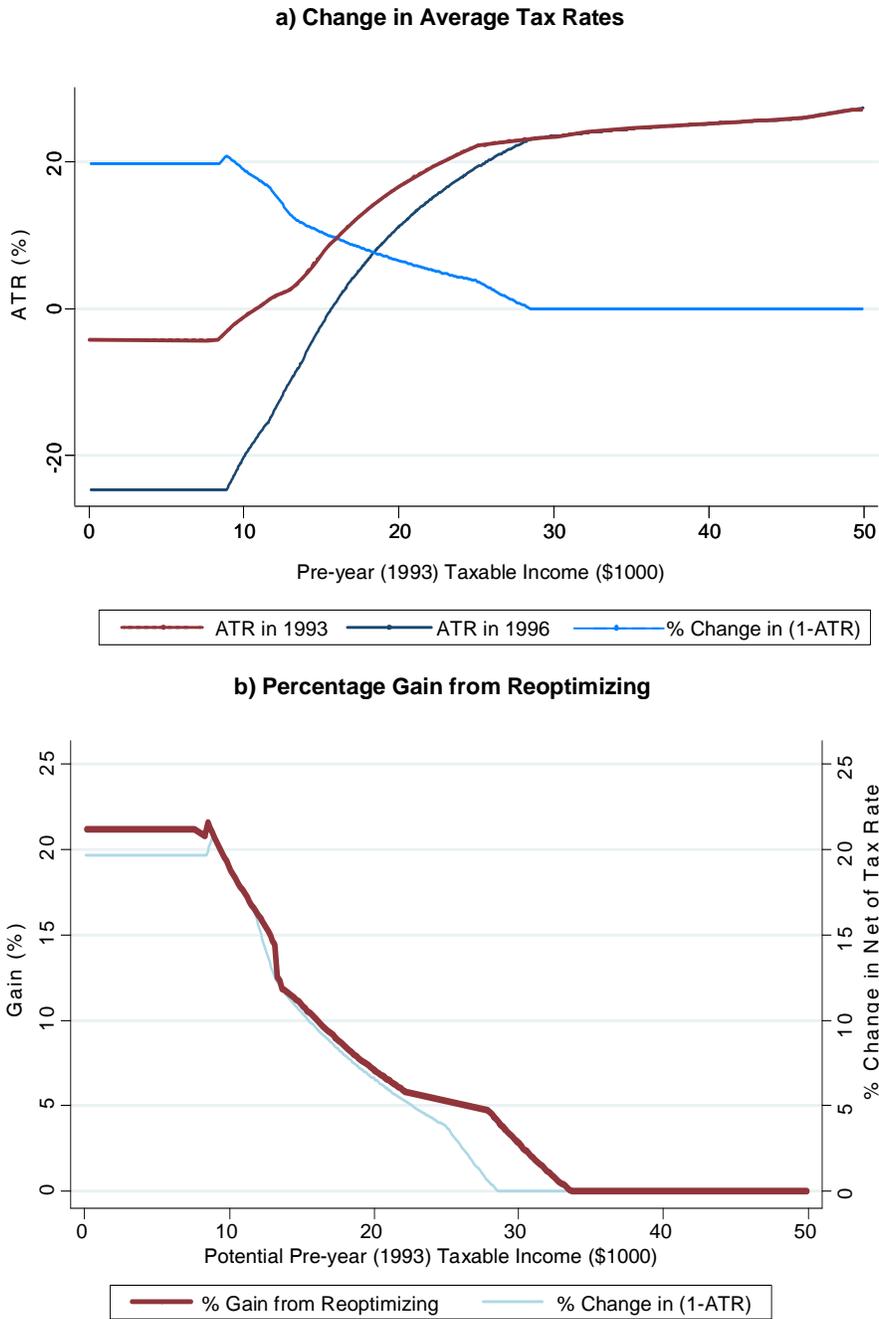
NOTE—These figures are based on the Tax Reform Act of 1986. Panel A shows changes in marginal tax rates by taxable income levels in 1985. Panel B plots the dollar gain from reoptimizing relative to TRA86 by income level with $\varepsilon = 1$. Panel C plots the percentage gain ($\Delta u\%$) from reoptimizing relative to TRA86 with $\varepsilon = 1$, which is defined as the dollar gain divided by consumption if the agent does not reoptimize. Panel D depicts the earnings change ($z^*(t_1) - z^*(t_0)$) required to reoptimize relative to the tax change. In Panels B-D, the dashed blue line (right y axis) replicates the change in the net-of-tax rate (1-MTR) shown in Panel A.

FIGURE 9
Clinton Earned Income Tax Credit Expansion



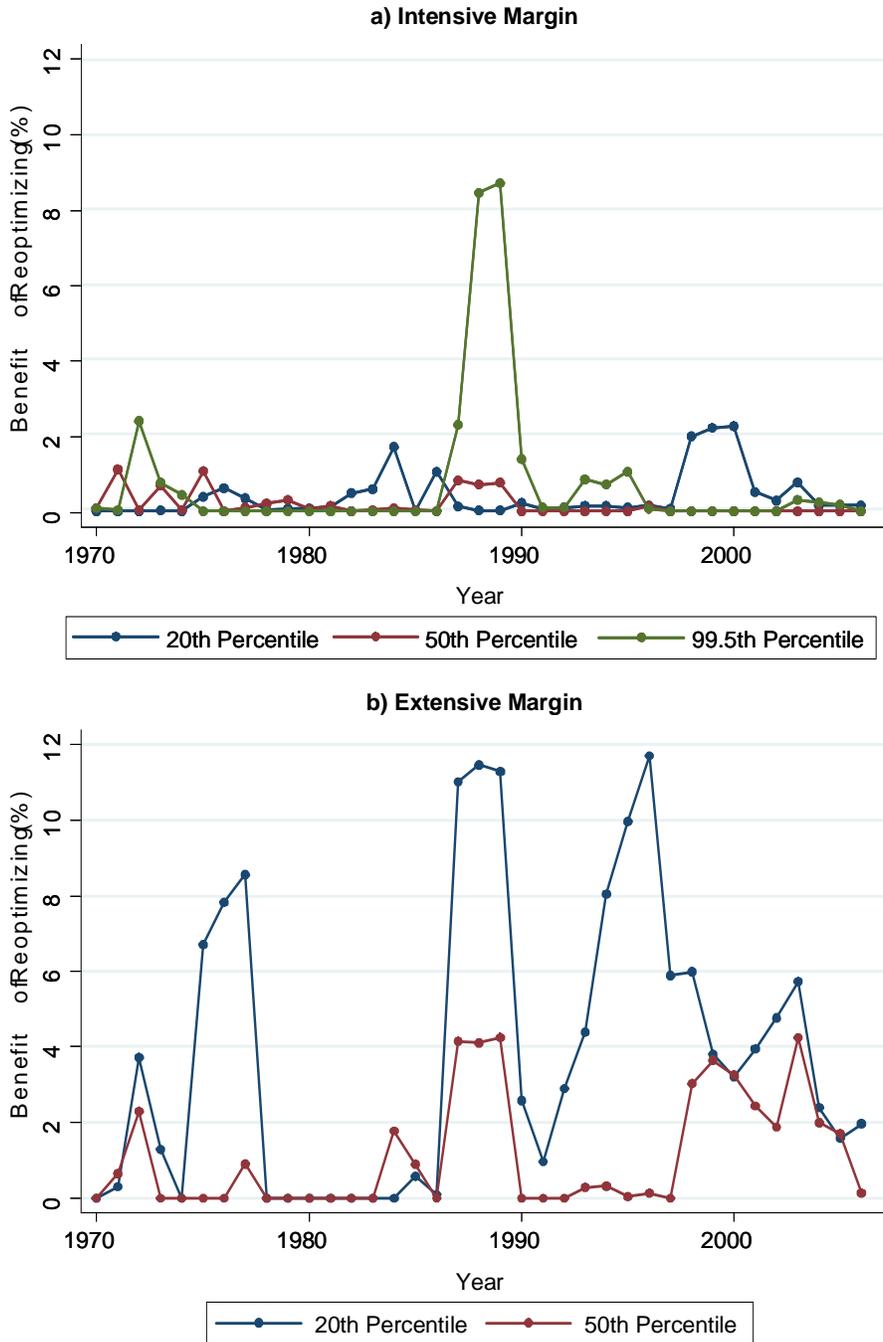
NOTE—These figures are based on the Clinton EITC Expansion enacted between 1993 and 1996. Panel A shows the changes in marginal tax rates by income during this period. Panel B plots the percentage utility gain ($\Delta u_{\%}$) from reoptimizing relative to EITC expansion by income level with $\varepsilon = 1$.

FIGURE 10
Clinton EITC Expansion: Extensive Margin



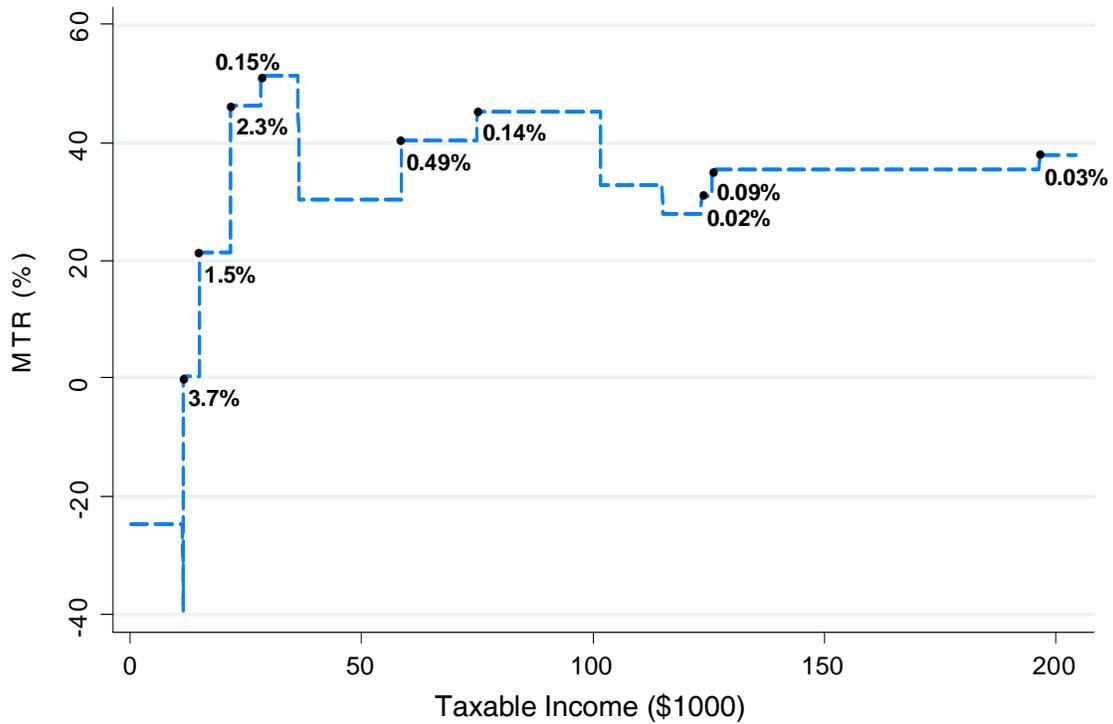
NOTE—These figures are the extensive margin analogue to Figure 9. Panel A shows changes in average tax rates by income during the Clinton EITC Expansion. Panel B plots the percentage gain ($\Delta u_{ext,\%}$) from reoptimizing relative to EITC expansion on the extensive margin. The x axis shows the income that an agent would optimally earn were he to work in 1993. For each such agent, I assume a fixed cost that would make him indifferent between working and not working in 1993. I compute the dollar gain from working and choosing the optimal labor supply in 1996 relative to staying out of the labor force. Finally, I define the percentage gain as the dollar gain from reoptimizing divided by the agent’s level of consumption when working in 1996.

FIGURE 11
 Percentage Gain from Reoptimizing in Response to Tax Changes by Year



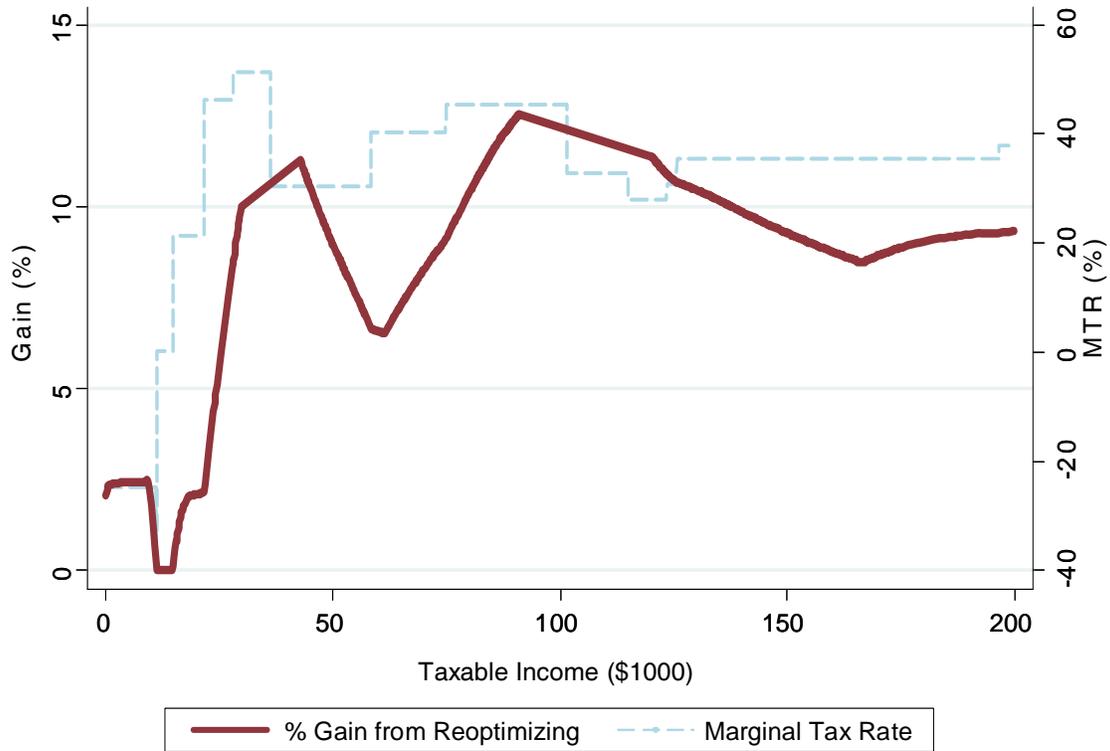
NOTE—These figures plot the percentage utility gain from reoptimizing in response to changes in taxes over three-year periods from the 1970s to 2000s for selected percentiles of the income distribution. In each year y , the point that is plotted shows the utility loss from selecting the optimal labor supply according to the tax system in year $y - 3$ instead of year y . Panel A depicts the percentage gain for the intensive margin ($\Delta u_{\%}$) while panel B depicts the percentage gain for the extensive margin ($\Delta u_{\text{ext},\%}$). The calculations assume $\varepsilon = 1$. See notes to Figures 8 and 10 for definitions of these percentage gain measures. The raw data underlying these figures are listed in Table 1.

FIGURE 12
Gains from Bunching at Kink Points in 2006 Tax Schedule



NOTE—This figure plots the 2006 marginal tax rate schedule in the U.S. The numbers near each convex kink shows the percentage gain ($\Delta u_{\%}$) from locating at that kink relative to locating at the optimum under the assumption that the tax rate in the previous bracket continues into next bracket. Each value is a mean of the percentage gains over all individuals whose wage rates would make it optimal for them to locate at that kink. The calculations assume $\varepsilon = 1$. The first two kinks (3.7% and 1.5%) correspond to the end of the phase-in and start of the phase-out regions of the EITC.

FIGURE 13
 Percentage Gain from Optimizing Relative to 2006 Tax Code: Intensive Margin



NOTE—This figure plots the percentage gain ($\Delta u\%$) from optimizing relative to 2006 tax code (versus choosing labor supply under the assumption that there are no income taxes) on the intensive margin with $\varepsilon = 1$.

FIGURE 14
Bounds on the Taxable Income Elasticity

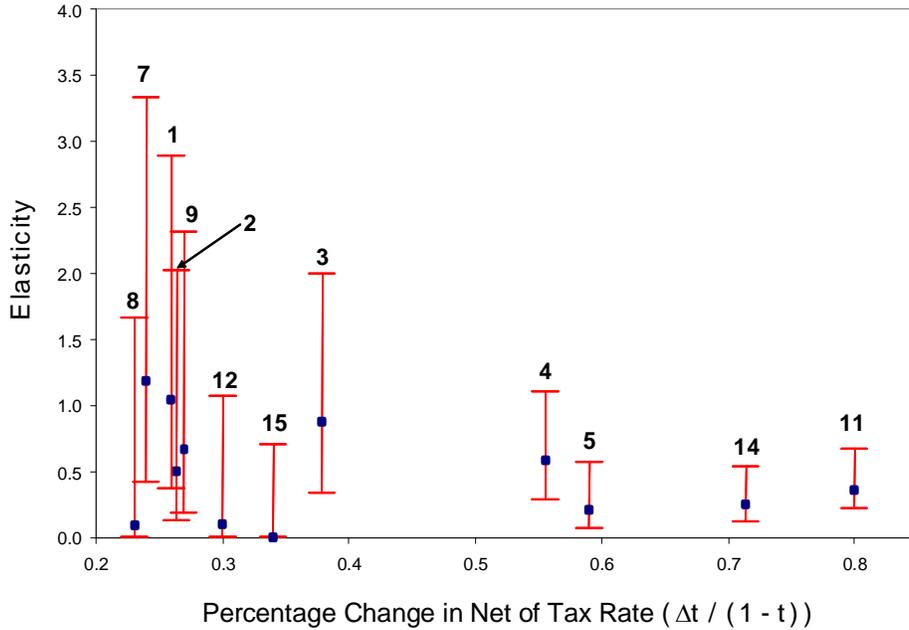


TABLE 2
Bounds on the Taxable Income Elasticity

Study	$\hat{\varepsilon}$	$\Delta t/(1-t)$	ε_L	ε_H	Source
(1)	(2)	(3)	(4)	(5)	(6)
1) Feldstein (1995)	1.04	26.0%	0.37	2.89	Table 2 High vs. Medium
2) Auten and Carroll (1997)	0.66	26.0%	0.19	2.32	Table 2, Col 2. $Dt/(1-t)$ from Feldstein (1995)
3) Goolsbee (1999)	1.00	36.5%	0.47	2.14	1985-1989, Table 2a,b
4) Goolsbee (1999)	0.58	55.6%	0.30	1.12	1922-1926, Table 4b,c
5) Goolsbee (1999)	0.21	59.0%	0.08	0.57	1931-1935, Table 5b,c
6) Gruber and Saez (2002)	0.40	5.8%	0.01	24.9	$\Delta t/(1-t)$ computed as mean 3-year change in NTR for median taxpayer
7) Prescott (2004)	1.18	24.0%	0.42	3.34	Table 2; U.K. and U.S. treated as control groups
8) Saez (2004)	0.09	23.1%	0.00	1.67	Top 10-5%, pg. 56 col. 7; elast from Table 4B
9) Saez (2004)	0.50	26.4%	0.12	2.02	Top 1%, pg. 56 col. 3; elast. From Table 2C, Col. 3
10) Kopczuk (2005)	0.25	5.8%	0.00	24.6	Page 22: taxable income elast. pre TRA86; $\Delta t/(1-t)$ as in Gruber-Saez
11) Davis and Henrekson (2005)	0.40	80.0%	0.23	0.69	Figure 1
12) Chetty et al. (2009)	0.10	30.0%	0.01	1.08	Figure 4, Figure 20: Top Kink
13) Chetty et al. (2009)	0.00	10.0%	0.00	8.00	Figure 4, Figure 20: Middle Kink
14) Gelber (2009)	0.25	71.4%	0.12	0.54	Table 1 (highest brackets), Table 3 Col. 2
15) Saez (2009)	0.00	34.0%	0.00	0.69	$\Delta t/(1-t)$ from first EITC Kink; 0 response for wage earners
Unified Bounds			0.47	0.54	

NOTE—The figure plots upper and lower bounds (shown in brackets) for selected elasticity estimates from studies in the taxable income elasticity literature. The bounds assume model uncertainty due to optimization frictions of $\delta = 1\%$. The blue squares show the point estimate of each study. The x axis in the figure is the percent change in the net of tax rate ($\Delta t/(1 - t)$) used for identification in each study. The numbers shown above each interval correspond to the numbers of the papers listed in Table 2.