Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply*

Raj Chetty

Harvard University and NBER

July 2010

Abstract

I derive bounds on price elasticities in a dynamic model that is mis-specified due to optimization frictions such as adjustment costs or inattention. The bounds are a function of the observed effect of a price change on demand, the size of the price change, and the degree of frictions. I measure the degree of frictions by the utility losses agents tolerate to make choices that deviate from the frictionless optimum. I apply these bounds to the literature on taxation and labor supply, allowing for frictions of 1% of consumption in choosing labor supply. Such small frictions reconcile the difference between micro and macro elasticities, extensive and intensive margin elasticities, and several other disparate findings. Pooling estimates from twenty existing studies yields bounds on the intensive margin Hicksian labor supply elasticity of $(0.47, 0.54)$.

*E-mail: chetty@fas.harvard.edu. I would like to thank Richard Blundell, David Card, Christopher Ferrall, John Friedman, Yuriy Gorodnichenko, James Hines, Caroline Hoxby, Guido Imbens, Patrick Kline, Adam Looney, Day Manoli, Ariel Pakes, Valerie Ramey, Emmanuel Saez, Laszlo Sandor, Harald Uhlig, and numerous seminar participants for very helpful comments. Gregory Bruich, Jane Choi, Nate Hilger, Tord Krogh, Zachary Liscow, Keli Liu, and Danny Yagan provided outstanding research assistance. Funding from NSF Grant SES 0645396 and the Sloan Foundation is gratefully acknowledged. The code used to produce the results in this paper is available on the author's website at www.economics.harvard.edu/faculty/chetty.

1 Introduction

The identification of structural parameters of stylized models is one of the central tasks of applied economics. Unfortunately, most models omit various frictions that make agents deviate systematically from their theoretical predictions. For instance, canonical models of labor supply or consumption behavior do not permit adjustment costs, inattentive agents, or status quo biases. How can structural parameters be identified when agents face such optimization frictions?

One natural solution is to estimate the structural parameters of a model that incorporates the frictions. This approach has two limitations in practice. First, it is difficult to incorporate all frictions in a tractable model. Second, estimating even simple dynamic models with frictions, such as Ss adjustment, requires strong econometric assumptions and is computationally challenging (Attanasio 2000). Motivated by these limitations, I propose an alternative solution in this paper: bounding structural preference parameters without identifying how frictions affect behavior.

I analyze a standard dynamic lifecycle model in which the effect of income-compensated (Hicksian) price changes on demand is determined by a structural parameter of utility ε. I introduce optimization frictions into this nominal model through an error term in the demand function whose conditional expectation is unknown. These optimization errors generate differences between mean observed demand and the mean optimal demand predicted by the frictionless model. Because the optimization errors need not be orthogonal to the price, the observed Hicksian elasticity $\widehat{\varepsilon}$ estimated from demand responses to a price change will generally differ from the structural elasticity parameter ε. Intuitively, the observed elasticity $\widehat{\varepsilon}$ confounds preferences (ε) with the effect of the frictions (e.g. adjustment costs). For example, agents may under-react to a price increase in the short-run because of adjustment costs.

This paper seeks to identify ε from estimates of $\widehat{\varepsilon}$. I focus on identifying ε because it is important for both positive and normative analysis. Long run effects of price changes are determined purely by ε in many models because the effect of frictions diminishes over time. Moreover, the recovery of preference parameters is essential for welfare analysis. I derive bounds on ε from observations of $\widehat{\varepsilon}$ by assuming that agents choose points near the frictionless optimum. Specifically, I allow agents to deviate arbitrarily from the nominal model's prediction as long as the expected utility cost of doing so is less than δ percent of expenditure.[1] This property is satisfied by standard dynamic models with adjustment costs, where agents remain on average within some utility threshold of their optimum. In the case of other frictions such as inattention or status quo biases, this restriction requires that agents respond to incentives that are sufficiently important.

Given an exogenously specified value of δ, the support of agent's optimization errors is bounded. This bounded support condition in turn produces bounds on the values of ε consistent with an observed elasticity $\widehat{\varepsilon}$. I derive a closed-form representation for bounds on the Hicksian price elasticity ε as a function of the observed Hicksian elasticity $\widehat{\varepsilon}$, the degree of frictions δ, and the size of the price change used for identification $\Delta \log p$. The bounds have several properties that shed light on what can be learned from reduced-form elasticity estimates in an environment with frictions. The bounds shrink at a quadratic rate with $\Delta \log p$ – studies identified from larger price changes contain much more information about ε. If $\widehat{\varepsilon} > 0$, the lower bound on the structural elasticity ε is strictly positive. If $\widehat{\varepsilon} = 0$, 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, analogous to power calculations used to evaluate statistical precision.

One can obtain tighter bounds on ε by calculating the least upper bound and the largest lower bound implied by a set of observed elasticities. The sensitivity of estimates of ε to frictions can be evaluated by computing these unified bounds as a function of δ. The smallest level of frictions $\delta_{\min}$ that reconciles a set of observed elasticities provides a quantitative measure of the "economic significance" of differences in estimates across studies. If $\delta_{\min}$ is small, the differences are not economically significant in that they can be explained by allowing for small frictions without making changes in the model's key assumptions. When $\delta = \delta_{\min}$, ε is point identified and can be expressed as a weighted average of observed elasticities, with greater weight placed on the elasticities identified using larger price changes. This result underscores a general lesson about identification with optimization frictions: pooling several small price changes – although useful in improving statistical precision – yields less information about the structural elasticity than studying a few large price changes.

[1] Permitting such deviations should be viewed as a minimal robustness requirement, as it is unlikely that any economic model predicts (average) behavior perfectly.

I apply these methods to investigate what can be learned about the structural labor supply elasticity from the literature on taxation and labor supply. The application consists of three components. First, I show that various disparate findings in the microeconometric labor supply literature can be synthesized by introducing small frictions into the simplest constant-elasticity neoclassical labor supply model. I calculate the utility costs of ignoring the tax reforms most commonly used for identification in microeconometric studies. I find that observed elasticities are large when the utility costs of failing to respond are large. For instance, the flow utility costs of ignoring the Tax Reform Act of 1986 (TRA86) are less than 2% of consumption for all except top income earners. Accordingly, empirical studies find changes in labor supply around TRA86 only for top incomes. In addition to explaining the heterogeneity in observed elasticities across income groups, frictions can also explain why we observe larger elasticities on the extensive margin than the intensive margin, limited bunching at kink points in non-linear budget set models, sharp responses to notches in budget sets, and converging elasticities across primary and secondary earners.

Second, I calculate bounds on the Hicksian labor supply elasticity using estimates from twenty studies that span various methodologies, ranging from microeconometric quasi-experiments to macroeconomic calibrations. Even though the observed Hicksian elasticity estimates vary widely, all twenty estimates are consistent with a single structural elasticity ε if one permits frictions of $\delta > \delta_{\min} = 0.8\%$ of consumption in choosing labor supply. Pooling the twenty studies yields bounds on the Hicksian labor supply elasticity of $(0.47, 0.54)$ when $\delta = 1\%$.

Finally, I show that frictions can explain the longstanding puzzle of why microeconometric elasticity estimates are much smaller than macro estimates. Micro studies estimate short-run responses to tax changes whereas macro studies estimate long-run responses. Frictions can attenuate short-run elasticities: for instance, if individuals draw low job switching costs once every ten years, micro estimates of changes over a one year horizon could be an order of magnitude smaller than long-run macro estimates. The bounds derived here show quantitatively that frictions of just 1% of consumption can generate the substantial observed differences between micro and macro elasticities. This reconciliation complements the well known explanation proposed by Richard Rogerson and others (e.g. Rogerson 1988, Rogerson 2006, Rogerson and Wallenius 2009).[2] These papers argue that macro elasticities are larger

[2] Others have proposed that differences in regulations (Alesina, Glaeser, Sacerdote 2005) or social insurance

because they incorporate both extensive and intensive margin labor supply responses, whereas micro studies focus on the intensive margin. While this insight clearly explains part of the puzzle, much of the difference in labor supply across countries with different tax regimes is driven by hours worked conditional on employment. That is, intensive margin macro elasticity estimates are much larger than their microeconometric counterparts. But micro and macro estimates of the extensive margin elasticity are actually quite similar. I show that optimization frictions generate precisely this pattern. Observed short-run extensive margin elasticities remain close to structural elasticities even with small frictions because the utility costs of failing to reoptimize are first-order on the extensive margin. In contrast, small frictions can attenuate observed short-run intensive margin responses substantially because the costs of failing to reoptimize on the intensive margin are second-order.

This paper builds upon and relates to the partial identification, near rationality, robust control, and durable goods literatures. The econometrics literature on partial or set identification considers problems such as missing data or imperfect instruments, where point identification is infeasible barring strong assumptions (Manski 2007, Nevo and Rosen 2008). The present paper uses set identification to estimate structural parameters with model mis-specification. Papers in the partial identification literature typically derive bounds by making assumptions such as stochastic dominance of wage distributions for labor force participants relative to non-participants or exploiting necessary conditions for Nash equilibria (Blundell et al. 2007, Pakes et al. 2007). Here, I derive bounds by assuming that agents are "near rational," as in the menu cost and near rationality literature in macroeconomics (Akerlof and Yellen 1985, Mankiw 1985, Cochrane 1989).[3] The focus on 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 I consider identification of the nominal model's parameters in the same setting.[4] Finally, the

systems (Ljungvist and Sargent 2006) between countries with low and high tax rates bias macro elasticity estimates upward . These explanations complement the one offered here, and could in particular help explain the largest values obtained by macro studies that do not account for such factors (e.g. Prescott 2004).

[3]Cochrane (1989) shows how to assess whether a given model is consistent with data if one permits small frictions. I develop a method of identifying the set of structural parameters consistent with the data with frictions. The difference between the two papers is analogous to the difference between hypothesis testing and confidence intervals in statistics.

[4]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.

bounds provide an alternative method of estimating preferences or production functions in models with adjustment costs. The approach proposed here requires fewer assumptions than existing methods of identifying such models (e.g. Eberly 1994, Attanasio 2000) because the observed elasticities it uses as inputs can be estimated using quasi-experimental techniques. However, it does not permit as rich an analysis of short-run counterfactuals because it only partially identifies the model's parameters.

The paper is organized as follows. The next section sets up a dynamic model with frictions. The bounds on price elasticities are derived in Section 3. Section 4 presents the application to labor supply and taxation. Section 5 concludes.

2 Demand Models with Frictions

Consider a dynamic model with N individuals who have heterogeneous tastes over two goods, x and y. The price of x in period t is p_t and the price of y is fixed at 1. Individual i has wealth Z_i and time-separable utility

$$\sum_{t=0}^{T} v_{i,t}(x_t, y_t) \tag{1}$$

To simplify the exposition, I focus on the following specification of flow utility in the text:

$$\begin{aligned} v_{i,t}(x_t, y_t) &= y_t + a_{i,t}\frac{x_t^{1-1/\varepsilon}}{1-1/\varepsilon} \text{ if } \varepsilon \neq 1 \\ v_{i,t}(x_t, y_t) &= y_t + a_{i,t}\log x_t \text{ if } \varepsilon = 1 \end{aligned} \tag{2}$$

This quasilinear utility specification has two convenient properties: (1) it is a money metric and (2) it permits heterogeneity in the levels of demand across agents and periods but generates a constant price elasticity of demand ε. I extend the main results to the general case where $v_{i,t}(x_t, y_t)$ is unrestricted in the appendix.

The utility function in (2) yields the following demand function for good x:

$$x_{i,t}^*(p_t) = (\frac{a_{i,t}}{p_t})^{\varepsilon} \tag{3}$$

Let $\alpha = \sum_i \sum_t \log x_{i,t}^*(p_t = 1)/N(T+1)$ denote the mean log demand in the population when $p_t = 1$ and $\nu_{i,t} = \log x_{i,t}^*(p_t = 1) - \alpha$ denote the deviation of individual i in period t from the mean. Then we can write agent i's demand function as

$$\log x_{i,t}^*(p_t) = \alpha - \varepsilon \log p_t + \nu_{i,t}$$

Our objective is to identify ε, the structural preference parameter that controls the price elasticity of demand. More compactly, I shall refer to ε as the "structural elasticity." Under the quasilinear utility specification in (2), the Hicksian, Marshallian, and Frisch elasticities are all equal to ε because $x^*_{i,t}$ depends only upon p_t and not on prices in other periods or wealth (Browning 2005). For more general utilities, the three elasticities differ. I show in the appendix that the bounds derived below apply to the Hicksian elasticity in the general case.

Consider identification of ε using a price change from p_A in period A to $p_B \neq p_A$ in period B.[5] The standard assumption made to identify ε from such variation is the following orthogonality condition on the error term $v_{i,t}$ in the demand function.

A1 Taste shocks $v_{i,t}$ are orthogonal to the identifying price variation: $\mathbb{E}\nu_{i,A} = \mathbb{E}\nu_{i,B}$.

Assumption A1 permits arbitrary changes in tastes $a_{i,t}$ across periods for any given agent, but requires that the aggregate distribution of tastes is identical in periods A and B. Under A1, which I assume holds throughout this paper,

$$\varepsilon = -\frac{\mathbb{E}\log x^*_{i,B}(p_B) - \mathbb{E}\log x^*_{i,A}(p_A)}{\log p_B - \log p_A} \tag{4}$$

Equation (4) shows that the observed demand response to a price change point identifies ε in the frictionless model in (1). I refer to the model in (1) as the "nominal" model, following the robust control literature. I now explore how the link between ε and the observed demand response in the nominal model is affected by optimization frictions using two examples.

Frictions - Example 1: Adjustment Costs. In many markets, an agent must pay a search or switching cost to change his level of consumption. Suppose that the path of the adjustment cost $k_{i,t}$ evolves deterministically and that tastes $a_{i,t}$ and prices p_t follow arbitrary stochastic processes. Let $\mathbb{E}_t$ denote the conditional expectation operator over prices and tastes given information available in period t. In this model, agent i chooses consumption $x_{i,t}$ in period t by solving:

$$\max_{x_s} \mathbb{E}_t \sum_{s=t}^{T} [a_{i,s}\frac{x_s^{1-1/\varepsilon}}{1-1/\varepsilon} - p_s x_s - k_{i,s}\cdot(x_s \neq x_{s-1})]. \tag{5}$$

Observed demand in this model, $x_{i,t}$, differs from the frictionless optimum $x^*_{i,t}$. Let the

[5] The analysis is unaffected if the identifying price variation is cross-sectional, provided that the variation in p_t is orthogonal to the variation in tastes across individuals $\nu_{i,t}$.

observed elasticity estimated from a price change between periods A and B be denoted by

$$\widehat{\varepsilon} = -\frac{\mathbb{E}\log x_{i,B}(p_B) - \mathbb{E}\log x_{i,A}(p_A)}{\log p_B - \log p_A} \tag{6}$$

In this model, $\widehat{\varepsilon}$ no longer identifies the structural elasticity ε. The observed elasticity $\widehat{\varepsilon}$ may be smaller or larger than ε depending upon the evolution of prices, adjustment costs, and tastes. For instance, suppose $A = 0$, $B = T = 1$, and agents do not anticipate any price or taste changes in period 1, so that the change from p_A to p_B is a surprise. Then $\widehat{\varepsilon} = f \cdot \varepsilon$, where f denotes the fraction of agents whose flow utility gain from choosing $x^*(p_B)$ instead of $x^*(p_A)$ exceeds their adjustment cost $k_{i,B}$. In this example, the observed elasticity $\widehat{\varepsilon} \leq \varepsilon$ because agents start at their optima ($x_{i,A} = x^*_{i,A}$). In contrast, suppose a history of small price increases before period A has led to $x_{i,A}(p_A) > x^*_{i,A}(p_A)$ for all agents. Here, a further price increase to $p_B > p_A$ could induce a large observed demand response, leading to $\widehat{\varepsilon} > \varepsilon$.

Although $\widehat{\varepsilon}$ generally differs from ε, the structural elasticity ε plays a central role in determining the long-run effects of changes in prices. For example, the effect of a permanent price change starting in period 0 is determined purely by ε. As another example, suppose all agents have zero adjustment costs from time to time ($k_{i,t} = 0$) and that mean observed demand equals mean optimal demand in period A ($\mathbb{E}\log x_{i,A} = \mathbb{E}\log x^*_{i,A}$). Then the observed effect of a permanent, unanticipated price change in period $A + 1$ approaches ε as B grows large irrespective of the path of tastes and adjustment costs. Intuitively, adjustment costs affect observed elasticities primarily in the short-run, as agents may delay adjustment until periods when they face low switching costs. Hence, studies that estimate long-run or steady state responses – such as macroeconomic studies that make cross-country comparisons – are more likely to identify ε. Unfortunately, such studies are also the most prone to omitted variable problems because of the lack of good counterfactuals. As a result, most modern microeconometric studies focus on estimating short-run responses.

How can one identify the structural elasticity ε from observed short-run responses to price changes when agents face adjustment costs? One natural approach is to estimate the model that incorporates adjustment costs in (5) instead of the frictionless model in (1), as in Eberly (1994) or Caballero, Engel and Haltiwanger (1995). Although refining the model to incorporate frictions is an ideal solution, the identification of dynamic models with adjustment costs runs into serious limitations in practice. Strong parametric assumptions on the evolution of adjustment costs $k_{i,t}$ and the stochastic processes governing tastes $a_{i,t}$ and prices p_t are required to derive tractable decision rules such as Ss behavior (Grossman and Laroque 1990). Even given these strong assumptions, estimating the model requires considerable data and is numerically demanding (Attanasio 2000).

Frictions - Example 2: Price Misperceptions. A growing body of evidence indicates that individuals misperceive prices in many domains (DellaVigna 2009). For instance, individuals are inattentive to tax rates and confuse average with marginal tax rates (Fujii and Hawley 1988, Chetty, Looney, and Kroft 2009, Chetty and Saez 2009). To model this class of deviations from (1), let $\widetilde{p}_{i,t}(p_t)$ denote agent i's perceived price as a function of the true price in period t. The agent chooses $x_{i,t}$ by solving

$$\max \sum_{s=t}^{T} [a_{i,s} \frac{x_s^{1-1/\varepsilon}}{1-1/\varepsilon} - \widetilde{p}_{i,s}(p_s) \cdot x_{i,s}] \tag{7}$$

The resulting demand function is

$$\log x_{i,t}(p_t) = \alpha - \varepsilon \log \widetilde{p}_{i,t}(p_t) + \nu_{i,t}$$

and the observed elasticity is

$$\widehat{\varepsilon} = \varepsilon \frac{\mathbb{E} \log \widetilde{p}_{i,B}(p_B) - \mathbb{E} \log \widetilde{p}_{i,A}(p_A)}{\log p_B - \log p_A}$$

Again, the observed elasticity $\widehat{\varepsilon}$ confounds the structural elasticity of interest ε with other parameters, in this case the effect of the price change on mean perceived prices. But if perceptions converge to the truth over time, long run responses are determined solely by ε.

Identifying ε by estimating the structural parameters of (7) requires specification of a theory of perceptions $\widetilde{p}_{i,t}(p_t)$. Unfortunately, there is little consensus on what determines perceptions. Perceptions may be determined by rational information acquisition (Stigler 1961, Sims 2003), advertisements (Nelson 1974), or persuasion by other agents (Mullainathan, Shleifer, and Schwartzstein 2008, DellaVigna and Gentzkow 2009). Each of these theories requires estimation of a different set of structural parameters, ranging from cost of acquiring information to the stochastic processes that govern triggers of attention.

Optimization Frictions and Partial Identification. The two examples above suggest that it is very challenging to accurately model and identify all the frictions that may cause agents to deviate systematically from standard stylized models. The challenge is compounded when one considers the array of other factors that may lead to deviations from the nominal model, such as status quo biases (Samuelson and Zeckhauser 1988) or satisficing behavior (Simon 1957). This problem motivates a less ambitious strategy: identifying ε without fully identifying the structure of optimization frictions. Identifying ε itself is useful (though not always sufficient) for both positive and normative analysis. As noted in the examples above, the effect of frictions diminishes in the long run under plausible conditions, making ε sufficient to predict long-run responses. For welfare analysis, it is essential to recover preferences, i.e. to identify ε. The structural elasticity ε and the observed elasticity $\widehat{\varepsilon}$ are together sufficient for approximate welfare calculations in many applications (Chetty, Looney, and Kroft 2009).[6]

The problem of identifying ε with unknown frictions can be viewed as a partial identification problem. Define agent i's "optimization error" as the log difference between his optimal demand as predicted under the nominal model in (3) and his observed demand:[7]

$$\phi_{i,t} = \log x_{i,t}(p_t) - \log x^*_{i,t}(p_t)$$

Then observed demand for agent i can be written as

$$\log x_{i,t}(p_t) = \alpha - \varepsilon \log p_t + \nu_{i,t} + \phi_{i,t} \tag{8}$$

Define $x_t(p_t) = [\prod_{i=1}^{N} x_{i,t}(p_t)]^{1/N}$ and $x^*_t(p_t) = [\prod_{i=1}^{N} x^*_{i,t}(p_t)]^{1/N}$ as the geometric means of observed and optimal demands. Mean observed (log) demand is

$$\begin{aligned}\log x_t(p_t) &= \mathbb{E}\log x_{i,t}(p_t) = \alpha - \varepsilon \log p_t + \mathbb{E}\phi_{i,t} \\ &= \log x^*_t(p_t) + \mathbb{E}\phi_{i,t}\end{aligned}$$

The optimization errors $\phi_{i,t}$ generated by frictions are not orthogonal to changes in prices. For example, in the adjustment cost model, mean observed demand may be at the optimum in period A ($\mathbb{E}\phi_{i,A} = 0$), but above the new optimum following a price increase in period B ($\mathbb{E}\phi_{i,B} > 0$). This is the fundamental difference between $\phi_{i,t}$ and the preference heterogeneity error $\nu_{i,t}$, which is plausibly orthogonal to certain types of price changes. Imposing the

[6] For questions that require full identification of the model's structure – such as predicting short-run responses to policy changes – it is still useful to start by identifying ε using as little structure as possible to obtain the most reliable estimates. One can then impose more structure to identify the model's remaining parameters.

[7] The optimization error is an error from the econometrician's perspective but not necessarily from the agent's perspective. In the adjustment cost model, the agent optimizes by choosing $x_{i,t}$ according to (5) instead of $x^*_{i,t}$.

orthogonality condition in A1 on $\nu_{i,t}$ but leaving $\mathbb{E}\phi_{i,t}$ unrestricted, the observed demand response to a price change is

$$\mathbb{E}\log x_{i,B} - \mathbb{E}\log x_{i,A} = -\varepsilon[\log p_B - \log p_A] + [\mathbb{E}\phi_{i,B} - \mathbb{E}\phi_{i,A}] \tag{9}$$

Without assumptions on $\phi_{i,t}$, ε is unidentified by the observed response because the $\mathbb{E}\phi_{i,B} - \mathbb{E}\phi_{i,A}$ is unknown. Intuitively, if one places no restrictions on perceptions or adjustment costs, an observed response to a price change can be reconciled with any structural price elasticity.

Restricting the Degree of Frictions. As in the partial identification literature, one can obtain bounds on ε by bounding the support of $\phi_{i,t}$ without making specific assumptions about $\mathbb{E}\phi_{i,t}$. The 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. I restrict the support of $\phi_{i,t}$ by requiring that agents make choices "near" the optimal choice under the nominal model. Let the maximum utility the agent can attain from periods t to T subject to the constraint that x_t is set at some level $\widetilde{x}_t$ be denoted by

$$u_{i,t}(\widetilde{x}_{i,t}) = \max_{x_s,y_s} \sum_{s=t}^{T} v_{i,t}(x_s, y_s) \text{ s.t. } [\sum_{s=t}^{T}(p_s x_s + y_s) = Z_t] \text{ and } [x_t = \widetilde{x}_{i,t}]$$

where $Z_t = Z - \sum_{s=0}^{t-1} p_s x_s + y_s$ denotes wealth remaining in period t. I impose the simple requirement that agents' deviations from the frictionless optimum cannot have too large of a utility cost (calculated under the nominal model):

$$u_{i,t}(x_{i,t}^*) - u_{i,t}(x_{i,t}) = u_{i,t}(x_{i,t}^*) - u_{i,t}(x_{i,t}^* e^{\phi_{i,t}}) < D_{i,t}$$

where $D_{i,t}$ is an exogenously specified threshold. It is convenient to normalize the dollar value $D_{i,t}$ by the optimal flow expenditure on good x and define $\delta_{i,t} = D_{i,t}/p_t x_{i,t}^*$. The parameter $\delta_{i,t}$ measures the utility loss that agent i tolerates as a percentage of his expenditure on good x in period t.

To obtain bounds on mean observed demand in the population, $\log x_t$, I restrict the average value of $\delta_{i,t}$ across agents in any period t to be below an exogenously specified threshold δ. For instance, $\delta = 1\%$ permits deviations from optimal demand with an average utility cost of up to 1% of expenditure in each period. Formally, the bounded support condition I impose is that the optimization errors in every period t, $\phi_t = \{\phi_{1,t}, ..., \phi_{N,t}\}$, lie within the set

$$\Phi_\delta = \{\phi_t : u_{i,t}(x_{i,t}^*) - u_{i,t}(x_{i,t}^* e^{\phi_{i,t}}) \leq \delta_{i,t} p_t x_{i,t}^* \ \forall i \text{ with } \textstyle\sum_i \delta_{i,t}/N \leq \delta\} \tag{10}$$

The parameter δ measures the degree of optimization frictions that one permits, and will vary across applications. A lower value of δ is appropriate when examining responses over longer horizons if agents converge to their new optima over time. In the context of inattention, a higher δ may be suitable for price changes that are less salient.

I refer to the models that generate optimization errors $\phi_{i,t} \in \Phi_\delta$ as a "δ class of models" around the nominal model. A δ class of models contains many models of behavior. The adjustment cost model in (5) lies in the δ class of models around (3) if the average adjustment cost as a percentage of consumption $\frac{1}{N}\sum_i k_{i,t}/p_t x^*_{i,t} \leq \delta/2$ in all periods t. The model of price misperceptions in (7) lies in the δ class of models around (3) if the expected utility losses due to misperceptions are less than δ – that is, if perceptions are not too inaccurate on average.

A δ class of models is more general than any of these specific models of behavior because it does not generate a 1-1 mapping from prices to observed demand. A single model maps a price path $p = (p_1, ..., p_T)$ to a single level of mean demand x_t in each period. A δ class of models instead maps a price path p to a set of mean demand levels in each period. Let

$$X_t(p_t, \delta) = \{x_t : \phi_t \in \Phi_\delta\}$$

denote the set of mean demands predicted by a δ class of models. When there is no heterogeneity in preferences and frictions across agents and over time ($\delta_{i,t} = \delta$, $a_{i,t} = a$, $Z_i = Z \Rightarrow u_{i,t} = u$ for all i, t), the choice set can simply by written as the set of demands that yield utility within δ units of the optimum $x^*(p_t) = x^*_{i,t}(p_t)$:

$$X(p_t, \delta) = \{x : u(x^*(p_t)) - u(x) \leq \delta p_t x^*(p_t)\} \tag{11}$$

I focus on this special case with homogeneous preferences and frictions (omitting time and agent subscripts) when illustrating the key results graphically. However, all the results below are established formally in the general case with heterogeneity. Figure 1 illustrates the construction of the choice set $X(p_t, \delta)$ when $\delta_{i,t} = \delta = 1\%$ and $a_{i,t} = a = e^{3.5}$. The figure plots $u(x_t)$ when $\varepsilon = 1$, $\log p_t = 1$, and $Z/(T+1) = 100$. The set of choices that yield utility within $\delta = 1\%$ of the optimum, $X(p_t, \delta) = [10.2, 14]$, is shown by the red interval on the x axis.

Now consider how a price increase from p_A to p_B affects mean observed demand in a δ class of models. Figure 2a illustrates the choice sets at the two prices, $X(p_A, \delta)$ and $X(p_B, \delta)$, with the same homogeneous preference parameters as in Figure 1. The structural elasticity

ε controls the movement of the choice sets with the price p, as illustrated by the dashed blue line. The black lines illustrate that various mean demand changes $[\log x_B(p_B) - \log x_A(p_A)]$ may be observed for a given value of ε. Each black line is generated by a different model. For instance, the flat black line could be generated by a model with status quo bias or satisficing consumers. Over-reaction could be observed in a model with adjustment costs, as discussed above. One may even observe an *increase* in demand, for instance if the price increase reflects a change in tax policy that raises tax rates but makes taxes less salient.

These examples show that optimization frictions destroy the 1-1 map from the observed response to the structural elasticity in (4). Let the range of structural elasticities consistent with a given observed elasticity $\widehat{\varepsilon}$ in a δ class of models be denoted by $(\varepsilon_L(\widehat{\varepsilon}, \delta), \varepsilon_U(\widehat{\varepsilon}, \delta))$. The objective of this paper is to characterize ε_L and ε_U in terms of empirically estimable parameters. The bounds $(\varepsilon_L, \varepsilon_U)$ measure the uncertainty in the structural elasticity due to potential 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 ε rather than other measures of dispersion because we typically lack a prior distribution over the models within the δ class. A natural 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 the bounds proceeds in three steps. I first characterize $X_t(p_t, \delta)$, the set of mean observed demands at a price p_t for a given value of ε. I then identify the set of structural elasticities ε consistent with an observed elasticity $\widehat{\varepsilon}$. Finally, I present methods of combining multiple observed elasticities to obtain more informative bounds on ε. Throughout, I provide results only on identification of bounds on ε. I briefly discuss below how inference about the bounds in finite samples can be handled using the techniques proposed in the recent partial identification literature.

3.1 Bounds on the Choice Set

The following lemma analytically characterizes $X_t(p, \delta)$ for small δ using a quadratic approximation to utility $u_{i,t}(x)$.

Lemma 1. For small δ, the set of mean observed demands is approximately

$$X_t(p_t, \delta) = \{x_t : |\log x_t - \log x_t^*| < [2\varepsilon\delta]^{1/2}\} \tag{12}$$

Proof. Taking a quadratic approximation to $u_{i,t}(x) = u_{i,t}(e^{\log x})$ around $\log x_{i,t}^*$ and exploiting the first-order condition under the nominal model $u_{i,t}'(x_{i,t}^*) = 0$ yields

$$u_{i,t}(x_{i,t}^*) - u_{i,t}(x) \simeq -\frac{1}{2}(x_{i,t}^*)^2 u_{i,t}''(x_{i,t}^*)(\log x - \log x_{i,t}^*)^2 \tag{13}$$

The definition of a δ class of models in (10) requires that

$$\begin{aligned} & u_{i,t}(x_{i,t}^*) - u_{i,t}(x_{i,t}) \leq \delta_{i,t} p_t x_{i,t}^* \\ \Rightarrow \quad & |\log x_{i,t} - \log x_{i,t}^*| \leq [-2\delta_{i,t} \frac{p_t}{x_{i,t}^*} \frac{1}{u_{i,t}''(x_{i,t}^*)}]^{1/2} \end{aligned} \tag{14}$$

With the quasilinear utility specification in (2), $u_{i,t}''(x_t) = \frac{\partial^2 v_{i,t}(x_t)}{\partial x_t^2}$ and the first order condition in the nominal model for $x_{i,t}$ is $\frac{\partial v_{i,t}}{\partial x}(x_{i,t}^*(p_t)) = p_t$. Implicitly differentiating this first order condition yields

$$u_{i,t}''(x_{i,t}^*) \frac{dx_{i,t}^*}{dp_t} = 1 \tag{15}$$

Substituting (15) into (14) gives the following restriction on demand for each agent:

$$|\log x_{i,t} - \log x_{i,t}^*| \leq [2\varepsilon\delta_{i,t}]^{1/2}$$

To derive bounds on mean observed demand x_t, use Jensen's inequality to obtain:

$$|\log x_t - \log x_t^*| = |\mathbb{E}\log x_{i,t} - \mathbb{E}\log x_{i,t}^*| \leq \mathbb{E}[2\varepsilon\delta_{i,t}]^{1/2} \leq [2\varepsilon\mathbb{E}\delta_{i,t}]^{1/2} = [2\varepsilon\delta]^{1/2}$$

It follows that mean observed demand x_t in a δ class of models satisfies

$$|\log x_t - \log x_t^*| \leq [2\varepsilon\delta]^{1/2}$$

Note that the approximation error in this equation vanishes as $\delta \to 0$ because the remainder of the Taylor approximation in (13) involves higher-order terms.

Lemma 1 captures three intuitions. First, the width of the choice set, which is $2[2\varepsilon\delta]^{1/2}$ log units, shrinks at a square-root rate as δ goes to zero. The choice set therefore becomes large relative to the degree of frictions as $\delta \to 0$:

$$\lim_{\delta \to 0} \frac{\log \max(X_t(p_t, \delta)) - \log \min(X_t(p_t, \delta))}{\delta} = \infty$$

This result implies that even small optimization frictions δ can generate substantial variation in observed behavior. For example, with a price elasticity of $\varepsilon = 1$ and $\delta = 1\%$, the choice set extends approximately +/-14% around $x^*(p_t)$, as illustrated in Figure 1. The root-δ shrinkage of the choice set is driven by the second-order losses of deviating from the maximum of a smooth function (Akerlof and Yellen 1985, Mankiw 1985).

Second, equation (14) shows that the width of the choice set is inversely related to the curvature of the objective function at the optimum, $u''_{i,t}(x^*_{i,t})$. More curved utilities generate a narrower choice set around the optimum because utility falls more sharply as one deviates from the optimum. A very useful property of the model is that $u''_{i,t}(x^*_{i,t})$ is pinned down by ε, the structural parameter of interest. Highly curved utilities generate small elasticities because the agent has a strong preference to locate near $x^*_{i,t}$. For example, suppose the demand for an essential medicine is perfectly price inelastic at a level $x^*_{i,t}$. 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 $x^*_{i,t}$ are infinitely large, the choice set $X_t(p_t, \delta)$ collapses to the singleton x^*_t for any δ when $\varepsilon = 0$, as illustrated in Figure 2b. The choice set expands as ε rises. This connection between ε and the curvature of utility is critical because it eliminates the need to estimate the additional parameter $u''_{i,t}(x^*_{i,t})$ when bounding ε.

Finally, the set of mean observed demands depends only upon the mean level of frictions δ, and not the distribution of $\delta_{i,t}$ in the population. Because each individual's choice set is proportional to $[\delta_{i,t}]^{1/2}$, the potential difference between mean observed and optimal demand is maximized when $\delta_{i,t} = \delta$ for all i, t. Given that we do not place any restrictions on the distribution of $\delta_{i,t}$, the worst case of $\delta_{i,t} = \delta$ determines the width of $X_t(p_t, \delta)$.

3.2 Bounds on the Structural Elasticity

Figure 3a depicts the largest structural elasticity ε that could have generated an observed elasticity $\widehat{\varepsilon}$. This elasticity ε_U generates the maximal shift in the choice sets consistent with the observed change in demand. When $\varepsilon = \varepsilon_U$, mean observed demand lies at the bottom of the choice set at price p_A ($\log x_A(p_A) = \log x^*_A(p_A) - (2\varepsilon\delta)^{1/2}$) and the top of the choice set at price p_B ($\log x_B(p_B) = \log x^*_B(p_B) + (2\varepsilon\delta)^{1/2}$). The upper bound ε_U therefore satisfies the condition

$$\widehat{\varepsilon} = -\frac{\log x_B(p_B) - \log x_A(p_A)}{\log(p_B) - \log(p_A)} = -\frac{\log x^*_B(p_B) - \log x^*_A(p_A) + 2(2\varepsilon\delta)^{1/2}}{\log(p_B) - \log(p_A)} = \varepsilon_U - 2\frac{(2\varepsilon_U\delta)^{1/2}}{\Delta \log p} \tag{16}$$

where $\Delta \log p = \log(p_B) - \log(p_A)$. Similarly, the lower bound structural elasticity ε_L consistent with $\widehat{\varepsilon}$, illustrated in Figure 3b, is defined by the equation

$$\widehat{\varepsilon} = \varepsilon_L + 2\frac{(2\varepsilon_L\delta)^{1/2}}{\Delta \log p} \tag{17}$$

The following proposition characterizes the solutions to (16) and (17).

Proposition 1. Under assumption A1, for small δ, the range of structural elasticities consistent with an observed elasticity $\widehat{\varepsilon}$ is approximately $(\varepsilon_L, \varepsilon_U)$ where

$$\varepsilon_L = \widehat{\varepsilon} + \frac{4\delta}{(\Delta \log p)^2}(1-\rho) \text{ and } \varepsilon_U = \widehat{\varepsilon} + \frac{4\delta}{(\Delta \log p)^2}(1+\rho) \tag{18}$$

$$\text{with } \rho = (1 + \frac{1}{2}\frac{\widehat{\varepsilon}}{\delta}(\Delta \log p)^2)^{1/2} \tag{19}$$

Proof. Equations (16) and (17) both reduce to the quadratic equation $(\widehat{\varepsilon} - \varepsilon)^2 = \frac{8\varepsilon\delta}{(\Delta \log p)^2}$. The upper and lower roots of this quadratic equation are the bounds.

Equation (38) maps the price change used for identification ($\Delta \log p$), the observed elasticity $\widehat{\varepsilon}$, and the degree of frictions δ to bounds on the structural elasticity. Figure 4 plots the bounds $(\varepsilon_L, \varepsilon_U)$ vs. $\widehat{\varepsilon}$ for four combinations of δ and $\Delta \log p$. The top two panels consider $\delta = 1\%$, while the lower two panels consider $\delta = 0.5\%$. The left panels have a price change of $\Delta \log p = 20\%$, while the right panels have $\Delta \log p = 40\%$. These bounds are computed using the formula in Proposition 1, which relies on a quadratic approximation to utility. To evaluate the quality of the approximation, I calculated the exact bounds with the utility in (2) numerically for a range of values of $\widehat{\varepsilon} < 1$, $\Delta \log p < 100\%$, and $\delta = 1\%$. In all cases, the exact and approximate bounds differ by less than 0.001, indicating that the simple analytical bounds in Proposition 1 are sufficiently accurate for most applications.[8]

The bounds in Proposition 1 cannot be directly applied to finite sample estimates of $\widehat{\varepsilon}$. When $\widehat{\varepsilon}$ is a finite-sample estimate, a 95% confidence set for ε can be obtained by computing

[8] Exact bounds can be computed numerically for any utility function by calculating choice sets and the minimal and maximal shifts in these sets consistent with a given observed elasticity.

ε_L using the lower limit of the 90% confidence interval for $\widehat{\varepsilon}$ and ε_U using the upper limit of the 90% confidence interval under certain regularity conditions (Imbens and Manski 2004).

Proposition 1 recovers information about a structural parameter of the true model (ε) using estimates of a mis-specified nominal model that ignores frictions. This approach is useful for three reasons. First, it permits identification of ε in standard adjustment cost models without placing structure on the taste shocks $a_{i,t}$, adjustment costs $k_{i,t}$, or expectations about price changes. Second, it permits identification of ε in environments where agents' choices differ systematically from the predictions of the nominal model in ways that we are unable to model. Finally, it extracts information about structural parameters of complex dynamic models purely from reduced-form estimates of observed elasticities, which can often be identified using transparent, quasi-experimental research designs.

The appendix extends Proposition 1 to the case where flow utility is not quasilinear. The formula for the bounds is unchanged, but the bounds apply to the Hicksian elasticity – that is, they map an observed compensated elasticity $\widehat{\varepsilon}^c$ to bounds on the structural compensated elasticity ε^c. All of the results that follow apply to Hicksian elasticities in a general model that does not restrict the functional form of flow utility.

Properties of the Bounds. The bounds offer some insights into what can be learned about structural elasticities from reduced-form estimates of observed elasticities. First, larger price changes are much more informative about ε because the bounds shrink at a quadratic rate with $\Delta \log p$. With a price change of 20%, an observed elasticity of $\widehat{\varepsilon} = 0.2$ is consistent with structural elasticities up to $\varepsilon_U = 2.3$. With $\Delta \log p = 40\%$ and $\widehat{\varepsilon} = 0.2$, $\varepsilon_U = 0.85$. The reason for this rapid shrinkage is that the movement in the choice sets for a given value of ε is larger when $\Delta \log p$ is larger, resulting in a narrower set of observed responses $\widehat{\varepsilon}$ consistent with any given ε.

Second, the bounds are asymmetric around the observed elasticity: $\varepsilon_U - \widehat{\varepsilon} > \widehat{\varepsilon} - \varepsilon_L$. This asymmetry is driven by the proportional relationship between the width of the choice sets and ε, as shown in Lemma 1. Small structural elasticities are inconsistent with large observed values of $\widehat{\varepsilon}$, making the lower bound relatively tight. Large structural elasticities generate wide choice sets – because they imply a relatively flat utility function around the optimum – and are consistent with many values of $\widehat{\varepsilon}$, making ε_U large. A related implication is that if ε is small, there will be little dispersion in observed elasticities across studies, whereas a large ε

may lead to substantial variation in observed elasticities.

Third, the lower bound is strictly positive ($\varepsilon_L > 0$) whenever $\widehat{\varepsilon} > 0$, irrespective of δ. If $\varepsilon = 0$, the choice sets collapse to a single point $x_t^*(p_A) = x_t^*(p_B)$ as shown in Lemma 1, and one will therefore never observe positive values of $\widehat{\varepsilon}$. Agents intent on maintaining a fixed value of x must face very large costs of deviating from the optimum and therefore will never do so. The following corollary of Proposition 1 establishes this result.[9]

Corollary 1. Under assumption A1, if $\widehat{\varepsilon} > 0$, the hypothesis that $\varepsilon = 0$ is rejected for any δ: $\widehat{\varepsilon} > 0 \Rightarrow \varepsilon_L(\widehat{\varepsilon}, \delta) > 0$.

Proof. Follows directly from the expression for ε_L in (38), where the second term can be shown to be strictly less than $\widehat{\varepsilon}$ for $\widehat{\varepsilon} > 0$.

Finally, consider the converse case of a study that detects zero observed behavioral response ($\widehat{\varepsilon} = 0$).[10] When $\widehat{\varepsilon} = 0$, the bounds take a particularly simple form. The lower bound is $\varepsilon_L = 0$. The upper bound can be conveniently expressed in terms of the utility cost of ignoring the price change for an optimizing agent with time-invariant preferences. Consider a hypothetical agent who has fixed tastes $a_{i,t}$ across periods A and B ($u_{i,B}(x) = u_{i,A}(x) = u_i(x)$) and is initially at his nominal optimum $x_i^*(p_A)$. Using a quadratic approximation analogous to that in Lemma 1, this agent's flow utility loss from failing to change demand to $x_i^*(p_B)$ is

$$\Delta u_i \equiv u_i(x_i^*(p_B)) - u_i(x_i^*(p_A)) \simeq -\frac{1}{2}u_i''(x_{i,A}^*)(\log x_{i,B}^* - \log x_{i,A}^*)^2(x_{i,A}^*)^2.$$

Using equation (15), for any a_i, the utility loss from failing to reoptimize in response to a price change as a percentage of the original expenditure level is

$$\Delta u_{\%}(\varepsilon) = \frac{\Delta u_i}{p_A x_i^*(p_A)} = \frac{1}{2}\varepsilon(\Delta \log p)^2 \tag{20}$$

The utility loss $\Delta u_{\%}(\varepsilon)$ is an increasing function of the structural elasticity ε. The following result shows that the upper bound on ε when $\widehat{\varepsilon} = 0$ can be expressed in terms of $\Delta u_{\%}(\varepsilon_U)$.

Corollary 2. Under assumption A1, when $\widehat{\varepsilon} = 0$, the upper bound structural elasticity $\varepsilon_U(\widehat{\varepsilon} = 0, \delta)$ satisfies

$$\Delta u_{\%}(\varepsilon_U) = 4\delta \tag{21}$$

[9] The same reasoning implies that an estimate of $\widehat{\varepsilon} < 0$ implies $\varepsilon > 0$, as one could never observe a negative response if $\varepsilon = 0$. Note that negative structural elasticities ($\varepsilon < 0$) are ruled out by agent optimization.

[10] Among 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.

Proof. When $\widehat{\varepsilon} = 0$, (38) implies

$$\varepsilon_U = 8\delta/(\Delta \log p)^2. \tag{22}$$

Combining (22) with (20) yields (21).

Corollary 2 provides a simple method of determining the range of structural elasticities for which one can be sure to detect a behavioral response, analogous to a statistical power calculation. Starting from the optimum, the percentage utility cost of ignoring a price change given an elasticity of ε must exceed 4δ to guarantee an observed elasticity $\widehat{\varepsilon} > 0$. The intuition for the 4δ condition is illustrated in Figure 5 for a case with no heterogeneity across agents. Let $d = x_A^*(p_A) - \min(X_A(p_A, \delta))$ denote the difference between the mean optimal demand and the lowest mean demand in the initial choice set. Panel A shows that at the upper bound ε_U, the difference between the optimal demands at the two prices is $x^*(p_A) - x^*(p_B) = 2d$. By definition, the percentage utility cost of choosing $\min(X_A(p_A, \delta))$ instead of $x^*(p_A)$ is δ. Given that 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. The 4δ condition is obtained because the cost of deviating from the optimum rises at a quadratic rate.

Equation (22) shows that ε_U shrinks at a quadratic rate with $\Delta \log p$ but only a linear rate with δ when $\widehat{\varepsilon} = 0$, as can be seen in Figure 4. Studying a price change that is twice as large yields more information about ε even if frictions are also twice as large, underscoring the value of placing greater weight on large treatments for identification.

3.3 Combining Multiple Observed Elasticities

One can obtain more information about the structural elasticity by combining multiple observed elasticities. Suppose we have a set of observed elasticities $\{\widehat{\varepsilon}_1, ..., \widehat{\varepsilon}_J\}$ from J empirical studies. Let $\Delta \log p_j$ denote the size of the price change used to identify observed elasticity j. Let ε_L^j and ε_U^j denote the lower and upper bounds implied by study j, derived using Proposition 1. Let $\varepsilon_L^{\max} = \max(\varepsilon_L^j)$ denote the largest lower bound and $\varepsilon_U^{\min} = \min(\varepsilon_U^j)$ denote the least upper bound. Then it follows that $\varepsilon \in (\varepsilon_L^{\max}, \varepsilon_U^{\min})$.

Inference for the unified bounds $(\varepsilon_L^{\max}, \varepsilon_U^{\min})$ can be handled using the techniques proposed by Chernozhukov, Hong, and Tamer (2007). Constructing a 95% coverage region for $(\varepsilon_L^{\max}, \varepsilon_U^{\min})$ is more difficult than constructing a confidence interval for bounds from a single study because of an order statistics problem. Intuitively, a confidence interval constructed only from the estimates of the studies that deliver $\varepsilon_L^{\max}$ and $\varepsilon_L^{\min}$ is too narrow because these studies may have obtained large or small elasticities purely due to sampling error.

By calculating $(\varepsilon_L^{\max}, \varepsilon_U^{\min})$ as a function of δ, one can assess how sensitive estimates of ε are to frictions. One value of special interest is the smallest δ that reconciles the observed elasticities, $\delta_{\min}$. When $\delta = \delta_{\min}$, the structural elasticity ε is point identified. To characterize this minimum-δ value of ε, let study 1 denote the study with the least upper bound when $\delta = \delta_{\min}$ and study 2 that with the highest lower bound, i.e. $\varepsilon_U^1(\delta_{\min}) = \varepsilon_U^{\min}(\delta_{\min})$ and $\varepsilon_L^2(\delta_{\min}) = \varepsilon_L^{\max}(\delta_{\min})$. The minimum-$\delta$ estimate of ε satisfies

$$\varepsilon = \varepsilon_U^1(\delta_{\min}) = \varepsilon_L^2(\delta_{\min}).$$

Using the definitions of ε_U and ε_L in (16) and (17) and solving these two equations for $\delta_{\min}$ and ε yields the following estimator under assumption A1:

$$\varepsilon_{\delta\text{-min}} = \frac{\Delta \log p_1}{\Delta \log p_1 + \Delta \log p_2}\widehat{\varepsilon}_1 + \frac{\Delta \log p_2}{\Delta \log p_1 + \Delta \log p_2}\widehat{\varepsilon}_2 \tag{23}$$

The minimum-δ estimate is a weighted average of observed elasticities from the pivotal studies. Studies identified from larger price changes receive more weight in this estimator both because they are more likely to be pivotal and because (23) places greater weight on the pivotal study that is identified from the larger price change. An advantage of this estimator for ε relative to calculating unified bounds is that it does not require exogenous specification of δ. This approach is attractive when deviations from the nominal model are known to be small and one therefore aims to identify the structural elasticity for the smallest level of frictions consistent with the data.

The value $\delta_{\min}$ is itself of interest because it can be used to formalize the notion of "economically significant" differences that is used loosely in the existing literature. A small $\delta_{\min}$ shows that small frictions can explain the differences across estimates, implying that modifications in the fundamental economic model are not necessary to reconcile the evidence. Economically significant differences in estimates require large $\delta_{\min}$ to be reconciled. In analogy with the standard practice of reporting the statistical significance of differences between estimates, reporting the $\delta_{\min}$ required to reconcile estimates may be a useful way to quantify the economic significance of a new finding.

In applications with large δ and uninformative unified bounds, one can obtain sharper identification of ε by imposing parametric assumptions on how frictions affect behavior. As an illustration, in the working paper version (Chetty 2009), I consider the assumption that the difference between mean observed and optimal demand is independently distributed across periods and has zero mean. Under this assumption, it is straightforward to establish that for any δ, the best (minimum-variance) linear unbiased estimate of ε is $\varepsilon_{\text{BLUE}} = \sum_j (\Delta \log p_j)^2 \widehat{\varepsilon}_j / \sum_j (\Delta \log p_j)^2$. In future work, it would be useful to identify other restrictions implied by models with frictions in order to obtain sharper identification of ε.

4 Application: Labor Supply and Taxation

The wage (or net-of-tax) elasticity of labor supply is a parameter of central interest for tax policy analysis and macroeconomic models. A large 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 applies the methodology developed above to explain a set of puzzles in this literature and identifies bounds on the labor supply elasticity using estimates from existing studies.

The labor supply literature estimates many different types of elasticities. I focus primarily on the intensive margin elasticity – the effect of tax changes on hours of work or earnings for those who are already in the labor force. I analyze extensive margin elasticities in section 4.2, showing that the bounds on extensive margin elasticities are an order of magnitude tighter than those on intensive margin elasticities. I restrict attention to studies that estimate Hicksian (income-constant) elasticities, but discuss the implications of the analysis for the Frisch elasticity in section 4.4. Finally, the literature measures "labor supply" in different ways. Traditional studies measure hours worked, but modern studies focus on taxable income (Feldstein 1995, Saez, Slemrod, and Giertz 2009). Taxable income elasticities may be larger than labor supply elasticities because they incorporate changes in reporting and avoidance behavior as well as changes in work effort (Slemrod 1995). I analyze a model where the hours and taxable income elasticities are the same, and pool estimates from both types of studies to bound the structural labor supply elasticity in this model.

4.1 Nominal Labor Supply Model

I begin by adapting the demand model above to labor-leisure choice. Consider a lifecycle model in which agents choose consumption (c_t) and hours of work (l_t) to solve

$$\max \sum_{t=0}^{T} v_t(c_t, l_t) \text{ s.t. } \sum_{t=0}^{T} [Y_{i,t} + (1-\tau_t) w l_t - c_t] = 0 \tag{24}$$

where τ_t denotes the tax rate in period t, w denotes the wage rate, and $Y_{i,t}$ denotes unearned (non-wage) income.[11] The bounds calculated below apply to the Hicksian elasticity without any restrictions on $v_t(c_t, l_t)$. However, for expositional convenience, consider the case in which the flow utility function is quasilinear and iso-elastic:

$$v_{i,t}(c_t, l_t) = c_t - a_{i,t} \frac{l_t^{1+1/\varepsilon}}{1+1/\varepsilon} \tag{25}$$

The optimal level of hours is $l^*_{i,t} = \left(\frac{(1-\tau_t)w}{a_{i,t}}\right)^{\varepsilon}$, or equivalently,

$$\log l^*_{i,t} = \alpha_{i,t} + \varepsilon \log(1-\tau_t) w \tag{26}$$

where $\alpha_{i,t} = -\varepsilon \log a_{i,t}$. Let $\log l^*_t = \mathbb{E} \log l^*_{i,t}$ denote the mean level of (log) labor supply in period t. Under the orthogonality assumption A1, the structural preference parameter ε is identified by the hours response to a change in tax rate from τ_A to τ_B:

$$\varepsilon = \frac{\log l^*_B - \log l^*_A}{\log(1-\tau_B) - \log(1-\tau_A)}$$

Earnings (or taxable income) are given by wl^*_t. In this one-dimensional model of labor supply, the hours elasticity ε is equal to the taxable income elasticity: $\varepsilon = \frac{\log wl^*_B - \log wl^*_A}{\log(1-\tau_B) - \log(1-\tau_A)}$. I shall therefore refer to ε simply as the structural labor supply elasticity. Identifying ε is important for designing optimal income tax policies (e.g. Saez 2001), predicting steady state effects of taxation on output and revenue (e.g. Trabandt and Uhlig 2009), and calibrating macroeconomic models (e.g. Prescott 2004).

Let agent i's observed level of labor supply in period t be denoted by $l_{i,t}$ and define mean observed labor supply as $\log l_t = \mathbb{E} \log l_{i,t}$. The observed labor supply elasticity is

$$\widehat{\varepsilon} = \frac{\log l_B(\tau_B) - \log l_A(\tau_A)}{\log(1-\tau_B) - \log(1-\tau_A)}.$$

[11] I assume that w is constant to simplify notation. Variation in w across individuals and time is isomorphic to variation in $a_{i,t}$ and does not affect the results below.

There are many frictions that may make $\widehat{\varepsilon}$ differ from ε, 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 labor supply elasticities account for such frictions. The methods developed above are therefore well suited to extracting the information these studies contain about ε.

Let $u_{i,t}(\widetilde{l}_{i,t})$ denote the maximum utility the agent can attain over periods t to T if he sets $l_{i,t} = \widetilde{l}_{i,t}$. I define a δ class of models around the nominal model in (24) by requiring that agents' average utility loss from setting $l_{i,t}$ suboptimally is less than δ percent of net-of-tax earnings $(1-\tau_t)wl^*_{i,t}$ in every period t:

$$u_{i,t}(l^*_{i,t}) - u_{i,t}(l_{i,t}) \leq \delta_{i,t}c^*_{i,t} \; \forall i \text{ with } \textstyle\sum_i \delta_{i,t}/N \leq \delta \tag{27}$$

In this model, Proposition 1 can be immediately applied to obtain bounds on ε, with $\Delta \log p = \Delta \log(1-\tau)$ and δ measuring frictions in choosing labor supply as a percentage of consumption.

4.2 Costs of Ignoring Tax Changes: A Synthesis of Micro Evidence

Before calculating bounds on ε, I show that frictions can explain several divergent findings in the labor supply literature. Motivated by Corollary 2, I calculate the utility costs of ignoring the tax changes used for identification in the microeconometric literature. I find that observed elasticities are large when the utility costs of ignoring tax changes are large, suggesting that frictions could explain the variation in observed elasticities.

Calculating Utility Costs: Methodology. Corollary 2 shows that the utility cost relevant for calculating bounds on ε is for an agent with time-invariant preferences. Therefore, I assume that agents' tastes are time-invariant ($a_{i,t} = a_i$) throughout this subsection. To calculate the utility costs of ignoring changes in a progressive income tax system, let $T_t(wl)$ denote an agent's tax liability as a function of his taxable income in year t. An agent with taste a_i has flow utility $u_i(l;T_t) = wl - T_t(wl) - a_i\frac{l^{1+1/\varepsilon}}{1+1/\varepsilon}$. I consider tax changes over a three year interval, following the convention in the literature (Gruber and Saez 2002). The utility loss in dollars from ignoring the tax changes that occur between years $t-3$ and t for an individual who sets labor supply at the optimum in year $t-3$ is:[12]

$$\Delta u_{i,t} = u_i(l^*_{i,t}; T_t) - u_i(l^*_{i,t-3}; T_t) \tag{28}$$

I calculate tax rates $T_t(wl)$ using the NBER TAXSIM calculator, including both employer and employee payroll taxes but ignoring state taxes. I consider a single tax filer with two children who has only labor income and no deductions other than those for children. I adjust for inflation in the wage w using the CPI over the relevant three-year period. Under these assumptions, I numerically calculate $l^*_{i,t}$ and $\Delta u_{i,t}$ for various values of a_i and years t with a structural elasticity of $\varepsilon = 0.5$, the midpoint of the bounds calculated in the next section. I also compute $\Delta u_{i,t,\%} = \Delta u_{i,t}/c^*_{i,t-3}$, the utility loss from failing to reoptimize as a percentage of pre-reform consumption, which corresponds to the measure used in Corollary 2.[13]

A small complication in using Corollary 2 to evaluate whether frictions could produce zero observed elasticities is that it only applies to linear budget sets. For individuals who are at an interior optimum both before and after the tax change, Corollary 2 still holds. In particular, a tax change could produce an observed elasticity $\widehat{\varepsilon} = 0$ if the level of frictions $\delta > \Delta u_{i,t,\%}(\varepsilon)/4$. For individuals who optimally locate at kinks between tax brackets, the tangency conditions used to derive Corollary 2 do not hold. However, even for those who would choose kinks, it is easy to see that a tax change could produce $\widehat{\varepsilon} = 0$ if $\delta > \Delta u_{i,t,\%}(\varepsilon)$. Hence, the utility cost of ignoring tax changes remains useful for assessing which reforms will generate behavioral responses even with a progressive tax system.

I now calculate the utility costs of ignoring some of the major tax reforms studied in the empirical literature. Each reform highlights a different set of stylized facts that can be explained by frictions.

Tax Reform Act of 1986: Low vs. High Incomes. Figure 6 evaluates the costs of ignoring the Tax Reform Act of 1986 (TRA86), one of the largest reforms in the U.S. tax code and the focus of many empirical studies. Panel A shows the marginal tax rate schedules in 1985 (thick

[12] To be clear, the results below do *not* assume that all agents are at the optimum in the base year; they only require that choices in the base year lie within a δ class of models. I calculate utility costs for agents who start at the optimum in year t because this calculation tells us whether $\widehat{\varepsilon} = 0$ is consistent with a given structural elasticity (Corollary 2).

[13] A stata program TAXCOST.ado that calculates the utility cost of ignoring tax reforms has been posted on the NBER server. TAXCOST takes exactly the same inputs as TAXSIM. By running TAXCOST instead of TAXSIM, researchers can calculate the utility costs of ignoring the tax changes they are using for identification. See http://obs.rc.fas.harvard.edu/chetty/taxcost.html for further information.

red line) and 1988 (thin blue line). The dashed blue line, which is replicated in all the panels as a reference, shows the percentage change in the marginal net-of-tax rate (NTR). TRA86 increased the NTR by 15-20% for those with incomes below $100,000 and by nearly 40% for those with incomes close to $200,000.

Panel B plots the utility cost (measured in dollars) of ignoring the tax change (Δu_i) vs. base year taxable income.[14] For instance, an individual whose taste parameter a_i placed him at an optimal taxable income of $100,000 prior to TRA86 would lose $1,000 by failing to reoptimize labor supply in response to the change in the tax code. Panel C plots the cost of ignoring the tax reform as a percentage of consumption, $\Delta u_{i,\%}$. Most individuals earning less than $100,000 lose less than 1% of consumption by ignoring TRA86 when choosing labor supply in 1988 with $\varepsilon = 0.5$. These small utility losses reflect the second-order costs of deviating from an optimum and are consistent with the formula in (20), which predicts that $\Delta u_{\%} = \frac{1}{2}\varepsilon \cdot (0.2)^2 = 1\%$ for a 20% change in the NTR. These calculations imply that frictions of $\delta < 1\%$ could lead to an observed elasticity of $\widehat{\varepsilon} = 0$ even if the underlying structural elasticity is $\varepsilon = 0.5$.

Finally, Panel D plots the change in taxable income ($wl^*_{i,t+3} - wl^*_{i,t}$) required to reoptimize relative to TRA86. With $\varepsilon = 0.5$, a taxpayer earning $100,000 prior to the reform would have to increase his pre-tax earnings by $13,000 in order to reach his new optimum. This substantial change would yield a utility gain (net of the disutility of added labor) of only $1,000. Given that the search costs of finding additional work that pays an extra $13,000 could well exceed $1,000, it is plausible that many individuals would not respond to TRA86 within a three-year horizon. This could explain why empirical studies of TRA86 find little or no change in earnings for low and middle income wage earners between 1985 and 1988 (e.g. Gruber and Saez 2002, Saez 2004).[15]

The costs of ignoring TRA86 are considerably larger for high income earners. An individual earning $200,000 in 1985 would lose $4,500 per year (4% of consumption) by ignoring the tax

[14] Values at non-convex kinks in the base year are interpolated to obtain a continuous curve. Since no individual would optimally locate at a non-convex kink, the utility cost is undefined at these points.

[15] The total lifetime gain from reoptimizing labor supply is much larger because the agent gains $1,000 every year. The key point here is that because the flow utility gains are relatively small, many agents may delay adjustment until a period where frictions (e.g. job switching costs) are lower. Thus, micro studies might not detect much change in labor supply between 1985 and 1988 even if TRA86 induced individuals to reoptimize in the long run.

reform. High income individuals gain a lot more from reoptimizing both because the dollars at stake rise with income and because the change in tax rates was larger for high incomes. This is consistent with the much larger elasticities observed for high income earners in studies of TRA86 (e.g. Auten and Carroll 1999, Saez 2004).[16]

EITC Expansions: Intensive vs. Extensive Margin. Figure 7 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 rose by 20% for single tax filers with two children earning below $10,000 as the phase-in subsidy was increased. In contrast, net-of-tax wages fell by roughly 15% for those with incomes between $15,000 and $30,000 because of the increase in the phase-out tax rate.

Panel B, which is analogous to Figure 6c, shows that most individuals lose less than 2% of consumption by ignoring the EITC expansions entirely. Corollary 2 therefore implies an observed response of $\widehat{\varepsilon} = 0$ would be consistent with $\varepsilon = 0.5$ if one permits $\delta = 1\%$ frictions in reoptimizing labor supply. Correspondingly, most studies find virtually no changes in labor supply in response to EITC expansions for individuals on the intensive margin (Meyer and Rosenbaum 2001, Hotz and Scholz 2003, Eissa and Hoynes 2006). The same empirical studies, however, do find a substantial response on the *extensive* margin: labor force participation rates for single women with children surged as a result of the EITC expansion. I now show that this difference between observed extensive and intensive margin elasticities can be explained by frictions.

I begin by introducing fixed costs into the nominal linear tax model in (24) to generate an extensive margin. Suppose that an individual must pay a fixed cost k to enter the labor force. Let $u_i(l) = (1-\tau_t)wl - a_i \frac{l^{1+1/\varepsilon}}{1+1/\varepsilon} - k \cdot [l > 0]$ denote the utility obtained from choosing l units of labor supply, and let $l_i^*(\tau_t)$ denote the optimal labor supply choice. Letting $\widetilde{l}_i(\tau_t)$ denote the optimal labor supply choice conditional on working, individuals with a_i such that

$$u_i(\widetilde{l}_i(\tau_t)) = (1-\tau_t)w\widetilde{l}_i(\tau_t) - a_i \frac{(\widetilde{l}_i(\tau_t))^{1+1/\varepsilon}}{1+1/\varepsilon} - k < 0 \quad (29)$$

will not work in period t. This condition implicitly defines a threshold $\overline{a}(\tau_t)$ such that individuals with disutility of labor $a_i > \overline{a}(\tau_t)$ set $l_i^*(\tau_t) = 0$. Let the distribution of a_i in

[16] High income individuals may also be more responsive because they have lower adjustment costs or higher structural elasticities. The analysis here is not intended to rule out these other explanations. It merely shows that the simple model in (24) with a constant elasticity ε across individuals can explain the data if one permits small frictions. One does not necessarily need more complex models to explain the evidence available to date.

the population be given by a smooth cdf $F(a_i)$ with positive support for all $a_i > 0$. Let $\theta^*(\tau_t) = F(\overline{a}(\tau_t))$ denote the optimal labor force participation rate and $\theta_t(\tau_t)$ denote the observed labor force participation rate. The structural extensive margin labor supply elasticity for a tax change from τ_A to τ_B is

$$\varepsilon_{\text{ext}}(\tau_A, \tau_B) \equiv \frac{\log \theta^*(\tau_B) - \log \theta^*(\tau_A)}{\log(1-\tau_B) - \log(1-\tau_A)} \simeq \frac{f(\overline{a}(\tau_A))\overline{a}}{F(\overline{a}(\tau_A))} \frac{d \log \overline{a}}{d \log(1-\tau_t)}$$

The corresponding observed extensive margin elasticity is $\widehat{\varepsilon}_{\text{ext}}(\tau_A, \tau_B) \equiv \frac{\log \theta(\tau_B) - \log \theta(\tau_A)}{\log(1-\tau_B) - \log(1-\tau_A)}$. Note that the structural intensive and extensive margin elasticities are independent parameters: the structural primitive that controls the intensive-margin elasticity is ε, whereas ε_{ext} depends on the taste distribution (F). Because the density $f(\overline{a}(\tau_A))$ varies with the tax rate, $\varepsilon_{\text{ext}}(\tau_A, \tau_B)$ varies with tax rates.

Now consider the utility costs of ignoring tax changes in this augmented model with fixed costs. For agents with $l_i^*(\tau_B) > 0$, the utility cost of ignoring a tax cut to $\tau_B < \tau_A$ depends upon whether they were in the labor force before the tax change:

$$\begin{aligned} \Delta u_i &= \{u_i(\widetilde{l}_i(\tau_B); \tau_B) - u_i(\widetilde{l}_i(\tau_A); \tau_B)\} + \mathbb{I}[l_i^*(\tau_A) = 0] u_i(\widetilde{l}_i(\tau_A); \tau_B) && (30) \\ &\simeq \frac{1}{2}(\frac{\Delta\tau}{1-\tau_A})^2 \widetilde{c}_{i,A}\varepsilon + \mathbb{I}[l_i^*(\tau_A) = 0](\frac{-\Delta\tau}{1-\tau_A}\widetilde{c}_{i,A} + u_i(\widetilde{l}_i(\tau_A); \tau_A)) && (31) \end{aligned}$$

where the first term in the second line uses a quadratic approximation similar to that in Corollary 2. To convert this measure into percentage units, I normalize Δu_i by optimal consumption when working, $\widetilde{c}_{i,A}$:

$$\Delta u_{i,\%} \equiv \frac{\Delta u_i}{\widetilde{c}_{i,A}} = \frac{1}{2}(\frac{\Delta\tau}{1-\tau_A})^2 \varepsilon + \mathbb{I}[l_i^*(\tau_A) = 0](\frac{-\Delta\tau}{1-\tau_A} + \frac{u_i(\widetilde{l}_i(\tau_A); \tau_A)}{\widetilde{c}_{i,A}}) \quad (32)$$

The first term in this expression, $\Delta u_{i,\text{int},\%} = \frac{1}{2}(\frac{\Delta\tau}{1-\tau_A})^2\varepsilon$, reflects the loss from working $\widetilde{l}_i(\tau_A)$ hours instead of $\widetilde{l}_i(\tau_B)$ hours. This expression is equivalent to that obtained in the pure intensive margin model in (24). The second term, $\Delta u_{i,\text{ext},\%} = -\frac{\Delta\tau}{1-\tau_A} + \frac{u_i(\widetilde{l}_i(\tau_A);\tau_A)}{\widetilde{c}_{i,A}}$, arises only for agents who were not working prior to the tax cut. This term reflects the utility loss from failing to reoptimize on the extensive margin – that is, from failing to enter the labor force.

The extensive margin utility loss $\Delta u_{i,\text{ext},\%}$ is linear in $\Delta\tau$ whereas the intensive margin utility loss $\Delta u_{i,\text{int},\%}$ is proportional to $(\Delta\tau)^2$. As a result, some agents on the extensive margin can bear very large utility costs from ignoring tax cuts (or increases). For instance, for the agent who has $u_i(\widetilde{l}_i(\tau_A);\tau_A)) = 0$ and is just on the margin of entering the labor force at tax rate τ_A, the utility cost of ignoring a 10% increase in his net-of-tax wage is 10% of consumption if he were initially not working. However, the cost of ignoring the same tax cut would be only 0.25% if he were initially working $\widetilde{l}_i(\tau_A)$ hours with $\varepsilon = 0.5$. The reason for this 40-fold difference is that agents on the extensive margin are not near their post-reform optimum to begin with. The first-order gains from a tax cut (e.g. a larger EITC refund) are automatically obtained on the intensive margin even if a worker does not change his hours. But non-workers get the first-order benefits of the tax cut only if they reoptimize and start to work.

Because of the first-order utility costs, the range of structural elasticities consistent with zero observed response is much smaller on the extensive margin. To see this, let δ denote the degree of frictions permitted as a fraction of consumption $\widetilde{c}_{i,A}$ for all agents, so that each agent's labor supply choice $l_{i,t}$ satisfies $u(l^*_{i,t}) - u(l_{i,t}) < \delta\widetilde{c}_{i,A}$.[17] Consider an agent with a_i such that $\frac{u_i(\widetilde{l}_i(\tau_A);\tau_A)}{\widetilde{c}_{i,A}} = -\delta$, so that he falls just outside his choice set by working at the initial tax rate τ_A. To observe $\widehat{\varepsilon}_{\text{ext}} = 0$, this agent must remain within his choice set in period B without entering the labor force. This requires that his percentage utility cost from ignoring the tax change is less than δ:

$$\begin{aligned} \Delta u_{i,\text{ext},\%} &= -\frac{\Delta\tau}{1-\tau_A} + \frac{u_i(\widetilde{l}_i(\tau_A);\tau_A)}{\widetilde{c}_{i,A}} < \delta \\ &\Rightarrow -\frac{\Delta\tau}{1-\tau_A} < 2\delta \end{aligned} \tag{33}$$

This result is the analog of Corollary 2 for the extensive margin. If the percentage change in the net-of-tax rate exceeds 2δ, we must see $\widehat{\varepsilon}_{\text{ext}} > 0$. A 20% change in the net-of-tax wage would produce no behavioral response on the extensive margin only with frictions of $\delta > 10\%$ of consumption. In contrast, Corollary 2 shows that a 20% change in the net-of-tax wage could produce $\widehat{\varepsilon} = 0$ on the intensive margin with a structural elasticity of $\varepsilon = 0.5$ even if $\delta = 0.25\%$ of consumption.

Equation (33) can be restated as requiring that $\Delta u_{i,\text{ext},\%} < \delta$ for the agent who has $u_i(\widetilde{l}_i(\tau_A);\tau_A)) = 0$ and is just on the margin of entering the labor force at tax rate τ_A.

[17] Heterogeneity in $\delta_{i,t}$ across agents does not affect the result in (33) provided that $\mathbb{E}\delta_{i,t}|a_i < \delta$ for each a_i to ensure that the choice set has the same width for the marginal agents at each level of the tax rate τ_t. This condition was not necessary in the intensive margin case because the marginal agent did not vary with τ_t there. Allowing the choice set to vary in width with $\widetilde{c}_{i,t}$, so that $u(l^*_{i,t}) - u(l_{i,t}) < \delta\widetilde{c}_{i,t}$ complicates the algebra significantly but simply changes the condition in (33) to $-\frac{\Delta\tau}{1-\tau_B} < 2\delta$.

Hence, the utility cost of ignoring tax changes again sheds light on whether they will generate behavioral responses in an environment with frictions.[18] Motivated by this result, Panels C and D of Figure 7 plot the costs of ignoring the Clinton EITC expansion for agents on the extensive margin. The x axis of these figures is the taxable income that the individual would optimally earn ($w\widetilde{l_i}(\tau_{1993})$) were he to work prior to the EITC expansion. On the extensive margin, the relevant tax rates are average rather than marginal tax rates. Panel C 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 D shows the utility cost of ignoring the EITC expansion for individuals on the margin of entering the labor force at various income levels in 1993. The figure plots the difference in utility between choosing labor supply optimally in 1996 and remaining out of the labor force in 1996 for individuals indifferent between working and not working in 1993.[19] The denominator used to calculate the percentage utility cost is the consumption level if the agent works in 1993, as in (32). 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 extra $2,000 EITC refund if he had stayed out of the labor force. The utility cost exceeds the change in the net-of-tax rate because it incorporates both the extensive ($\Delta u_{i,\text{ext},\%}$) and intensive margin utility costs ($\Delta u_{i,\text{int},\%}$).

Because the utility losses from ignoring tax changes are so large for agents on the extensive margin, small tax changes are likely to induce rapid changes in labor force participation even with frictions. Hence, small frictions can explain why microeconometric studies of short-run responses have consistently found much larger labor supply elasticities on the extensive margin than the intensive margin (Pencavel 1986, Blundell and MaCurdy 1999).

Tax Reforms from 1970-2006. Figure 8 extends the analysis of tax reforms to cover all tax changes from 1970-2006. I compute the percentage utility loss ($\Delta u_{\%}$) from ignoring tax changes at the 20th, 50th, and 99.5th percentile of the household income distribution. The

[18] The critical threshold is δ instead of 4δ as on the intensive margin because the utility losses from ignoring price changes are first-order on the extensive margin.

[19] For each income level shown on the x axis, I find the taste parameter a_i that would make that earnings level optimal in 1993. I then choose a fixed cost k so that this individual's net utility from working in 1993 zero, placing him on the extensive margin. The figure plots the net utility from working in 1996 for such agents.

value plotted for year t is the percentage utility cost of choosing the level of labor supply that was optimal given the tax system in year $t-3$.[20] Panel A shows that on the intensive margin, there is no tax change since 1970 for which the utility cost of failing to reoptimize exceeds 1% of consumption for the median individual. The utility costs of ignoring tax reforms are substantial only for the top 1% of income earners around TRA86. Correspondingly, the largest observed elasticities in the historical time series are for top incomes around TRA86; for other groups, intensive margin elasticities are near zero (Saez 2004). Panel B shows that in contrast, there are several tax changes that would generate utility losses of 5-10% of consumption if ignored on the extensive margin.[21] The utility costs are particularly large for individuals who earn low incomes when working, consistent with the literature finding the largest extensive margin responses for this group.

While the flow utility costs of ignoring the tax *changes* that have occurred over the past four decades in the U.S. are small on the intensive margin, the utility costs of ignoring taxes in steady state are large. For example, the utility cost of ignoring a marginal tax rate of $\tau = 40\%$ and working $l^*(\tau = 0)$ hours is approximately $\frac{1}{2} \cdot \frac{1}{2} \cdot (0.4)^2 = 4\%$ of consumption per year when $\varepsilon = 0.5$. This calculation underscores the point that short-run responses to tax reforms in the U.S. may not be very informative about how the tax system affects labor supply on the intensive margin in steady state.

Bunching at Kinks and Non-Linear Budget Set Models. A basic challenge in estimating non-linear budget set models of labor supply (e.g. Hausman 1985, Blomquist and Newey 2002) is that they predict far more bunching at kinks in the tax schedule than what is observed in practice. This is illustrated in Figure 9, which plots the income tax schedule in 2006 (dashed blue line) for a single filer with two children. The solid grey curve shows the income distribution predicted by the frictionless model in (24) when $\varepsilon = 0.5$ and tastes a_i are uniformly distributed. The frictionless model predicts sharp spikes at each kink in a kernel density plot of the income distribution.[22] However, empirical income distributions for wage earners exhibit

[20] For 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 three years and out of the labor force for cases where it falls. This is the relevant calculation to determine the utility costs of failing to reoptimize on the extensive margin.

[21] I exclude the 99.5 percentile from this figure for scaling reasons and because few individuals enter the labor force at the 99.5 percentile of the income distribution.

[22] Adding noise to the income process can make the spikes (which arise from point masses) more diffuse, but plausible levels of noise do not eliminate bunching in the frictionless model (Saez 2002).

no such bunching at kinks (Saez 2002, 2009). The lack of bunching leads to the rejection of NLBS models, and has forced researchers to adopt various ad hoc solutions, such as smoothing the budget set to remove kinks or restricting the compensated elasticity to be positive (e.g. MaCurdy et al. 1990, Ziliak and Kniesner 1998).

Optimization frictions provide a simple explanation for why individuals do not bunch at kinks: the utility losses from ignoring kinks are very small for most individuals. The number next to each convex kink in Figure 9 shows the utility gain (as a percentage of consumption) from locating at that kink point relative to optimizing under the incorrect assumption that the rate in the previous bracket continues into the next bracket.[23] The utility losses are around 1% of consumption even at kinks that are predicted to produce large spikes in the income distribution. Hence, introducing small frictions in choosing l could generate income distributions that exhibit no bunching at kinks, a conjecture that is verified by simulations in Chetty et al. (2009). More generally, introducing optimization errors – by permitting agents to deviate systematically from their frictionless optima provided that the utility losses fall below some threshold – could provide a more disciplined and widely applicable method of estimating NLBS models.

Saez (2009) documents that the one kink at which there is 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 tax returns because of the lack of double reporting. Unlike changes in hours of work, misreporting generates a *first-order* utility gain because it transfers resources from the government to the taxpayer. The large utility gains from misreporting taxable income could explain why the self-employed bunch at this kink. Interestingly, Saez finds no bunching at the second kink of the EITC schedule (where the phase-out region begins) even for the self-employed. The first kink in the EITC schedule maximizes the size of the EITC refund while minimizing payroll tax liabilities. There is no reason to locate at the second kink if bunching is driven by the first-order gains from income manipulation.

Other Findings. In addition to the results described above – (1) larger observed elasticities for top incomes, (2) larger extensive margin elasticities, (3) rejection of NLBS models and no

[23] There are many values of a_i that can induce individuals to locate at each kink. The numbers in the figure are (unweighted) mean percentage losses for agents who would optimally locate at the kink.

bunching at kinks for wage earners in the U.S., and (4) bunching at kinks for the self-employed – frictions can also explain the following findings:

(5) Notches in budget sets, where a \$1 change in earnings leads to a discontinuous jump in consumption, generate substantial behavioral responses. For example, income cutoffs to qualify for Medicaid (Yelowitz 1995) and social security benefits in some pension systems (Gruber and Wise 1999) induce sharp reductions in labor supply. To calculate the utility cost of ignoring a notch, suppose that earning $wl_t > K$ leads to a penalty of P. Then the utility cost of setting $l_t > K/w$ for an individual with $l_t^* \leq K/w$ is $\Delta u > P$. Because the utility cost of ignoring a notch increases at a first-order rate with the size of the penalty P, eligibility cutoffs for large transfer programs will affect observed behavior substantially even with frictions.

(6) Elasticities are historically larger for secondary earners than primary earners, but have converged over time (Eissa and Hoynes 2004, Blau and Kahn 2007). The utility cost calculations above indicate that small variations in frictions could explain the differences in observed elasticities across primary and secondary earners even if they have the same structural elasticity. For instance, primary earners hold full-time jobs that tend to have rigid schedules (e.g. manufacturing jobs) while secondary earners traditionally held more flexible, part-time jobs. As secondary earners take up full-time jobs similar to those of primary earners, their observed elasticities shrink, even though the preferences of men and women (ε) may be constant and identical.[24]

(7) Chetty and Saez (2009) show that providing information about the tax code amplifies observed responses. If information reduces optimization frictions, it may amplify observed behavioral responses for a given structural elasticity ε.

(8) Using Danish data, Chetty et al. (2009) document that larger tax changes generate larger observed elasticties. The analysis here shows that larger tax changes are much less likely to generate $\widehat{\varepsilon} = 0$, consistent with this finding.

[24] Recent studies have called for lower tax rates on women because they are more elastic (Alesina et al. 2007, Kleven et al. 2009). If the differences in observed elasticities are due to frictions rather than tastes, taxes may distort the behavior of men as much as women in the long run, and there is no efficiency rationale for gender-based taxation.

4.3 Bounds on the Labor Supply Elasticity

Having shown that frictions provide a useful method of synthesizing the findings in the labor supply literature, I now calculate the bounds on the intensive margin Hicksian elasticity implied by twenty representative studies. One should keep three caveats in mind when interpreting the results of this exercise. First, I assume a common structural elasticity across all the studies, ignoring potential variation in preferences across income levels, demographic groups, or countries. Second, I take the estimates of each study at face value by assuming that they provide unbiased estimates of observed elasticities. Econometric issues such as omitted variables, mean reversion, and endogeneity of tax variation across countries may bias some of the elasticity estimates (Saez, Slemrod, and Giertz 2009). Any such biases would pass through to the bounds. Finally, I do not correct the bounds for the non-linearities in agents' budget sets created by the progressive tax system. Instead, I assume that agents face a linear budget set whose slope is given by their marginal tax rate (MTR) and apply Proposition 1 using $\Delta \log(1-\text{MTR})$ in place of $\Delta \log p$. This simple approach produces valid bounds on ε for agents who remain in the interior of budget segments. However, the bounds cannot be directly applied to agents who locate at kinks. Given that most of the modern labor supply literature estimates elasticities from changes in the behavior of agents away from kinks, this problem is unlikely to affect the bounds calculated below significantly.[25]

Table 1 lists the studies, divided into four groups based on methodological approach: (A) studies that use hours to measure labor supply; (B) studies that use taxable income to measure labor supply and focus on workers in the middle of the income distribution; (C) studies that use taxable income but focus exclusively on top income earners; and (D) macroeconomic studies that rely on cross-country comparisons or long-term trends for identification of hours elasticities. The table lists the point estimate and standard error of the observed Hicksian elasticity from the authors' preferred specification and the change in the net-of-tax rate used for identification. To calculate $\Delta \log(1-\text{MTR})$, I either use the reported percentage change in

[25] Recent studies that identify near-zero observed elasticities from bunching at kinks (e.g. Saez 2009, Chetty et al. 2009) are an exception. I incorporate these studies into the linear-demand framework by exploiting the fact that they also study *movements* in the kinks over time, which create reductions in marginal rates for the subgroup of individuals located between the old and new bracket cutoffs. The results of these studies imply that these individuals do not increase labor supply significantly when their marginal tax rates are lowered. This constitutes an observed elasticity estimate based on choices at interior optima, permitting application of Proposition 1.

the net-of-tax rate for the "treatment" group in the study or calculate it based on the variation in marginal tax rates in the authors' sample. Details on the calculations and sources for each study are given in the appendix.

The observed elasticity estimates vary substantially across the four categories. Microeconometric studies of the full population almost uniformly find very small elasticities: the mean observed hours elasticity among the studies considered in the table is 0.17 and the mean observed taxable income elasticity is 0.08. Interestingly, the largest observed elasticity in each of these two groups is obtained from the study that focuses on the largest change in tax policy: the abolition of the income tax for a year in Iceland (Bianchi, Gudmundson, and Zoega 2001) and a Swedish tax reform in 1991 termed the "tax reform of the century" (Gelber 2009). This pattern is consistent with the view that frictions are less likely to attenuate short-run responses to very large price changes. Studies of top income earners find much larger elasticities, with a mean of 0.85. Macroeconomic studies also find large elasticities, with a mean of 0.71. Note that the differences in point estimates of observed elasticities across the studies cannot be explained by statistical imprecision. The upper limits of the 95% confidence intervals for $\widehat{\varepsilon}$ for the studies in groups A and B are virtually all below the lower limits of the 95% confidence intervals for $\widehat{\varepsilon}$ for the studies in groups C and D.

Columns 6-7 of Table 1 show the bounds $(\varepsilon_L, \varepsilon_U)$ implied by each study's point estimate of the observed elasticity with frictions in choosing labor supply of $\delta = 1\%$ of consumption. The bounds are calculated using the formula in (38). The width of the bounds varies tremendously with the size of the variation used for identification. Traditional studies in the labor economics literature produce very wide bounds because they pool relatively small variation in wage and tax rates for identification. For instance, even though MaCurdy (1981) estimates an intensive margin Hicksian elasticity of only 0.15, his estimate is consistent with a structural Hicksian elasticity as large as $\varepsilon = 5.63$. The reason is that MaCurdy's estimates are identified from changes in wage rates of approximately 10%, which are not big enough to overcome small frictions. The studies in the more recent taxable income literature tend to yield narrower bounds because they use sharp changes in tax policy as quasi-experiments.

Figure 10 gives a visual representation of the bounds in columns 6-7 of Table 1. For scaling purposes, I exclude studies that use variation in net-of-tax rates of less than 20% for identification. 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 elasticity estimates, all twenty elasticities are consistent with a single structural elasticity ε if one permits 1% frictions. Second, when pooled together, the twenty studies in Table 1 yield informative bounds on the Hicksian elasticity. The unified lower bound across the twenty studies when $\delta = 1\%$ 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 (2009) analysis of the Swedish tax reform of 1991. This exercise shows that combining several studies can yield informative bounds even though any one study by itself produces wide bounds. By estimating elasticities in a broad range of environments, one can narrow the bounds on ε sharply.

Figure 11 shows how the unified bounds vary with the degree of frictions. The dark shaded region (between the solid red lines) shows the values of ε consistent with the observed elasticities in column 3 of Table 1 for $\delta \in (0, 5\%)$. The bounds naturally widen as δ rises, but remain somewhat informative even with $\delta = 5\%$, where $\varepsilon_L = 0.21$ and $\varepsilon_U = 1.23$. Given that individuals are unlikely to tolerate utility losses equivalent to 5% of consumption on average, we can rule out $\varepsilon < 0.21$ (as suggested by some microeconometric studies) or $\varepsilon > 1.23$ (as used in some macro calibrations) based on existing evidence. The smallest value of δ that can reconcile the twenty observed elasticities is $\delta_{\min} = 0.8\%$. That is, the differences in the observed elasticity estimates are "economically significant" only if frictions in choosing labor supply are less than 0.8% of consumption on average. The corresponding minimum-δ point estimate of the structural elasticity is $\varepsilon_{\delta\text{-min}} = 0.50$.

The calculations above ignore statistical imprecision in the point estimates of $\widehat{\varepsilon}$. Columns 8-9 of Table 1 show a 95% confidence set for the ε implied by each study. These columns use the lower endpoint of the 90% confidence interval (CI) for $\widehat{\varepsilon}$ to calculate ε_L and the upper endpoint of the 90% CI to calculate ε_U (Imbens and Manski 2004). In many cases, the 95% confidence sets are only slightly wider than the bounds obtained when ignoring sampling error. For instance, ε_U for Gelber's study rises from 0.54 to 0.6. Pooling the twenty studies, the unified lower bound constructed from the lower endpoints of the 95% CI's (the largest ε_L in Column 8) is 0.32 when $\delta = 1\%$. The analogous unified upper bound constructed from the upper endpoints of the 95% CI's is (the smallest ε_U in Column 9) is 0.6. While these calculations suggest that accounting for statistical imprecision would widen the bounds modestly, the unified bounds of $(0.32, 0.6)$ do not provide a 95% confidence set for ε because of the order-statistics problem in inference with multiple moment inequalities noted above.[26] The dashed blue lines in Figure 11 display analogous calculations of the unified bounds as a function of δ. The set of ε consistent with the observed elasticities expands modestly if one uses the lower and upper endpoints of the 90% CI's for $\widehat{\varepsilon}$ to calculate the bounds on ε in lieu of the point estimates of $\widehat{\varepsilon}$.

Overall, these calculations suggest that the greater source of imprecision about ε is uncertainty about the economic model of behavior due to frictions rather than noise due to sampling error. Many recent studies in Table 1 obtain extremely precise estimates of $\widehat{\varepsilon}$ by pooling multiple reforms or using large micro datasets, but are not much more informative about ε than prior work. To obtain more information about ε, it would be best to examine steady state responses to large tax changes that involve minimal frictions – e.g. examining the effects of a widely publicized tax policy on a subgroup with adjustable labor supply.

4.4 Reconciliation of Micro and Macro Elasticities

I now build on the preceding findings to tackle the puzzle of why micro elasticity estimates are smaller than macro elasticities. I begin by summarizing the key stylized facts to be explained. The macro elasticity is the elasticity of *aggregate* labor supply with respect to net-of-tax rates. Let h denote average hours worked conditional on working and N the number of individuals who work. Then aggregate labor supply is $L = Nh$ and the macro elasticity is $\eta \equiv \frac{d\log L}{d\log(1-\tau)} = \frac{d\log N}{d\log(1-\tau)} + \frac{d\log h}{d\log(1-\tau)} = \varepsilon_{\text{ext}} + \varepsilon$, where ε_{ext} and ε are the structural extensive and intensive margin elasticities in the nominal labor supply model specified in section 4.2.

Intensive and extensive margin elasticities have both been estimated using microeconometric methods (e.g. tax reforms and lifecycle wage variation) and macro methods (e.g. cross-country comparisons and trends). Let $\widehat{\varepsilon}^m$ denote the observed intensive margin elasticity estimated using micro methods and $\widehat{\varepsilon}^M$ denote the observed intensive margin elasticity estimated using macro methods. Similarly, let $\widehat{\varepsilon}^m_{\text{ext}}$ and $\widehat{\varepsilon}^M_{\text{ext}}$ denote the observed extensive margin elasticities in micro and macro studies. Finally, let $\widehat{\eta}$ denote the observed macro elasticity estimated using macro methods. The stylized facts on micro and macro elasticities that we

[26] A 95% confidence set for the unified bounds could be constructed numerically using the methods proposed by Chernozhukov, Hong, and Tamer (2007).

seek to explain are the following:

1. $[\widehat{\eta} > \widehat{\varepsilon}^m]$ Microeconometric estimates of intensive margin elasticities ($\widehat{\varepsilon}^m$) are much smaller than observed macro elasticities. Davis and Henrekson (2005) estimate an elasticity of $\widehat{\eta} = 0.6$ using cross-country variation in tax rates and Prescott (2004) and Ohanian et al. (2008) find that elasticities near 1 fit time series trends in total hours worked across countries with different changes in tax rates. In contrast, the mean value of $\widehat{\varepsilon}^m = 0.12$ among the microeconometric studies in Table 1, if one excludes studies of top income earners, who constitute a small fraction of hours in the economy.

2. $[\widehat{\varepsilon}^M_{\text{ext}} = \widehat{\varepsilon}^m_{\text{ext}}]$ Micro and macro studies agree on the magnitude of extensive margin elasticities. Nickell (2003) and Davis and Henrekson (2005) use data on labor force participation across countries to estimate $\widehat{\varepsilon}^M_{\text{ext}} = 0.2$. Microeconometric studies of low income individuals, secondary earners, and retirees also find observed extensive margin elasticities around $\widehat{\varepsilon}^m_{\text{ext}} = 0.2$ (Gruber and Wise 1999, Coile and Gruber 2000, Eissa and Hoynes 2004, Blau and Khan 2007).

3. $[\widehat{\varepsilon}^M > \widehat{\varepsilon}^m]$ Facts 1 and 2 imply that most of the difference between micro and macro elasticities is driven by difference in intensive margin responses. As shown in Table 1, macro estimates of intensive-margin elasticities are much larger than micro estimates both in the traditional labor supply literature and the more recent taxable income literature for all but very top incomes.

Explaining the Facts. Rogerson and Wallenius (2009) show that small intensive margin elasticities ε are consistent with large macro elasticities η if the extensive margin elasticity ε_{ext} is large enough. This insight – namely that macro elasticities combine extensive and intensive margin responses – is clearly central for understanding fact 1. It does not, however, directly explain facts 2 and 3. In particular, the extensive margin elasticities that the Rogerson and Wallenius simulations require to explain the gap between micro and macro elasticities are considerably larger than both micro and macro estimates of $\widehat{\varepsilon}_{\text{ext}}$. Moreover, as they note, their model does not explain the differences between micro and macro intensive margin elasticity estimates.

Optimization frictions can explain these facts. Micro studies examine short-run responses to tax changes, whereas macro studies focus on long-run or steady state responses. Frictions such as job switching costs, status quo biases, or inattention can lead to delayed responses and attenuate short-run elasticity estimates. For instance, suppose 10% of agents in the economy have low job switching costs each year and the remainder have large switching costs. Then micro studies that examine changes in labor supply over a one year horizon could underestimate the long-run elasticity by an order of magnitude. Moreover, micro studies typically focus on relatively small tax changes that apply to subgroups of the population, whereas macro studies examine larger economy-wide changes. Large, economy-wide changes are more likely to overcome frictions because they induce coordinated changes in work patterns (Chetty et al. 2009).

Frictions can attenuate short-run intensive margin elasticities substantially because the utility gain from reoptimizing labor supply in response to tax changes on the intensive-margin is second-order. The bounds in Table 1 show that frictions of 1% of consumption can generate the observed quantitative differences between micro and macro elasticities on the intensive margin. But the same 1% frictions do not affect short-run estimates of $\widehat{\varepsilon}^m_{\text{ext}}$ significantly because the gains from reoptimizing on the extensive margin are first-order, exceeding 10% of consumption for some of the tax changes studied in the micro literature. This explains why micro and macro studies find similar elasticities on the extensive margin: $\widehat{\varepsilon}^m_{\text{ext}} = \widehat{\varepsilon}^M_{\text{ext}} = \varepsilon_{\text{ext}} = 0.2$. Thus, combining frictions with the insight that the macro elasticity η incorporates both ε_{ext} and ε reconciles all three facts.[27] This synthesis of micro and macro evidence suggests that $\eta \simeq 0.7$ when $\delta = 1\%$.

Implications for the Frisch Elasticity. The analysis above reconciles micro and macro estimates of the Hicksian (income-constant) elasticity of labor supply. While the Hicksian elasticity is relevant for welfare analysis and predicting long-run effects of tax policies, the Frisch (marginal utility constant) elasticity is more relevant for understanding business-cycle fluctuations. An analogous puzzle exists in the context of Frisch elasticities: micro studies (e.g., MaCurdy 1981, Altonji 1982, Ziliak and Kniesner 1999, Blundell, Duncan, and Meghir 1998) find much smaller elasticities than macro studies (e.g., Kydland and Prescott 1982, King and Rebelo 1999, Chang and Kim 2007).

Can optimization frictions reconcile the differences in observed Frisch elasticities? At first brush, the answer may appear to be no. The mean fluctuations in wage rates at business

[27] The explanation here does not contradict that offered by Rogerson and Wallenius (2009). Rather, it shows that one must incorporate additional factors to fully explain the difference between micro and macro elasticities, as suggested by Rogerson (2006).

cycle frequencies are small and temporary, as in micro studies, and thus would appear to be no more likely to overcome frictions. However, this simple view neglects two important factors: heterogeneity and extensive margin responses. Although mean changes in wage rates are small, recessions reduce wage rates for certain individuals very sharply. Frictions could explain why such large concentrated shocks generate substantial changes in intensive margin labor supply while the smaller fluctuations in wage rates over the life cycle used in micro studies do not. Moreover, most of the change in hours over the business cycle is driven by changes in labor force participation (Coleman 1984, Rogerson 1988, Cho and Cooley 1994). Frictions could explain why most of the observed response to fluctuations in wages over the business cycle is on the extensive rather than the intensive margin.

I defer a quantitative exploration of whether frictions can reconcile differences in observed Frisch elasticities across micro and macro studies to future work.[28] Instead, I conclude by calculating the structural intensive margin Frisch elasticity that should be used in such analyses. What can be learned about the structural Frisch elasticity ε^F from the bounds obtained on the Hicksian elasticity ε^c above?

The answer to this question depends upon the assumptions one imposes on preferences. Under the quasilinear utility specification in (25), there are no income effects and $\varepsilon^F = \varepsilon^c$. In an intertemporal labor supply model with arbitrary time-separable utility of the form in (24), the Frisch labor supply elasticity is related to the Hicksian elasticity by the following equation (Ziliak and Kniesner 1999, Browning 2005):

$$\varepsilon^F = \varepsilon^c + \rho(\frac{d[wl^*_{i,t}]}{dY_{i,t}})^2 \frac{Y_{i,t}}{wl^*_{i,t}} \tag{34}$$

where ρ is the elasticity of intertemporal substitution (EIS), $\frac{d[wl^*_{i,t}]}{dY_{i,t}}$ measures the effect of non-wage (unearned) income on labor supply (the marginal propensity to earn out of unearned income), and $\frac{Y_{i,t}}{wl^*_{i,t}}$ is the ratio of non-wage income to wage income. This equation shows that $\varepsilon^F > \varepsilon^c$, a restriction generated by optimization with concave utilities. Therefore we know $\varepsilon^F > 0.47$ using the lower bound on the Hicksian elasticity from Table 1 without placing any restrictions on the flow utility function.

[28] Chang and Kim (2007) and Gourio and Noual (2009) show that models with heterogeneity and extensive-margin responses can generate substantial differences between micro and macro elasticities. Adding frictions to such models could potentially help explain why (1) most of the response is on the extensive margin and (2) the intensive-margin macro response is non-trivial even though most micro studies find intensive-margin elasticities very close to zero.

One can obtain more information about ε^F by calibrating the other parameters in (34). For plausible parameter values, the structural intensive margin Frisch elasticity turns out be quite close to the structural Hicksian elasticity of $\varepsilon^c = 0.5$. In the aggregate economy, the ratio of non-wage to wage income is approximately $\frac{Y_{i,t}}{wl^*_{i,t}} = \frac{1}{2}$. Table 2 shows the values of the Frisch elasticity implied by a Hicksian elasticity of $\varepsilon^c = 0.5$ and $\frac{Y_{i,t}}{wl^*_{i,t}} = \frac{1}{2}$ for various combinations of ρ and $-\frac{d[wl^*_{i,t}]}{dY_{i,t}}$. To calibrate these two parameters, note that balanced growth requires that income and substitution effects cancel, implying $\frac{d[wl^*_{i,t}]}{dY_{i,t}}\frac{Y_{i,t}}{wl^*_{i,t}} = \varepsilon^c \Rightarrow \frac{d[wl^*_{i,t}]}{dY_{i,t}} = -0.25$. Both micro and macro studies find an EIS of $\rho \leq 1$ (Hall 1988, Vissing-Jorgensen 2002, Guvenen 2006). The largest Frisch elasticity consistent with these parameters is $\varepsilon^F = 0.63$, a value that can explain observed business cycle fluctuations in labor supply in a model with search frictions (Hall 2009).[29]

5 Conclusion

There are many frictions that induce agents to deviate from the optimal choices predicted by standard economic models. This paper has shown that the model mis-specification that arises from the omission of these frictions can be handled using the tools of set identification. Abstractly, I exchange the standard orthogonality condition on the error term for a bounded support condition based on the utility costs of errors. I derive an analytical representation for bounds on structural price elasticities that is a function of the observed elasticity, size of the price change used for identification, and the degree of optimization frictions.

Applying the bounds to the literature on taxation and labor supply offers a critique and synthesis of this literature. The critique is that many microeconometric studies of labor supply are uninformative about the labor supply elasticity because they cannot reject very large values of ε with frictions of even 1% of consumption in choosing labor supply. The synthesis is that several patterns in this literature can be reconciled by allowing for such small frictions. Most importantly, frictions can explain the large differences between micro vs. macro elasticities and extensive vs. intensive margin elasticities. Combining estimates from twenty studies yields bounds on the intensive margin compensated (Hicksian) elasticity of $\varepsilon \in (0.47, 0.54)$

[29] This estimate of ε_F does not suffer from Imai and Keane's (2004) critique that microeconometric estimates of ε_F using lifecycle wage variation (e.g. Altonji 1982) are biased downward because they ignore the returns to human capital accumulation. Here, ε_F is identified from studies that use variation in taxes that is orthogonal to returns to human capital.

with frictions of 1% of consumption. A simple calibration suggests Frisch elasticities in a similar range. Even with frictions as large as 5% of consumption, existing estimates imply $\varepsilon < 1.23$. In sum, frictions justify a substantially larger ε than the estimates of most micro studies but not the very large values of ε used to calibrate some macro models.

The bounding methodology developed here can be applied in other contexts where the values of key parameters are debated, such as the elasticity of intertemporal substitution, the marginal propensity to consume out of income, or the effects of the minimum wage on employment. Such analyses would shed light on which disagreements are economically significant 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.

Alesina, Alberto F., Edward L. Glaeser, and Bruce Sacerdote (2005). "Work and Leisure in the U.S. and Europe: Why So Different?" in: Kenneth Rogoff and Mark Gertler (eds.), *NBER Macroeconomics Annual* 2005, pp 1-100. Cambridge: MIT Press

Alesina, Alberto F., Andrea Ichino, and Loukas Karabarbounis (2007). "Gender Based Taxation and the Division of Family Chores," NBER Working Paper No. 13638.

Altonji, Joseph G. (1982). "The intertemporal substitution model of labor market fluctuations: An empirical analysis." *Review of Economic Studies* XLIX: 783–824.

Altonji, Joseph G. and Christina H. Paxson (1992). "Labor Supply, Hours Constraints, and Job Mobility," *Journal of Human Resources,* 27(2): 256-278.

Attanasio, Orazio P. (2000). "Consumer Durables and Inertial Behaviour: Estimation and Aggregation of (S,s) Rules for Automobile Purchases," *Review of Economic Studies*, 67(4): 667-696: 319-337.

Auten, Gerald and Robert Carroll (1999). "The Effect of Income Taxes on Household Behavior," *Review of Economics and Statistics*, 81: 681-693.

Bianchi, Marco, Björn R. Gudmundsson, and Gylfi Zoega (2001). "Iceland's Natural Experiment in Supply-Side Economics," *American Economic Review*, 91(5): 1564-1579.

Blau, Francine D. and Lawrence M. Kahn (2007). "Changes in the Labor Supply Behavior of Married Women: 1980-2000," *Journal of Labor Economics*, 25: 393-438.

Blomquist, Soren and Whitney Newey (2002). "Nonparametric Estimation with Nonlinear Budget Sets," *Econometrica* 70(6): 2455-2480.

Blundell, Richard and Thomas MaCurdy (1999). "Labor Supply: A Review of Alternative Approaches," in Orley Ashenfelter and David Card (eds.), *Handbook of Labor Economics.* Vol. 3. Amsterdam: North-Holland.

Blundell, Richard, Alan Duncan, and Costas Meghir (1998). "Estimating Labor Supply Responses Using Tax Reforms," *Econometrica*, 66(4): 827-862.

Blundell, Richard, Amanda Gosling, Hidehiko Ichimura, and Costas Meghir (2007). "Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds," *Econometrica* 75(2): 323-363.

Browning, Martin (2005). "A working paper from April 1985: Which demand elasticities do we know and which do we need to know for policy analysis?" *Research in Economics*, 59(4): 293-320.

Caballero, Ricardo, Eduardo Engel, and John Haltiwanger (1995). "Plant Level Adjustment and Aggregate Investment Dynamics,"*Brookings Papers on Economic Activity* 2, Macroeconomics: 1-54.

Chang, Yongsung and Sun-Bin Kim (2007). "Heterogeneity and Aggregation: Implications for Labor-Market Fluctuations," *American Economic Review*, 97(5): 1939-1956.

Chernozhukov, Victor, Han Hong, and Elie Tamer (2007). "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75(5): 1243-1284.

Chetty, Raj (2009). "Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply," NBER Working Paper No. 15616.

Chetty, Raj, Adam Looney and Kory Kroft (2009). "Salience and Taxation: Theory and Evidence," *American Economic Review* 9(4): 1145-1177.

Chetty, Raj and Emmanuel Saez (2009). "Teaching the Tax Code: Earnings Responses to a Field Experiment with EITC Recipients," NBER Working Paper No. 14836.

Chetty, Raj, John Friedman, Tore Olsen, and Luigi Pistaferri (2009). "Adjustment Costs, Firm Responses, and Labor Supply Elasticities: Evidence from Danish Tax Records," NBER Working Paper No. 15617.

Cho, Jang-Ok and Cooley, Thomas F. (1994). "Employment and Hours Over the Business Cycle," *Journal of Economic Dynamics and Control*, 18(2): 411-432.

Cochrane, John (1989). "The Sensitivity of Tests of the Intertemporal Allocation of Consumption to Near-Rational Alternatives" *American Economic Review*, 79: 319-337.

Coile, Courtney and Jonathan Gruber (2000). "Social Security and Retirement," NBER Working Papers No. 7830.

Coleman, Thomas Sedgwick (1984). "Essays on Aggregate Labor Markets Business Cycle Fluctuations," unpublished dissertation, University of Chicago.

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.

Del Boca, Daniela, and Annamaria Lusardi (2003). "Credit Market Constraints and Labor Market Decisions," *Labour Economics,* 10(6): 681-703.

DellaVigna, Stefano (2009). "Psychology and Economics: Evidence from The Field," *Journal of Economic Literature* 47: 315-372.

DellaVigna, Stefano and Gentzkow, Matthew (2009). "Persuasion: Empirical Evidence," forthcoming *Annual Review of Economics.*

Eberly, Janice C. (1994). "Adjustment of Consumers' Durables Stocks: Evidence from Automobile Purchases," *Journal of Political Economy*, 102(3): 403-436.

Eissa, Nada and Hilary Hoynes (1998). "The Earned Income Tax Credit and the Labor Supply of Married Couples," NBER Working Paper 6856 (December).

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.

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.

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," University of Pennsylvania working paper.

Gourio, François and Pierre-Alexandre Noual (2009). "The Marginal Worker and the Aggregate Elasticity of Labor Supply," Boston University Dept. of Economics Working Papers Series, WP2006-009.

Goolsbee, Austan (1999). "Evidence on the high-income Laffer curve from six decades of tax reform," *Brookings Papers on Economic Activity*, 2: 1-47.

Grossman, Sanford J. and Guy Laroque (1990). "Asset Pricing and Optimal Portfolio Choice in the Presence of Illiquid Durable Consumption Goods," *Econometrica*, 58(1): 25-51.

Gruber, Jonathan and David A. Wise (1999). *Social Security and Retirement Around the World*, National Bureau of Economic Research, Inc.

Gruber, Jonathan and Emmanuel Saez (2002). "The Elasticity of Taxable Income: Evidence and Implications," *Journal of Public Economics,* 84: 1–32.

Guvenen, Fatih (2006). "Reconciling Conflicting Evidence on the Elasticity of Intertemporal Substitution: A Macroeconomic Perspective," *Journal of Monetary Economics*, 53(7): 1451-1472.

Hall, Robert E. (1988). "Intertemporal Substitution in Consumption," *Journal of Political Economy*, 96(2): 339-357.

Hall, Robert E. (2009). "Reconciling Cyclical Movements in the Marginal Value of Time and the Marginal Product of Labor," *Journal of Political Economy* 117(2): 281-323.

Hansen, Lars Peter and Thomas J. Sargent (2007). *Robustness*, Princeton: Princeton University Press.

Hausman, Jerry A. (1985). "Taxes and Labor Supply," in Alan Auerbach and Martin Feldstein (eds.), *Handbook of Public Economics* Vol. 1, 213-263, Amsterdam: North-Holland.

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.

Imai, Susumu and Michael P. Keane (2004). "Intertemporal Labor Supply and Human Capital Accumulation." *International Economic Review* 45(2): 601-641.

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.

King, Robert G. and Sergio T. Rebelo (1999). "Resuscitating Real Business Cycles," in: J.B. Taylor and M. Woodford (eds.), *Handbook of Macroeconomics*, 927-1007, Amsterdam: North-Holland.

Kleven, Henrik and Esben A. Schultz (2009). "Estimating the Elasticity of Reported Income Using a Large Panel of Danish Taxpayers" LSE Working Paper.

Kleven, Henrik, Claus Kreiner, and Emmanuel Saez (2009). “The Optimal Income Taxation of Couples,” *Econometrica*, 77(2): 537-560.

Kopczuk, Wojciech (2009). “Polish Business Flat Tax and its Effect on Reported Incomes,” Columbia University mimeo.

Kydland, Finn and Edward Prescott (1982). “Time to Build and Aggregate Fluctuations,” *Econometrica*, 50(6): 1345-1370

Ljungqvist, Lars and Thomas J. Sargent (2006). “Do Taxes Explain European Employment? Indivisible Labor, Human Capital, Lotteries, and Savings.” in: Daron Acemoglu, Kenneth Rogoff and Michael Woodford (eds.), *NBER Macroeconomics Annual* 2006, 181-246. Cambridge: MIT Press

MaCurdy, Thomas E. (1980). “An Empirical Model of Labor Supply in a Life-Cycle Setting,” NBER Working Paper No. 421.

MaCurdy, Thomas (1981). “An Empirical Model of Labor Supply in a Life-Cycle Setting,” *Journal of Political Economy,* 89: 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.

Mankiw, N. Gregory (1985). “Small Menu Costs and Large Business Cycles: A Macroeconomic Model,” *Quarterly Journal of Economics*, 100(2): 529–38.

Manski, Charles F. (2007). *Identification for Prediction and Decision*, Cambridge: Harvard University Press.

Meyer, Bruce and Dan Rosenbaum (2001). “Welfare, the Earned Income Tax Credit, and the Labor Supply of Single Mothers,” *Quarterly Journal of Economics*, 116(3): 1063-1114.

Mullainathan, Sendhil, Joshua Schwartzstein, and Andrei Shleifer (2008). “Coarse Thinking and Persuasion,” *Quarterly Journal of Economics*, 123(2): 577-619.

Nelson, Phillip (1974). “Advertising as Information,” *Journal of Political Economy*, 82(4): 729-254.

Nevo, Aviv and Adam Rosen (2008). “Identification with Imperfect Instruments,” NBER Working Paper No. 14434.

Nickell, Stephen (2003). “Employment and Taxes,” CESifo Working Paper No. 1109.

Ohanian, Lee, Andrea Raffo, and Richard Rogerson (2008). “Long-term changes in labor supply and taxes: Evidence from OECD countries, 1956-2004,” *Journal of Monetary Economics*, 55(8): 1353–1362.

Pakes, Ariel, Jack Porter, Katherine Ho, and Joy Ishii (2007). “Moments Inequalities and Their Application.” Harvard University mimeo.

Pencavel, John (1986). “Labor Supply of Men: A Survey.” in Orley Ashenfelter and Richard Layard (eds.), *Handbook of Labor Economics* Vol. 1. Amsterdam: North Holland.

Prescott, Edward (2004). “Why Do Americans Work So Much More Than Europeans?” *Federal Reserve Bank of Minneapolis Quarterly Review*, 28(1): 2–13.

Rogerson, Richard (1988). “Indivisible Labor, Lotteries and Equilibrium,” *Journal of Mone-*

tary Economics, 21(1): 3–16.

Rogerson, Richard (2006). "Understanding Differences in Hours Worked," *Review of Economic Dynamics*, 9: 365-409.

Rogerson, Richard and Johanna Wallenius (2009). "Micro and Macro Elasticities in a Life Cycle Model with Taxes," *Journal of Economic Theory*, 144(6): 2277-2292.

Saez, Emmanuel (2001). "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 68: 205-229.

Saez, Emmanuel (2002). "Do Taxpayers Bunch at Kink Points?" UC-Berkeley mimeo.

Saez, Emmanuel (2004). "Reported Incomes and Marginal Tax Rates, 1960-2000: Evidence and Policy Implications," *Tax Policy and the Economy* 18: 117-174. Cambridge: MIT Press.

Saez, Emmanuel (2009). "Do Taxpayers Bunch at Kink Points?" *American Economic Journal: Economic Policy*, forthcoming.

Saez, Emmanuel, Joel Slemrod and Seth Giertz (2009). "The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review," NBER Working Paper No. 15012.

Samuelson, William and Richard Zeckhauser. (1988). "Status Quo Bias in Decision Making." *Journal of Risk and Uncertainty* 1: 7-59.

Simon, Herbert A. (1957). *Models of Man*. New York: Wiley.

Sims, Christopher A. (2003). "Implications of rational inattention," *Journal of Monetary Economics*, 50(3): 665-690.

Slemrod, Joel B. (1995). "Income Creation or Income Shifting? Behavioral Responses to the Tax Reform Act of 1986," *American Economic Review*, 82(5): 175-180.

Stigler, George J. (1961). "The Economics of Information," *Journal of Political Economy*, 69(3): 213-225.

Trabandt, Mathias and Harald Uhlig (2009). "How Far Are We From The Slippery Slope? The Laffer Curve Revisited," NBER Working Paper No. 15343.

Vissing-Jorgensen, Annette (2002). "Limited Asset Market Participation and the Elasticity of Intertemporal Substitution," *Journal of Political Economy*, 110(4): 825-853.

Yelowitz, Aaron S. (1995). "The Medicaid Notch, Labor Supply, and Welfare Participation: Evidence from Eligibility Expansions," *Quarterly Journal of Economics*, 110(4): 909-39.

Ziliak, James P. and Thomas J. Kniesner (1999). "Estimating Life Cycle Labor Supply Tax Effects," *Journal of Political Economy*, 107(2): 326-359.

Appendix A: Bounds on Elasticities with Income Effects

This appendix shows that the bounds in Proposition 1 apply to the Hicksian elasticity when utility does not take the specific quasilinear form in (2). For simplicity, I ignore heterogeneity across agents and assume all agents have a flow utility function $v(x_t, y_t)$. When utility is not quasi-linear, the δ class of models is defined using the expenditure function. The agent's expenditure function is

$$e(p,U) = \min \sum_{s=t}^{T} (p_s x_s + y_s) \text{ s.t. } \sum_{s=t}^{T} v(x_s, y_s) \geq U$$

Let the optimal choice of x_t be denoted by $x_t^c(p,U)$, the Hicksian demand function. Let the agent's constrained expenditure function conditional on consuming $\widetilde{x}_t$ units of good x_t be denoted by

$$\widetilde{e}(\widetilde{x}_t) = \min \sum_{s=t}^{T} (p_s x_s + y_s) \text{ s.t. } \sum_{s=t}^{T} v(x_s, y_s) \text{ and } x_t = \widetilde{x}_t \tag{35}$$

The δ class of models is defined by the condition

$$\widetilde{e}(x_t) - \widetilde{e}(x_t^c) < \delta p_t x_t^c$$

Note that this definition collapses to that in (11) when utility is quasilinear. I first establish an analog of Lemma 1 with income effects.

Lemma A1. For small δ, the set of observed demands is approximately

$$X_t(p_t, \delta) = \{x_t : |\log x_t - \log x_t^c| < [2\varepsilon^c(p_t)\delta]^{1/2}\} \tag{36}$$

where x_t^c denotes the Hicksian demand function and $\varepsilon^c(p_t) = -\frac{\partial x^c}{\partial p_t}\frac{p_t}{x^c}$ denotes the compensated (Hicksian) price elasticity of demand at price p_t in period t.

Proof. Using a quadratic approximation to the constrained expenditure function and recognizing that $\widetilde{e}_{\widetilde{x}_t}(x_t^c) = 0$ from agent optimization, we obtain

$$\widetilde{e}(x_t) - \widetilde{e}(x_t^c) \simeq \frac{1}{2}(x_{i,t}^c)^2(\log x - \log x_t^c)^2 e_{xx}(x_t^c)$$

and hence we can rewrite (36) as

$$|\log x - \log x_t^c| < [2\delta \frac{p_t}{x_t^c}\frac{1}{\widetilde{e}_{xx}(x_t^c)}]^{1/2}$$

The remainder of the proof establishes that $\frac{\partial x^c}{\partial p_t} = -1/\widetilde{e}_{\widetilde{x}_t\widetilde{x}_t}(x_t^c)$. Let λ_1 denote the multiplier on the utility constraint and λ_2 the multiplier on the $x_t = \widetilde{x}_t$ constraint in (35). Then

$$\widetilde{e}_{\widetilde{x}_t} = \frac{\partial L}{\partial \widetilde{x}_t} = \lambda_2 \Rightarrow \widetilde{e}_{\widetilde{x}_t\widetilde{x}_t} = \frac{\partial \lambda_2}{\partial \widetilde{x}_t}$$

The first order condition for agent optimization of x_t in (35) is

$$p_t = \lambda_1 v_x(\widetilde{x}_t) + \lambda_2 \tag{37}$$

Hence

$$\widetilde{e}_{\widetilde{x}_t\widetilde{x}_t} = \frac{\partial \lambda_2}{\partial \widetilde{x}_t} = -\lambda_1 v_{xx} - \frac{\partial \lambda_1}{\partial \widetilde{x}_t} v_x$$

Next, consider the constrained expenditure minimization problem when $\widetilde{x}_t(p_t) = x_t^c(p_t)$. In this case $\lambda_2 = 0$ and implicitly differentiating (37) w.r.t. p_t yields

$$\begin{aligned} 1 &= \lambda_1 v_{xx} \frac{\partial x^c}{\partial p_t} - \frac{d\lambda_1}{dp_t} v_x \\ &= \lambda_1 v_{xx} \frac{\partial x^c}{\partial p_t} - \left(\frac{\partial \lambda_1}{\partial \widetilde{x}_t}\frac{\partial x_t^c}{\partial p_t} + \frac{\partial \lambda_1}{\partial p_t}|\widetilde{x}_t\right) v_x \end{aligned}$$

The solution to the constrained expenditure minimization problem is unaffected by p_t when x_t is fixed and hence $\frac{\partial \lambda_1}{\partial p_t}|\widetilde{x}_t = 0$. Therefore the preceding condition implies

$$1 = (\lambda_1 v_{xx} - \frac{\partial \lambda_1}{\partial \widetilde{x}_t} v_x)\frac{\partial x^c}{\partial p_t}$$

It follows that $\frac{\partial x^c}{\partial p_t} = -1/\widetilde{e}_{\widetilde{x}_t\widetilde{x}_t}(x_t^c)$, completing the proof.

Next, I establish the analog of Proposition 1. I bound the Hicksian elasticity at a particular price point p_t using observed elasticities around that price level because $\varepsilon^c(p_t)$ may vary with p_t when $v(x, y)$ is left unrestricted. Because the Hicksian elasticity may also vary with the period t, we must consider a price change *within* a given period t to obtain a well defined elasticity. Consider an experiment that raises the price of x_t from p_A to p_B in period t while providing the agent with income compensation so that his utility remains unchanged. Then the lower and upper bounds on the observed Hicksian price elasticities remain as depicted in Figure 3, replacing the Marshallian demand curve $x(p_t)$ with the Hicksian demand curve $x^c(p_t)$. If $\varepsilon^c(p_t)$ does not vary locally between p_A and p_B, the upper and lower bounds on $\varepsilon^c(p_t)$ are characterized by equations (16) and (17):

$$\widehat{\varepsilon}^c(p_t) = \varepsilon^c(p_t) \pm 2\frac{(2\varepsilon_L \delta)^{1/2}}{\Delta \log p}$$

Proposition A1. For any flow utility function $v(x, y)$, under assumption A1, for small δ, the range of structural Hicksian elasticities consistent with an observed Hicksian elasticity $\widehat{\varepsilon}^c(p_t)$ is approximately $(\varepsilon_L^c(p_t), \varepsilon_U^c(p_t))$ where

$$\begin{aligned} \varepsilon_L^c(p_t) &= \widehat{\varepsilon}^c(p_t) + \frac{4\delta}{(\Delta \log p)^2}(1-\rho) \text{ and } \varepsilon_U^c(p_t) = \widehat{\varepsilon}^c(p_t) + \frac{4\delta}{(\Delta \log p)^2}(1+\rho) \qquad (38) \\ \text{with } \rho &= (1 + \frac{1}{2}\frac{\widehat{\varepsilon}^c(p_t)}{\delta}(\Delta \log p)^2)^{1/2} \qquad (39) \end{aligned}$$

Proof. Identical to the proof of Proposition 1.

Appendix B: Sources and Calculations for Studies in Table 1

This appendix describes how the values in columns 3–5 of Table 1 are calculated for each study. The studies do not always directly report the relevant inputs, especially for the net-of-tax change $\Delta \log(1-\tau)$. I adhere to the following principles to obtain consistent values of $\Delta \log(1-\tau)$ across the studies: (1) for studies whose estimates are identified from a single quasi-experiment (e.g. Feldstein 1995), I define $\Delta \log(1-\tau)$ as the change in the NTR for the group that the authors' define as the "treated" group; (2) for studies that pool multiple short run changes and do not explicitly isolate a treatment group (e.g. Gruber and Saez 2002), I define $\Delta \log(1-\tau)$ as the 90th percentile of the distribution of changes in net-of-tax rates in the sample; and (3) for studies that pool time series and cross-sectional variation in the level of the tax rate (e.g. Davis and Henrekson 2005) rather than focusing on short-run changes, I define $\Delta \log(1-\tau)$ as the difference between the NTR at the 10th and 90th percentile in the sample.

A. Hours Elasticities

1. MaCurdy (1981): $\hat{\varepsilon}$: reported in text on page 1083. s.e.($\hat{\varepsilon}$): imputed from the t-statistic for δ reported in row 5 of Table 1 as 0.15/0.98 because estimate of compensated elasticity is approximately equal to δ. $\Delta \log(1-\tau)$: mean change in wage rates $W(t) - W(t-1)$ reported in second block of Table A1 of NBER working paper number 421.

2. Eissa and Hoynes (1998): $\hat{\varepsilon}$: average of the compensated wage elasticities for married men and women with children from 1984 to 1996, computed using the Slutsky equation $\varepsilon_{l^c,w} = \varepsilon_{l,w} - \frac{wl}{y}\varepsilon_{l,y}$, where $\varepsilon_{l,w}$ and $\varepsilon_{l,y}$ are reported in Table 8, IV2 for men and Table 9, IV2 for women. Mean values of y and wl are computed from summary statistics reported in Table 3. s.e.($\hat{\varepsilon}$): calculated from standard errors reported in Table 8, IV2 and Table 9, IV2 for coefficients on log(net wage) and virtual income. $\Delta \log(1-\tau)$: unweighted mean percent change in the net-of-tax rate in the phase-in and phase-out regions of the EITC schedule for the reforms reported in Table 1 (TRA 1986, OBRA 1990, and OBRA 1994).

3. Blundell, Duncan, and Meghir (1998): $\hat{\varepsilon}$, s.e.($\hat{\varepsilon}$): Table 4, row 1. $\Delta \log(1-\tau)$: defined as $\Delta \log(\text{wage} \times \text{NTR}) = \Delta \log(\text{wage}) + \Delta \log(\text{NTR})$ in order to incorporate both wage and tax variation, where $\Delta \log(\text{wage})$ is the mean percentage change in annual wage rate computed from Table 15 and $\Delta \log(\text{NTR})$ is the difference between the 10th and 90th percentiles of log NTR's across all the cohort-year-education cells listed in Table 2.

4. Ziliak and Kniesner (1999): $\hat{\varepsilon}$, s.e.($\hat{\varepsilon}$): Table 1, column 3. $\Delta \log(1-\tau)$: change in net-of-tax rate between the 10th and 90th percentiles, calculated as $\log(1 - .292 + 1.64 \cdot 0.069) - \log(1 - .292 - 1.64 \cdot 0.069)$ where 0.069 is the standard deviation of the tax rates and 0.292 is the mean marginal tax rate reported in row 6 of Table D1.

5. Bianchi, Gudmundsson, and Zoega (2001): $\hat{\varepsilon}$, s.e.($\hat{\varepsilon}$): mean percent change in hours for men and women (columns 3-4 of Table 5) divided by the percent change in the net-of-tax rate. Standard error computed from the standard errors reported for the changes in hours. Note that the elasticity estimates provided by the authors are computed using *average* tax rates, necessitating use of the computation described above. $\Delta \log(1-\tau)$: log change from tax rate of 0 in 1987 to a 35 percent flat tax in 1988 as described in the text.

B. Taxable Income Elasticities

6. Gruber and Saez (2002): $\hat{\varepsilon}$, s.e.($\hat{\varepsilon}$): average of the estimates in column 2 of Table 9 for individuals with taxable income between \$10,000 and \$50,000 and those with taxable income between \$50,000 and \$100,000. $\Delta \log(1-\tau)$: 90th percentile of distribution of changes in net-of-tax rate for individuals with taxable income between \$10,000 and \$100,000, computed as unweighted mean of 90th percentiles implied by means and standard deviations in each cell of columns 3 and 4 of Table 3.

7. Saez (2004): $\hat{\varepsilon}$, s.e.($\hat{\varepsilon}$): Table 4B, column 6 for the top 5 to 1 percent of tax units. $\Delta\log(1-\tau)$: difference between 10th and 90th percentiles of NTR's for the top 5 to 1 percent of tax units listed in column 8 of Table B2.

8, 9. Chetty et al. (2009): $\hat{\varepsilon}$, s.e.($\hat{\varepsilon}$): observed elasticities at middle and top kinks reported in section 5. $\Delta\log(1-\tau)$: size of tax changes at the middle and top tax kinks as reported in section 5.

10. Gelber (2009): $\hat{\varepsilon}$, s.e.($\hat{\varepsilon}$): Table 3, column 2. $\Delta\log(1-\tau)$: percent change in net-of-tax rate from 1989 to 1991 for the highest tax brackets reported in Table 1.

11. Kleven and Schultz (2009): $\hat{\varepsilon}$, s.e.($\hat{\varepsilon}$): reported in text. $\Delta\log(1-\tau)$: change in net-of-tax rate for 1987 Danish tax reform reported in text.

12. Saez (2009): $\hat{\varepsilon}$, s.e.($\hat{\varepsilon}$): Table 2, column 6 for wage earners with two or more children. $\Delta\log(1-\tau)$: change in NTR at first kink in the EITC benefit schedule from 1995 to 2004.

C. Top Income Elasticities

13. Feldstein (1995): $\hat{\varepsilon}$: high minus medium tax rate specification in Table 2. s.e.($\hat{\varepsilon}$): not reported in the paper; computed by rescaling the standard error in Auten and Carroll (1999) by the ratio of sample sizes in the two studies ($\frac{14{,}425}{3{,}735}$) cited by Feldstein on page 566. $\Delta\log(1-\tau)$: reported in Table 2 for the high tax rate group.

14. Auten and Carroll (1999): $\hat{\varepsilon}$, s.e.($\hat{\varepsilon}$): Table 2, Col 6. $\Delta\log(1-\tau)$: change in NTR for high tax rate group in Table 2 of Feldstein (1995).

15. Goolsbee (1999): $\hat{\varepsilon}$, s.e.($\hat{\varepsilon}$): Table 2b, column 1. There is a typographical error in the standard error reported in Table 2b, so I use the standard error reported in the text on page 21 instead. $\Delta\log(1-\tau)$: Table 2b, row C for 1985 to 1989.

16. Saez (2004): $\hat{\varepsilon}$, s.e.($\hat{\varepsilon}$): Table 2C, column 3 for the top 1 percent of tax units. $\Delta\log(1-\tau)$: difference between the 10th and 90th percentiles of NTR's for the top 1 percent of tax units listed in column 3 of Table B2.

17. Kopczuk (2009): $\hat{\varepsilon}$, s.e.($\hat{\varepsilon}$): reported as estimates from preferred specification in text. $\Delta\log(1-\tau)$: change in the predicted marginal net-of-tax rate for individuals predicted to switch to the flat tax regime based on 2002 (pre-reform) earnings levels.

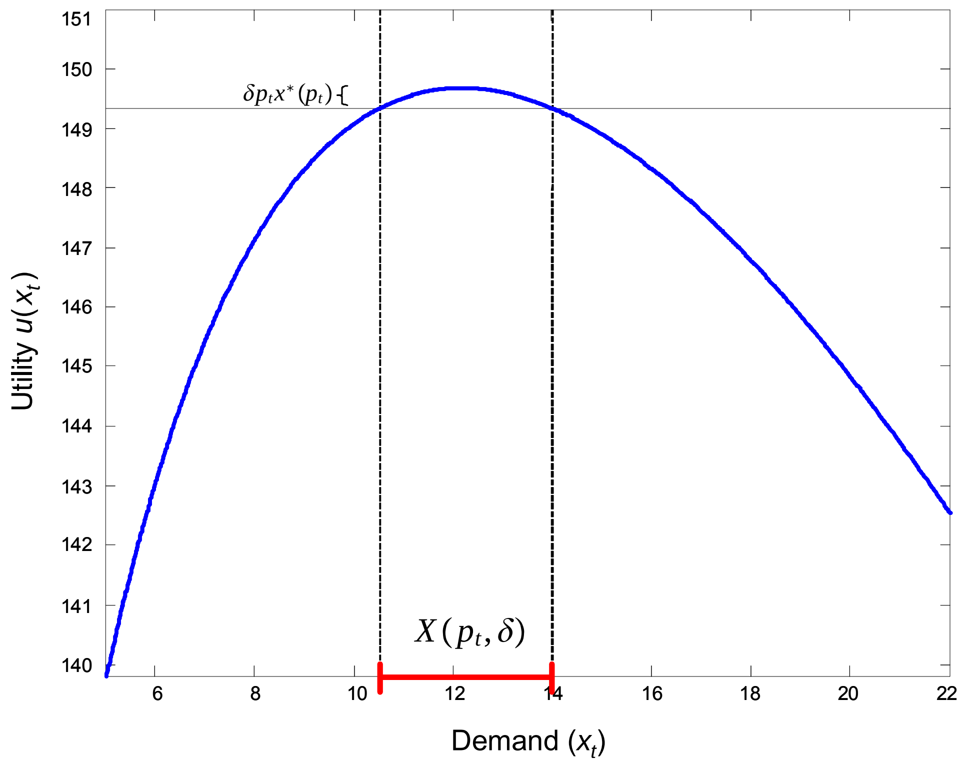
D. Macroeconomic/Trend-Based Elasticities

18. Prescott (2004): $\hat{\varepsilon}$: computed from hours and tax changes reported in Table 2, treating the United Kingdom and United States as control groups. I first compute the differences in log(1−tax rate) between 1993-1996 and 1970-1974 for each country. I then take the differences in the log of predicted hours in 1993-1996 and 1970-1974 for each country and divide by the corresponding percent change in the net-of-tax rate to define a country-specific observed elasticity. Finally, I compute the average of these country-level elasticities (excluding the U.K. and U.S.) to obtain $\hat{\varepsilon}$. s.e.($\hat{\varepsilon}$): not reported because Prescott calibrates a model to fit these data rather than directly estimating $\hat{\varepsilon}$. $\Delta\log(1-\tau)$: computed as the mean absolute value of the country-level differences in log(1−tax rate) as described above, excluding the U.K. and U.S.

19. Davis and Henrekson (2005): $\hat{\varepsilon}$: Figure 2.2 reports $\frac{dl}{d(1-\tau)} = 9.1$; $\hat{\varepsilon} = 9.1\frac{100\%-50.8\%}{1140}$ computed at sample means of $l = 1140$ hours and $\tau = 50.8\%$ reported in Table 2.1 for Sample A. s.e.($\hat{\varepsilon}$): calculated from the standard error reported for the regression coefficient in Figure 2.2. $\Delta\log(1-\tau)$: computed as the change in tax rates from the 10th to the 90th percentile of observations shown in Figure 2.2.

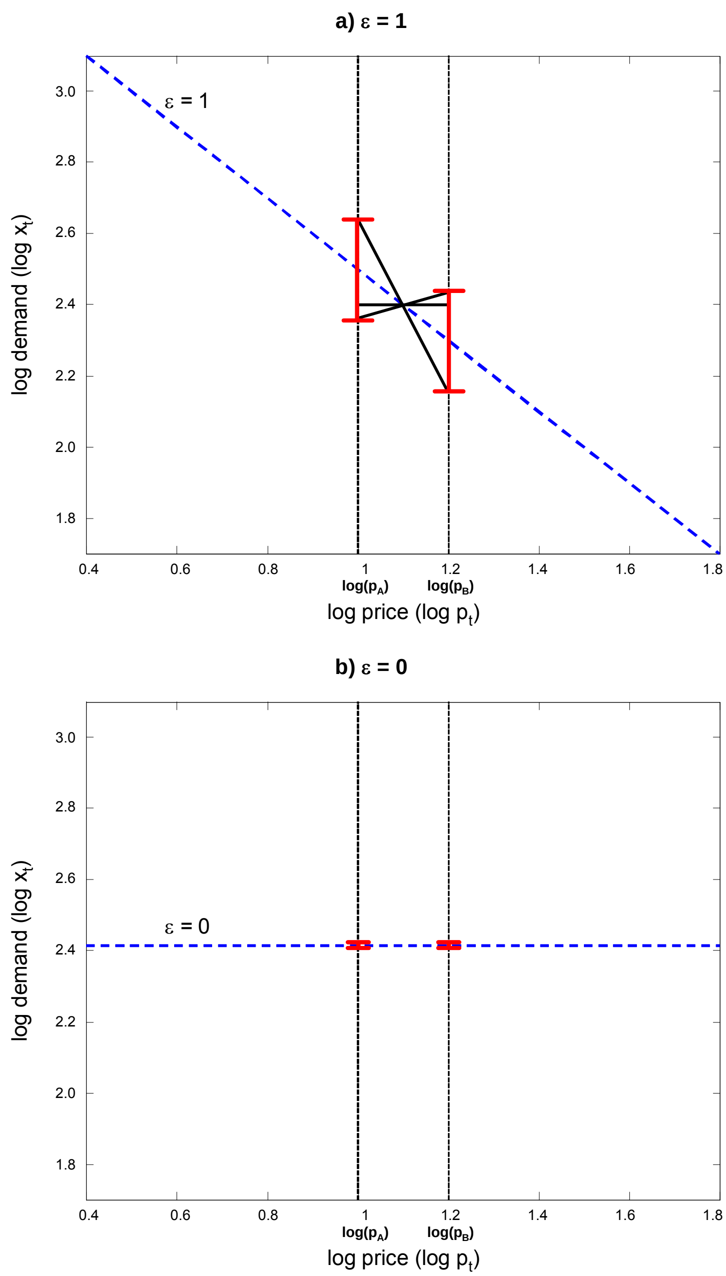
20. Blau and Kahn (2005): $\hat{\varepsilon}$, s.e.$(\hat{\varepsilon})$: average of the elasticities with respect to own log wage reported in column 3 (1989-91 to 1979-81) and column 7 (1999-2001 to 1989-91) of Table 10. $\Delta \log(1-\tau)$: average change in own log imputed wage reported in columns 4 and 5 of Table 2.

FIGURE 1
Choice Set in a δ Class of Models



NOTE–This figure illustrates the choice set $X(p_t, \delta)$ in a δ class of models when all agents have $\delta_{i,t} = \delta = 1\%$ and $a_{i,t} = a = \exp(3.5)$. The blue curve plots flow utility $u(x_t) = 100 - p_t x_t + a \log x_t$ with $\log p_t = 1$. The set of demand levels that yield utility within $0.01 p_t x^*(p_t)$ dollars of the maximum is shown by the red interval on the x axis.

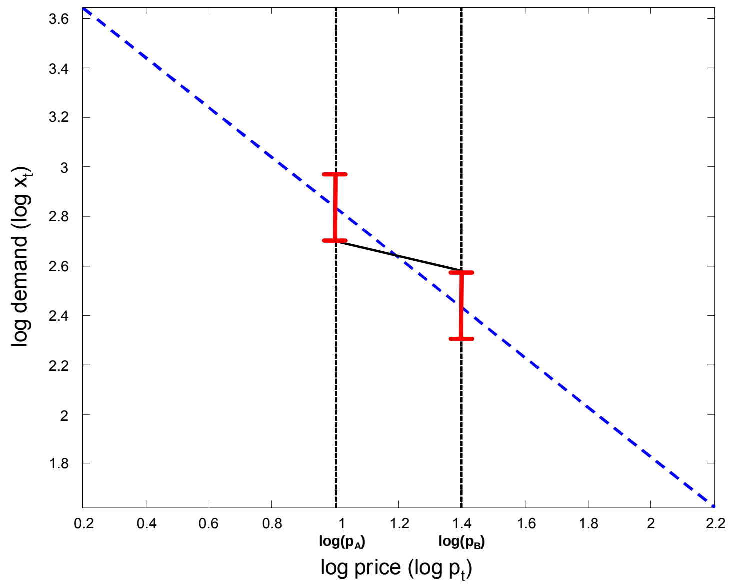
FIGURE 2
Identification with Optimization Frictions



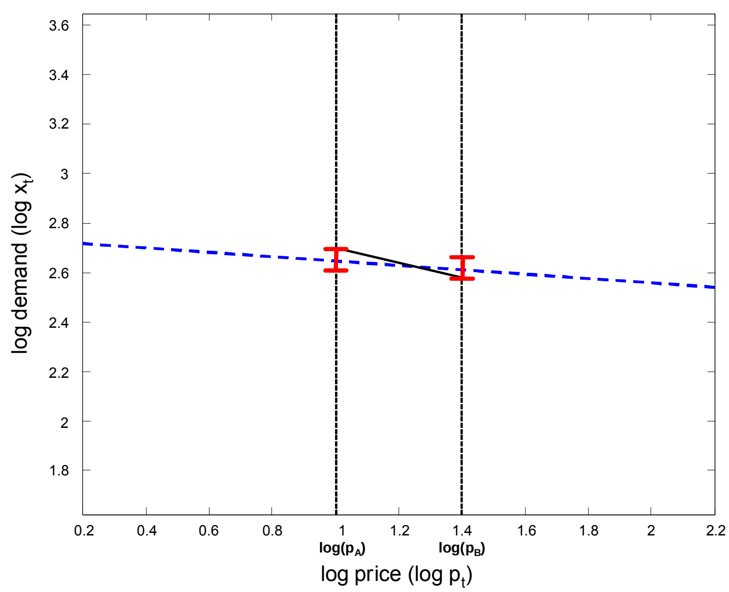
NOTE–This figure plots the choice sets at two price levels, $X(p_A, \delta)$ and $X(p_B, \delta)$, with $\log p_A = 1$ and $\log p_B = 1.2$. In panel A, $\varepsilon = 1$; in panel B, $\varepsilon = 0$. All other parameters are specified as in Figure 1. The dashed blue line shows the optimal demand $x^*(p_t)$. The black lines in Panel A illustrate some of the responses $(\log x_B(p_B) - \log x_A(p_A))$ that may be observed for a price increase from p_A to p_B with a structural elasticity of $\varepsilon = 1$ and frictions of $\delta = 1\%$ of consumption.

FIGURE 3
Bounding the Structural Elasticity with Optimization Frictions

a) Upper Bound on Structural Elasticity

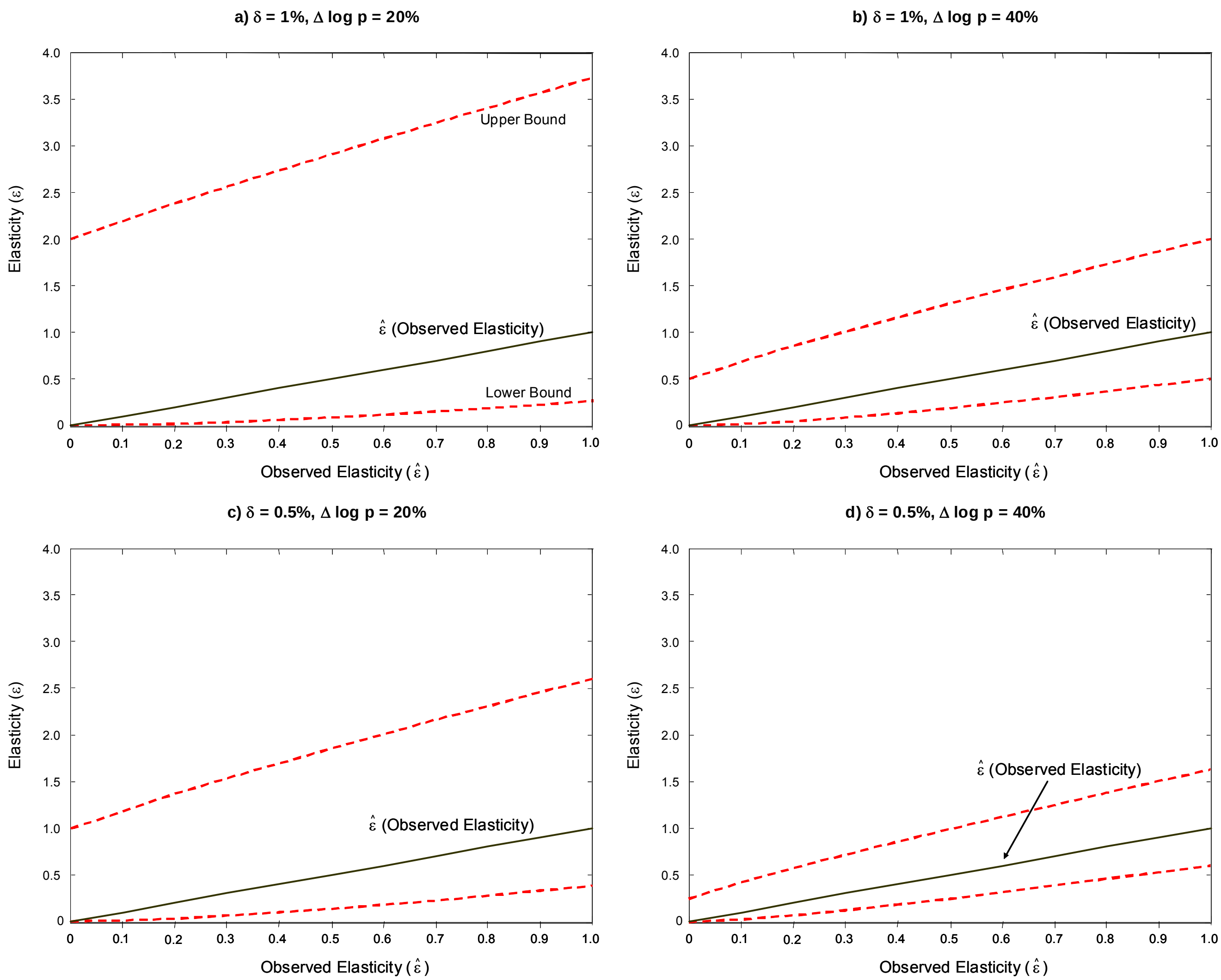


b) Lower Bound on Structural Elasticity



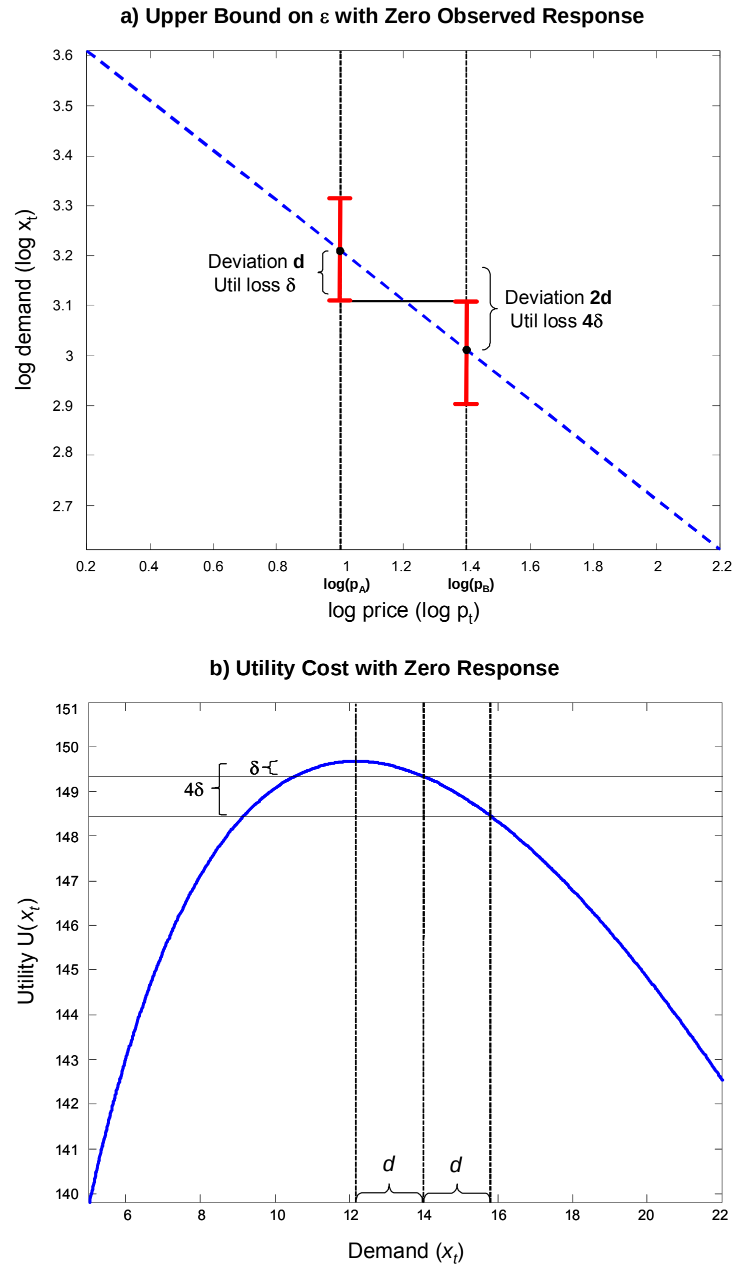
NOTE–The solid black line in each panel depicts the observed demand response for a price increase from p_A to p_B with an observed elasticity $\hat{\varepsilon} = 0.3$, $\log p_A = 1$, and $\log p_B = 1.4$. Panel A depicts the highest structural elasticity, $\varepsilon_U = 1$, that could have generated this observed response with $\delta = 1\%$. The blue dashed line depicts the optimal demand $x^*(p_t)$ with $\varepsilon = 1$. Panel B analogously depicts the lowest structural elasticity, $\varepsilon_L = 0.1$, that could have generated the same observed response.

FIGURE 4
Bounds on Structural Elasticities as a Function of Observed Elasticities



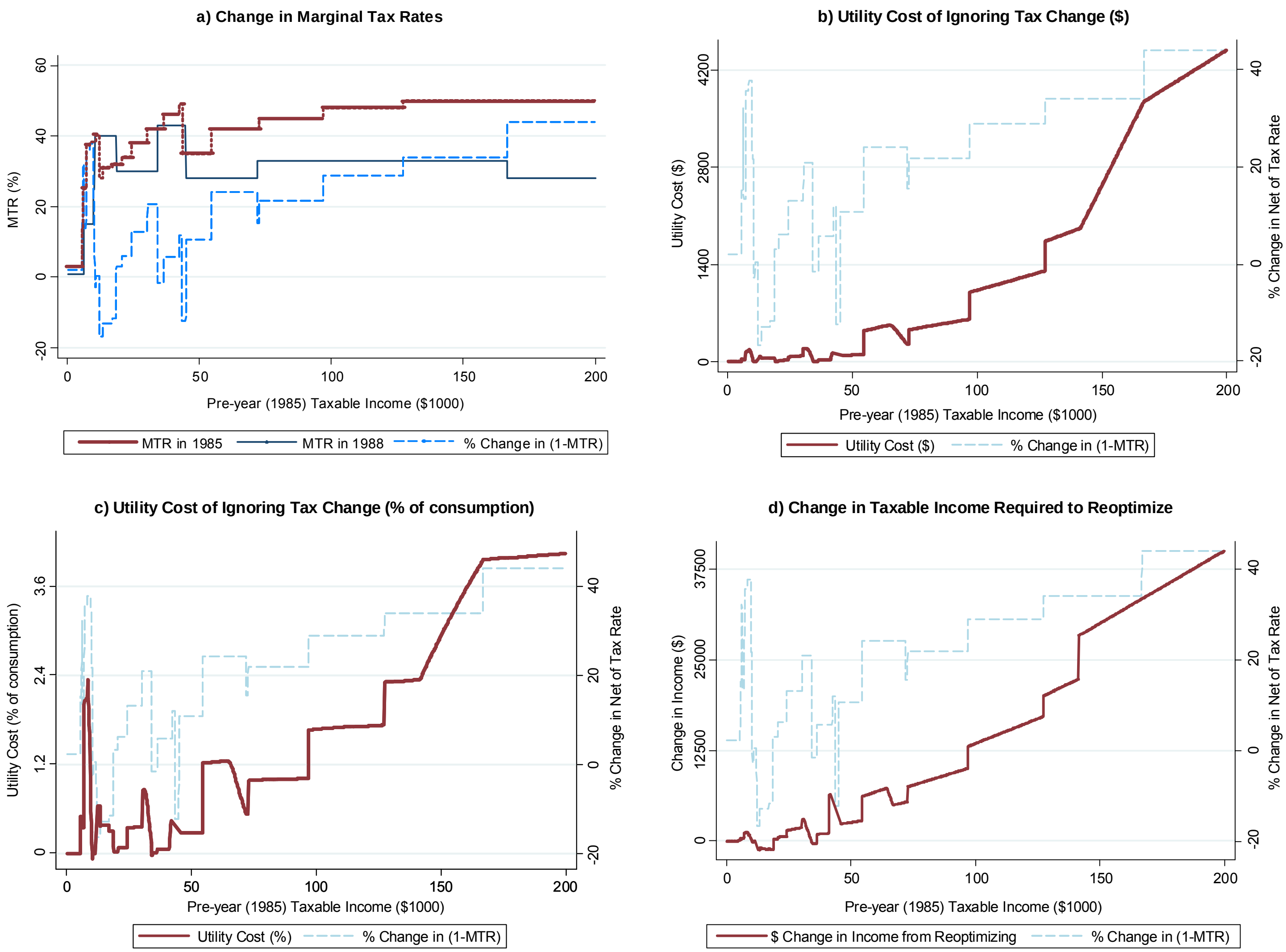
NOTE–This figure plots the bounds $(\varepsilon_L, \varepsilon_U)$ vs. $\hat{\varepsilon}$ for four combinations of δ and $\Delta \log p$, computed using the formula in Proposition 1. In the top two panels, the degree of optimization frictions is $\delta = 1\%$. The lower two panels consider $\delta = 0.5\%$. The left panels have a price change of $\Delta \log p = 20\%$, while the right panels have $\Delta \log p = 40\%$.

FIGURE 5
Upper Bound on Structural Elasticity with Zero Observed Response



NOTE–This figure illustrates the result in Corollary 2. Panel A shows the largest structural elasticity ($\varepsilon_U = 0.5$, dashed blue line) consistent with zero observed response ($\widehat{\varepsilon} = 0$, solid black line) to a price increase from p_A to p_B when $\delta = 1\%$. Panel A shows that under a quadratic approximation to $u(x)$, $x^*(p_B) - x^*(p_A) = 2d$, where d is the difference between the optimal demand and the lowest demand in the choice set. Panel B replicates Figure 1 and shows that the utility cost of being $2d$ units away from the optimum equals 4δ.

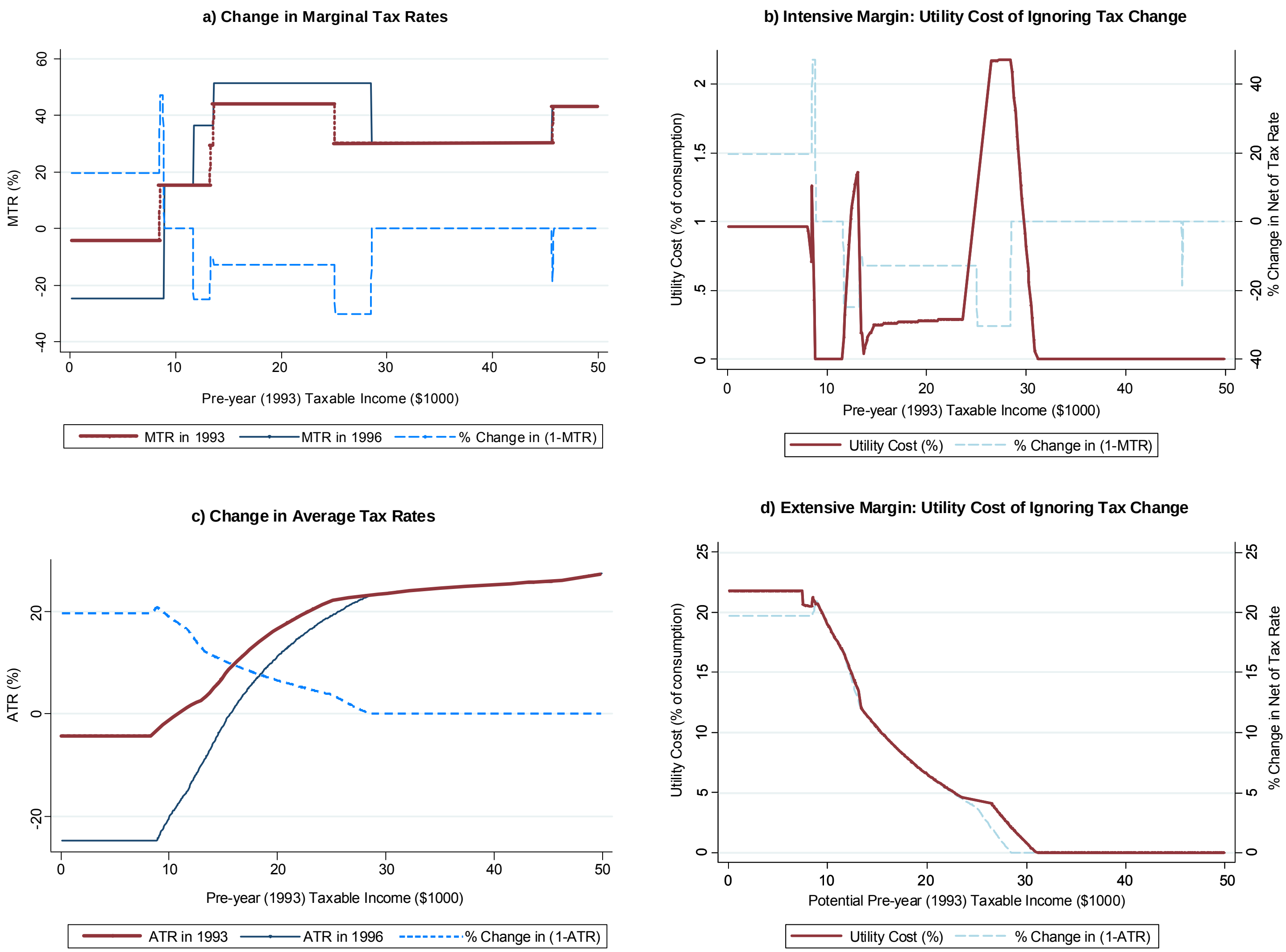
FIGURE 6
Tax Reform Act of 1986



NOTE–These figures are based on the Tax Reform Act of 1986. Panel A shows how marginal tax rates changed between 1985 and 1988 for single filers with two children. Panel B plots the utility cost Δu_i, measured in dollars of consumption, from failing to reoptimize labor supply on the intensive margin in response to the tax change with $\varepsilon = 0.5$. Panel C plots the same utility cost as a percentage of pre-reform consumption ($\Delta u_{i,\%}$), defined as the dollar cost in Panel B divided by the agent's consumption in 1985. Panel D shows the change in earnings ($wl^*_{i,1988} - wl^*_{i,1985}$) required to reoptimize relative to the tax change. In Panels B-D, the dashed blue line (right y axis) replicates the percentage change in the net-of-tax rate (1-MTR) shown in Panel A.

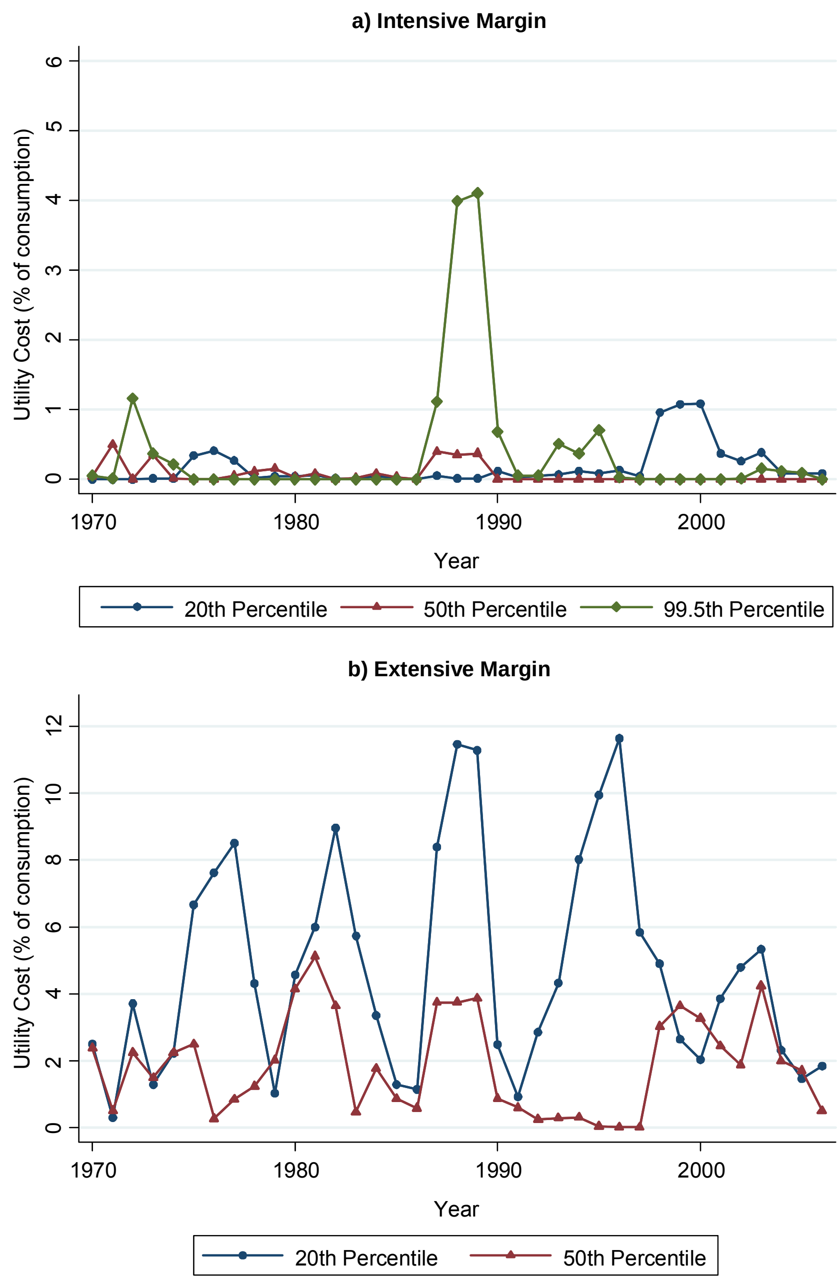
FIGURE 7

Clinton Earned Income Tax Credit Expansion



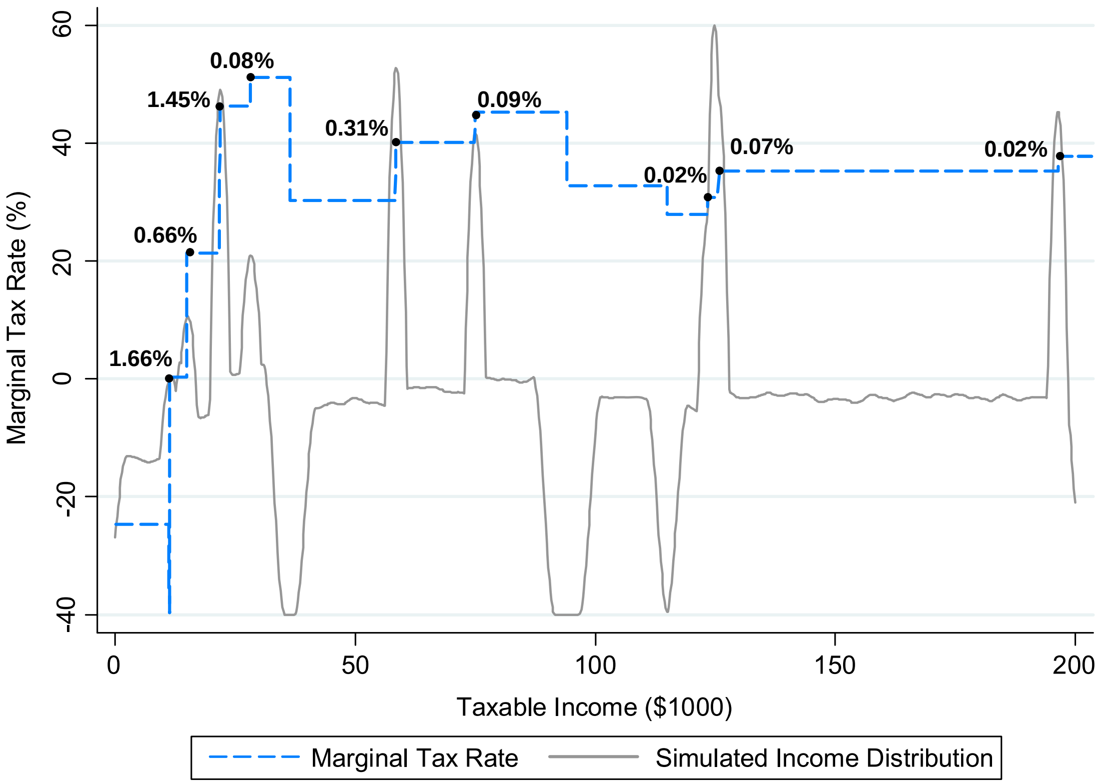
NOTE–These figures are based on the Clinton EITC Expansion enacted between 1993 and 1996. Panel A shows how marginal tax rates changed over this period for single filers with two children. Panel B plots the utility cost as a percentage of pre-reform consumption ($\Delta u_{i,\%}$) from failing to reoptimize labor supply in response to the tax change for agents who were already working in 1993 with $\varepsilon = 0.5$. Panels C and D are the extensive margin analogues to Panels A and B. Panel C shows changes in average tax rates by taxable income levels from 1993 to 1996. Panel D plots the utility cost ($\Delta u_{i,\text{ext},\%}$) from failing to reoptimize labor supply in response to these tax changes for agents at the margin of entering the labor force prior to the reform. This panel plots the utility gain from entering the labor force in 1996 for agents whose taste a_i and fixed cost k made them indifferent between working and not working in 1993 at the income levels on the x axis. This dollar utility gain is divided by optimal consumption when working in 1993 to obtain a percentage measure.

FIGURE 8
Utility Cost of Ignoring Tax Changes by Year



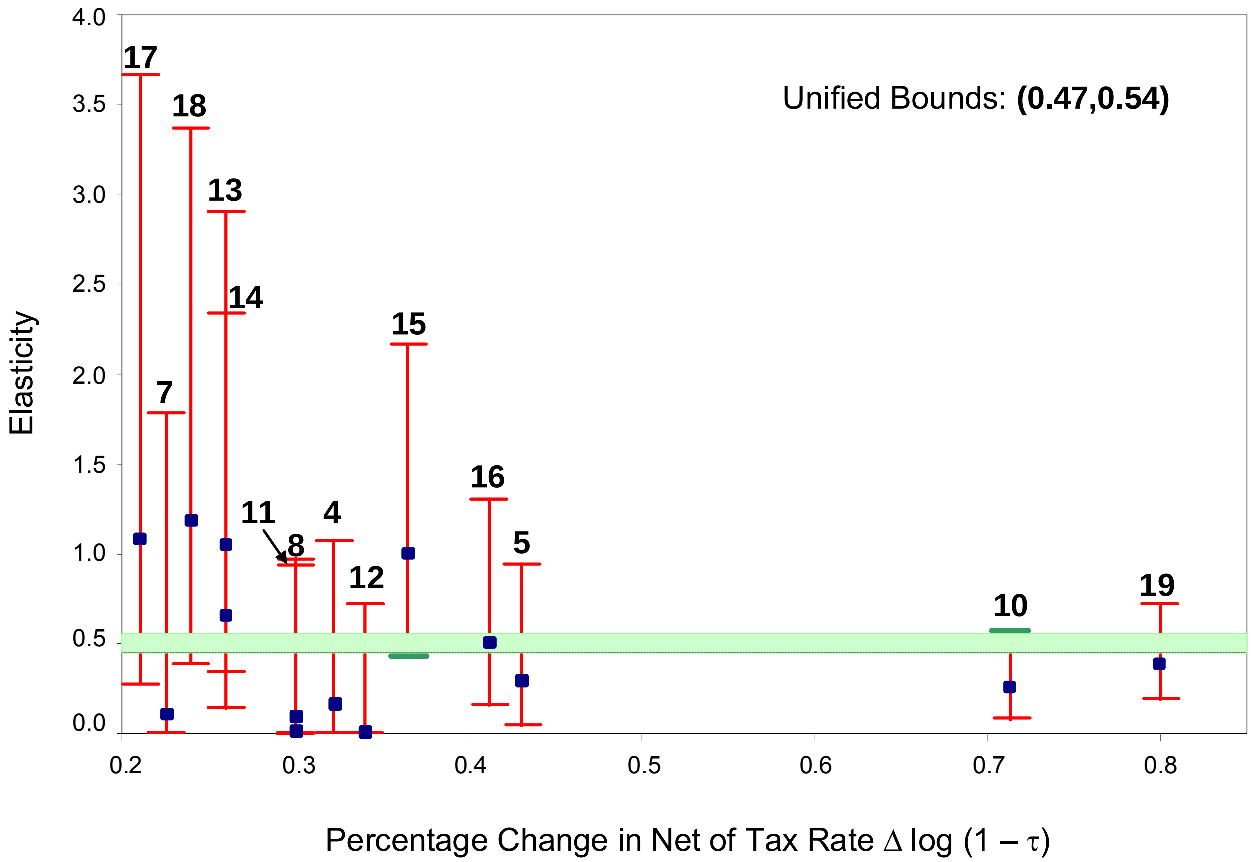
NOTE–These figures plot the utility cost of ignoring changes in taxes over three-year periods from the 1970 to 2006 for selected percentiles of the income distribution. In each year *y*, the point that is plotted shows the utility loss (as a percentage of consumption) from choosing labor supply optimally according to the tax system in year $y-3$ instead of year y. Panel A depicts the utility cost of failing to reoptimize labor supply on the intensive margin ($\Delta u_{i,\text{int},\%}$) with a structural intensive-margin elasticity of $\varepsilon = 0.5$, calculated as in Figure 6c. Panel B depicts the percentage utility cost of failing to reoptimize labor supply on the extensive margin ($\Delta u_{i,\text{ext},\%}$), calculated as in Figure 7d.

FIGURE 9
Gains from Bunching at Kinks in 2006 Tax Schedule



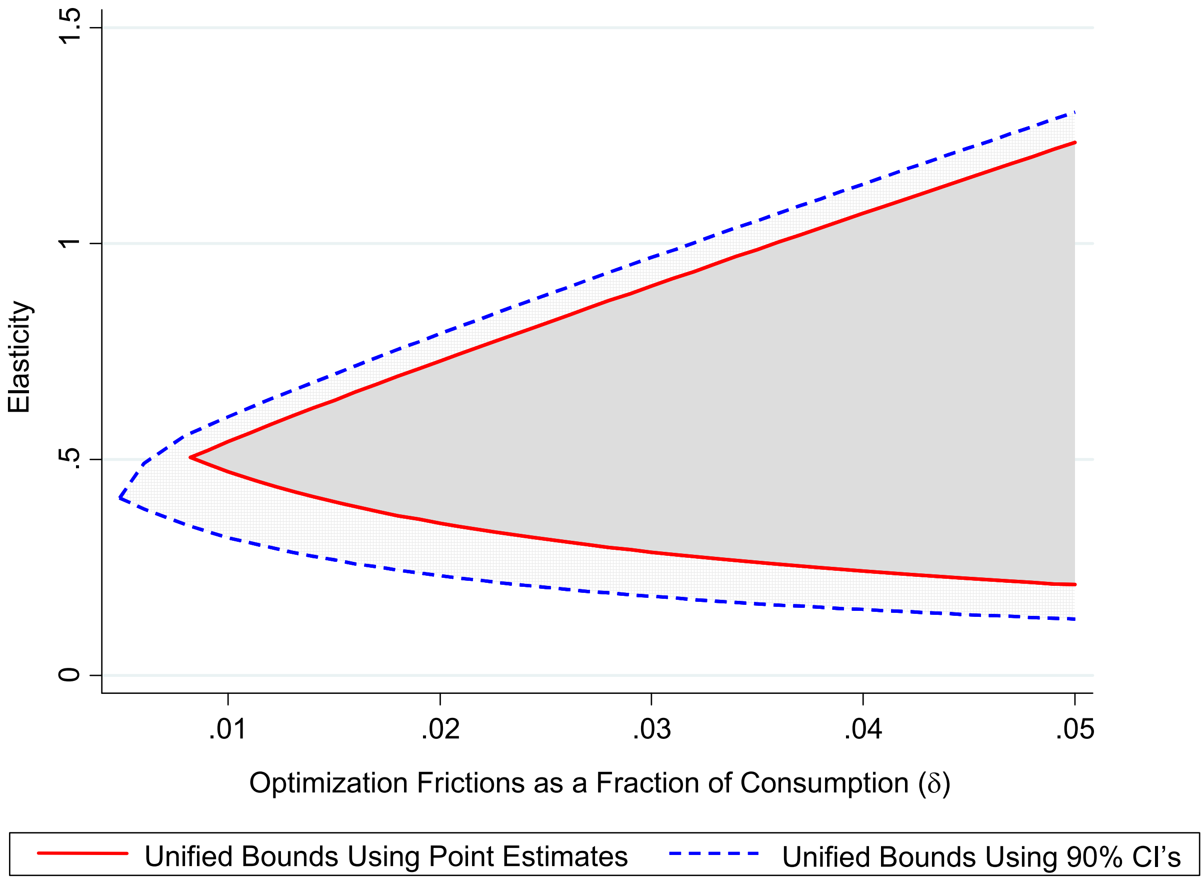
NOTE–The dashed blue curve shows the 2006 marginal tax rate schedule in the U.S. The solid grey curve shows the distribution of taxable income predicted by the frictionless labor supply model with $\varepsilon = 0.5$. This curve assumes a uniform distribution of a_i and plots an Epanechnikov kernel density of the simulated earnings distribution with a bandwidth of $1000. The numbers near each convex kink are the percentage utility gain ($\Delta u_{\%}$) from locating at that kink when $\varepsilon = 0.5$. To compute $\Delta u_{\%}$ at a given kink, I first define $\Delta u_{i,\%}$ as the utility gain for an individual with taste parameter a_i from locating at that kink relative to optimizing under the (incorrect) assumption that the tax rate in the previous bracket continues into the next bracket. I then define $\Delta u_{\%}$ as the unweighted mean of $\Delta u_{i,\%}$ over all individuals whose a_i would make it optimal for them to locate at that kink. The first two kinks (1.66% and 0.66%) correspond to the end of the phase-in and start of the phase-out regions of the EITC.

FIGURE 10
Bounds on Intensive-Margin Labor Supply Elasticities with $\delta = 1\%$



NOTE–The red intervals show the bounds on the structural intensive-margin elasticity ε implied by each of the studies with corresponding numbers listed in Table 1. The blue squares show the point estimate of each study. The x axis is the log change in the net of tax rate ($\Delta\log(1-\tau)$) used for identification in each study. Papers with $\Delta\log(1-\tau) < 20\%$ are excluded from this figure for scaling purposes. The shaded region shows the range of structural elasticities consistent with all the observed elasticities (the unified bounds).

FIGURE 11
Unified Bounds on Labor Supply Elasticity vs. Degree of Frictions



NOTE–This figure shows how the unified bounds on the structural intensive-margin elasticity ε vary with the level of frictions δ. The solid red lines plot the unified bounds implied by point estimates of the twenty observed elasticities $\hat{\varepsilon}$ in Table 1. These unified bounds are defined only for $\delta > \delta_{\min} = 0.8\%$ because δ's below this threshold cannot reconcile the observed elasticities. The dotted blue lines plot the unified bounds constructed from the upper and lower endpoints of each study's 90% confidence interval.

TABLE 1

Bounds on Intensive-Margin Hicksian Labor Supply Elasticities with δ = 1% Frictions

Study	Identification	$\hat{\varepsilon}$	s.e.($\hat{\varepsilon}$)	Δlog(1-τ)	ε_L	ε_U	95% CI ε_L	95% CI ε_U
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A. Hours Elasticities								
1. MaCurdy (1981)	Lifecycle wage variation, 1967-1976	0.15	0.15	0.12	0.00	5.63	0.00	6.11
2. Eissa and Hoynes (1998)	US EITC Expansions, 1984-1996	0.14	0.05	0.11	0.00	6.56	0.00	6.72
3. Blundell, Duncan, and Meghir (1998)	UK Tax Reforms, 1978-1992	0.14	0.09	0.19	0.01	2.54	0.00	2.81
4. Ziliak and Kniesner (1999)	Lifecycle wage, tax variation 1978-1997	0.15	0.07	0.32	0.02	1.05	0.00	1.24
5. Bianchi, Gudmundson, and Zoega (2001)	Iceland 1987 Zero Tax Year	0.29	0.03	0.43	0.09	0.91	0.07	0.98
	Mean observed elasticity	0.17						
B. Taxable Income Elasticities								
6. Gruber and Saez (2002)	US Tax Reforms 1979-1991	0.14	0.14	0.13	0.00	5.02	0.00	5.44
7. Saez (2004)	US Tax Reforms 1960-2000	0.09	0.04	0.23	0.00	1.75	0.00	1.88
8. Chetty et al. (2009)	Denmark, Top Kinks, 1994-2001	0.03	0.01	0.30	0.00	0.95	0.00	0.97
9. Chetty et al. (2009)	Denmark, Middle Kinks, 1994-2001	0.00	0.01	0.10	0.00	8.00	0.00	8.02
10. Gelber (2009)	Sweden, 1991 Tax Reform	0.25	0.03	0.71	0.12	**0.54**	0.09	0.60
11. Kleven and Schultz (2009)	Denmark, 1987 Tax Reform	0.01	0.01	0.30	0.00	0.91	0.00	0.94
12. Saez (2009)	US, 1st EITC Kink, 1995-2004	0.00	0.02	0.34	0.00	0.70	0.00	0.77
	Mean observed elasticity	0.08						
C. Top Income Elasticities								
13. Feldstein (1995)	US Tax Reform Act of 1986	1.04	0.26	0.26	0.37	2.89	0.17	3.50
14. Auten and Carroll (1999)	US Tax Reform Act of 1986	0.66	0.16	0.26	0.19	2.32	0.09	2.70
15. Goolsbee (1999)	US Tax Reform Act of 1986	1.00	0.15	0.37	**0.47**	2.14	0.32	2.47
16. Saez (2004)	US Tax Reforms 1960-2000	0.50	0.18	0.41	0.20	1.28	0.05	1.69
17. Kopczuk (2009)	Poland, 2002 Tax Reform	1.07	0.28	0.21	0.31	3.64	0.13	4.32
	Mean observed elasticity	0.85						
D. Macroeconomic/Trend-Based Elasticities								
18. Prescott (2004)	Cross-country Tax Changes, 1970-1996	1.18		0.24	0.42	3.34		
19. Davis and Henrekson (2005)	Cross-country Tax Variation, 1995	0.39	0.04	0.80	0.23	0.69	0.17	0.77
20. Blau and Kahn (2007)	US cohort wage trends, 1980-2000	0.56	0.14	0.16	0.07	4.33	0.03	4.74
	Mean observed elasticity	0.71						
				Unified Bounds:	**0.47**	**0.54**	0.32	0.60

Note: Values in columns 6-9 are calculated using formula in Proposition 1. See Appendix for sources and details underlying calculations in columns 3-5.

TABLE 2

Frisch Elasticities Implied by Hicksian Elasticity of ε_H = 0.5

		Income Effect: $-d[wl^*]/dY$							
		0.00	**0.05**	**0.10**	**0.15**	**0.20**	**0.25**	**0.30**	**0.35**
	0.00	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
	0.20	0.50	0.50	0.50	0.51	0.52	0.53	0.54	0.55
	0.40	0.50	0.50	0.51	0.52	0.53	0.55	0.57	0.60
EIS	**0.60**	0.50	0.50	0.51	0.53	0.55	0.58	0.61	0.65
(ρ)	**0.80**	0.50	0.50	0.52	0.54	0.56	0.60	0.64	0.70
	1.00	0.50	0.51	0.52	0.55	0.58	0.63	0.68	0.75
	1.20	0.50	0.51	0.52	0.55	0.60	0.65	0.72	0.79
	1.40	0.50	0.51	0.53	0.56	0.61	0.68	0.75	0.84

Note: This table shows the Frisch elastictity implied by various combinations of the EIS and income effect. The calculations assume that the ratio of unearned to earned income is $Y/wl^* = 1/2$ and the Hicksian (compensated) elasticity is $\varepsilon_H = 1/2$. The values are computed using the equation $\varepsilon_F = \varepsilon_H + \rho(d[wl^*]/dY)^2(Y/wl^*)$.