PRELIMINARY AND INCOMPLETE

Salience and Taxation

Raj Chetty, UC-Berkeley

Adam Looney, Federal Reserve Board

Kory Kroft, UC-Berkeley

March 2007*

Abstract

A central assumption in public finance is that individuals optimize fully with respect to the incentives created by tax policies. In this paper, we test this assumption using two empirical strategies. First, we conducted an experiment at a large grocery store where we posted tax-inclusive prices for certain products (showing both the pre-tax and post-tax price) over a three week period. Using scanner data, we find that posting the tax-inclusive prices reduced demand by 6-8% among the treated products relative to control products and nearby control stores. Second, using data on aggregate alcohol consumption by state from 1970-2003, we find that state-level increases in excise taxes (which are included in posted prices) reduce alcohol consumption significantly more than increases in sales taxes (which are added at the register and hence less salient). These results indicate that tax salience affects behavioral responses, contrary to canonical models. We propose a simple alternative model of boundedly rational agents to explain why salience matters. In the model, small (second-order) costs of cognition can lead agents to ignore fairly large taxes, even though these taxes may have large (first-order) effects on tax revenue and social welfare. We show that small costs of cognition may explain several stylized facts, such as greater awareness of average tax rates instead of marginal rates and larger responses on the extensive margin. Using this framework, we derive Harberger-type formulas for the efficiency cost and incidence of taxes when salience matters.

*E-mails: chetty@econ.berkeley.edu, adam.looney@frb.gov, kroft@econ.berkeley.edu.

We are very grateful to Sofia Berto Villas-Boas and Reed Johnson for help in implementing the experiment. Thanks to George Akerlof, David Ahn, Alan Auerbach, Stefano Della Vigna, Amy Finkelstein, Michael Greenstone, Jon Gruber, Shachar Kariv, Botond Koszegi, James Poterba, Matthew Rabin, Emmanuel Saez, Jesse Shapiro, Uri Simonsohn, and numerous seminar participants for helpful comments and discussions. Matt Grandy, Matt Levy, Ankur Patel, and James Sly provided outstanding research assistance. Funding was provided by NSF grant SES 0452605. The analysis and conclusions set forth are those of the authors and do not indicate concurrence by other members of the research staff or the Board of Governors of the Federal Reserve.

1 Introduction

A central assumption in public finance is that agents optimize fully with respect to the incentives created by tax schedules. For example, Ramsey's (1927) seminal analysis of optimal commodity taxation assumes that agents respond to tax changes in the same way that they to price changes. More recent analyses of taxation assume that agents optimize fully against complex, non-linear tax schedules. Examples include Mirrlees' (1971) classic treatment of income taxation, more recent work on optimal taxation in dynamic models (e.g., Judd 1985, Golosov et. al. 2003), or analyses of social insurance (e.g. Diamond and Sheshinski 1995, Chetty 2006, Shimer and Werning 2006).

The assumption that individuals optimize fully against the tax code also plays an important role in positive analyses of tax policy and empirical public finance. Tax schedules are quite complex in practice and are often not transparent. Income tax schedules are typically highly non-linear; benefit-tax linkages for social insurance programs are opaque (e.g. social security taxes and benefits); and taxes on commodities vary significantly and are often not displayed as part of posted prices (hotel city taxes, vehicle excise fees, sales tax rules). The efficiency and welfare consequences of such tax policies depend on how individuals optimize against these schedules.

Despite the importance of the full-optimization assumption in the analysis of taxation, there is relatively little evidence on whether individuals actually respond to taxes in this manner. In this paper, we investigate this issue by analyzing the effect of "salience" on behavioral responses to taxation. We define the "salience" of a tax as the cost of computing the tax-inclusive price of a good. For example, in the context of commodities, a tax that is included in the posted price is more salient than a tax that is only added at the time of payment (such as a sales tax), since computing the price including the latter tax requires additional computation. We test whether salience commodity tax has different effects on demand depending on whether it is included in or excluded from the posted price. Our empirical analysis can be motivated by the following simple specification for demand: $D = f(p + \theta t)$. The traditional assumption in the public finance literature is that $\theta = 1$ – taxes and prices affect behavior in the same way. We test whether $\theta = 1$ in the context of commodity taxes using two complementary methods: (1) a field experiment in a grocery store and (2) "natural experiments" in observational data on alcohol consumption.

The field experiment was implemented in collaboration with a major chain at a large grocery store over a three-week period in early 2006. In this store (as in most other retail outlets in the United States), prices posted on the shelf exclude sales tax; if the good is subject to tax, it is added to the bill only at the register. Our intervention was to post tags showing the tax-inclusive price below the original pre-tax price tag for all products in three groups (cosmetics, hair care accessories, and deodorants) . This intervention can be interpreted as a test of whether increasing the salience of the tax-inclusive price has an effect on demand behavior.

We analyze the effect of this intervention using a quasi-experimental differences-in-differences research design. Define the "treatment" group as the set of products we tagged (all products in the 3 categories – approximately 1,000 products total) in the treatment store during the three week treatment period. We compare demand for the products in the treatment group to demand in three "control groups": (1) a set of control products in nearby aisles that we did not tag (toothpaste, skin care products, shaving products); (2) a pair of control stores in nearby cities whose customers have similar demographic characteristics to the treatment store; and (3) control time periods in the months before the experiment took place.

Using scanner data, we find that quantity sold and total revenue in the group of products we treated fell by 6-8% during the intervention relative to the three control groups. This estimate is statistically significant with $p < 0.01$. This result is robust to controlling for price variation, varying the set of control groups. Estimation of "placebo effects" to implement non-parametric permutation tests also indicates that the showing tax-inclusive prices significantly reduced demand. To interpret the magnitude of the 6-8% decline in demand, we estimate product-level price elasticities for the products in the treatment group using the high-frequency price variation in the data in the pre-period. We find an (intertemporal) price elasticity of demand between 1 and 1.5, consistent with earlier estimates of category-level price elasticities in grocery stores (e.g. Hoch et. al. 1995). Given that the sales tax rate is 7.375%, we conclude that showing the tax-inclusive price reduced demand nearly as much as a price increase of an equivalent amount. This suggests that the vast majority of individuals ignore the sales tax on these products when making consumption decisions (i.e., $\theta \simeq 0$).

A concern with the experiment is that posting 1,000 new tags may have reduced demand simply because it temporarily violated familiar norms. This issue motivates our second empirical strategy, which compares the effect of price changes with tax changes using observational data. This test can be interpreted as a method of estimating the value of θ on average in the field over a longer horizon. We implement this test using annual data on aggregate alcohol consumption by state. We focus on alcohol because it is subject to two state-level taxes in the U.S.: an excise tax that is included in the posted price and a sales tax that is added at the register (and hence less salient). Exploiting state-level reforms in these two tax rates between 1970 and 2003, we find that increases in the excise tax reduced alcohol consumption approximately four times more than increases in the sales tax of similar magnitude.[1] This difference in elasticities persists over relatively long horizons (e.g. 2 or 3 years). Combining the two elasticity estimates suggests that $\theta < \frac{1}{\square}$ even in the longer run.

Both strands of evidence indicate that behavioral responses to taxation depend substantially on whether taxes are included in poste prices. There are two potential explanations for this finding. One is that customers are uninformed about the sales tax rate or which goods are subject to sales tax since it is not typically shown in prices. An alternative hypothesis is that salience matters: the customers know what is taxed, but choose to focus on the posted price when deciding what to buy because computing the tax-inclusive price for each good involves a cognitive or time cost. To distinguish between these competing hypotheses, we surveyed customers entering the grocery store about their knowledge of sales taxes. The median individual correctly reported the tax status of 7 out of the 8 products on the survey, and reported the average sales tax rate within 0.5 percentage points of the true rate. Since most individuals are in fact well informed about taxes when their attention is drawn to the subject, we conclude that they must choose not to compute tax-inclusive prices when shopping. As a result, the salience of tax-inclusive prices affects consumption

[1] An obvious concern with this comparison is that the sales tax applies to a broader range of goods than the alcohol excise tax. However, since food is generally exempt from the sales tax, increases in the sales tax change the relative price of alcohol relative to food and non-alcoholic drinks, which are presumably the most important substitutes. See section 5 for additional details..

decisions.

The empirical analysis indicates that the benchmark "unbounded rationality" model where individuals always take all taxes into account does not fit the data. What alternative model explains why salience matters? The second part of this paper focuses on developing a model that can match the empirical results while providing a tractable framework for public finance. We propose a simple bounded rationality model where agents face a cognitive cost of computing tax-inclusive prices (as in Simon 1955, Akerlof and Yellen 1985). An attractive feature of this model is that it has a well defined notion of welfare – utility minus cognitive costs. As a result, revealed preference can be used to evaluate the welfare consequences of tax policies in this model, as in the Ramsey-Harberger tradition.

We use the model to establish a series of results. First, we show that second-order (small) cognitive costs can lead agents to ignore details of a broad range of tax policies. Intuitively, when agents are close to an interior optimum to begin with, the marginal welfare gain from reoptimizing relative to the true tax rate is small – an application of the envelope theorem, as in Akerlof and Yellen (1985). For example, a simple calibration using quasilinear utility shows that the cost of miscalculating the tax rate by 10 percentage points on an item on which the agent spends $10,000 is only $50. As a result, small cognitive costs (limited attention) can lead individuals to rationally ignore taxes on goods such as cosmetics or alcohol unless they are already included in posted prices.

Second, we show that even though agents may incur second-order utility losses from ignoring details of tax policies, these policies can still have first-order (large) effects on social welfare and revenue. For example, a 10% tax increase can raise a significant amount of revenue for the government regardless of whether the agent reoptimizes his behavior. If the agent does reoptimize, the tax increase could create substantial deadweight burden because of the fiscal externality that the agent imposes on the government by changing his behavior.[2] Hence, bounded-rationality and salience can be important in the analysis of many large-scale tax policies from a social perspective, even though they may be unimportant from an

[2] This does not necessarily imply that 'hidden' tax changes are less distortionary or that cognitive frictions improve the efficiency of taxation (in contrast with the results of Liebman and Zeckhauser 2004). This is because perceptions of tax rates presumably adapt to actual rates, and income effects can make non-salient tax changes costly by distorting consumption allocations.

individual's perspective.

Third, we provide a (preliminary) characterization of some of the simplifying heuristics a boundedly rational agent uses when faced with a complex tax schedule. Some aspects of these heuristics match stylized facts in the literature. For instance, the utility gain from knowing the average income tax rate is likely to exceed that of knowing the marginal income tax rate under plausible conditions. This may explain why people appear more cognizant of average tax rates than marginal rates, as documented by Liebman (1998). The model also predicts that agents should respond more on the extensive margin than on the intensive margin to tax changes, and is consistent with limited "bunching" at kink points (Saez 2002). The model also predicts that behavioral responses to taxation should be larger among the rich, consistent with the evidence in Goolsbee (2000) and Saez (2004). More generally, the model suggests that actual marginal rates – which are the focus of much of the existing theoretical and empirical literature – may be less important in determining behavior than broad, long run tax perceptions. This reasoning could potentially explain why the estimated elasticity of labor supply with respect to the tax-rate is larger in studies that compare across countries (e.g. Prescott 2004, Davis and Henrekson 2006) than in studies that focus on changes in behavior in short windows around tax reforms (e.g. Gruber and Saez 2002, Saez 2003).

Finally, we apply the theoretical framework to conduct a positive analysis of taxation when agents are boundedly rational and salience matters. We derive Harberger-type formulas for the incidence and efficiency costs of taxation. The deadweight cost of taxation differs from the standard Harberger expression in three ways: (1) the perceived tax rate is what matters; (2) there is an additional term related to the difference between the perceived and actual tax rates that reflects a distortionary income effect that arises from budgeting errors when taxes are not salient; and (3) the relevant elasticity is not the net-of-tax elasticity generally estimated in the empirical literature, but the "fundamental" price elasticity. In the analysis of incidence, we show that the classic tax-neutrality result breaks down: when individuals are boundedly rational, incidence will in general depend on whether the tax is levied on consumers or firms. The effect of elasticities on incidence is amplified when taxes are not salient: for more inelastic goods, consumers are less likely to pay attention to taxes, and are likely to bear even more of the incidence. These findings also have new implications for the distributional incidence of commodity and income taxation across income groups.

The remainder of the paper is organized as follows. Section 2 discusses related literature. Section 3 presents a simple two-type model as an organizing framework for our empirical analysis. Section 4 discusses the field experiment, section 5 presents the evidence on alcohol sales, and section 6 presents the survey evidence. In section 7, we develop the model of boundedly-rational agents and show how it can explain our empirical findings as well as other stylized facts. Section 8 analyzes the efficiency consequences and incidence of taxation in this framework. Section 9 concludes.

2 Related Literature

Our work builds on and relates to several strands of the literature in behavioral economics, macroeconomics, and public finance. First, empirical studies have documented the importance of salience and limited attention in a variety of economic contexts: up-front appliance costs vs. subsequent electricity costs (Hausman and Joskow 1982); non-linear pricing (Shin 1985); internet price search engines (Ellison and Ellison 2004); prices vs. shipping fees (Morgan and Hossain 2005); financial markets (Barber, Odean and Zheng 2005; DellaVigna and Pollet 2006); the pass-through of manufacturer rebates for car purchases (Busse, Silva-Risso, and Zettlemeyer 2006); and rankings of colleges and hospitals (Pope 2006). Similarly, studies in marketing have shown that the partitioning of prices into "base prices" and additional fees or into monthly payments vs. total payments has real effects on demand (e.g. Gourville 1998, Morwitz et.al. 1998).

Salience has received less attention in the public finance literature. A small body of studies has demonstrated that individuals often misunderstand the difference between marginal vs. average tax rates in the income tax schedule. Brown (1968) and Fujii and Hawley (1988) find that individual's self-reported marginal income tax rate often differs from the marginal tax rate implied by their demographic and income characteristics. de Bartolome (1995) shows using a lab experiment that many MBA students confuse the average rate with the marginal rate when making $1 "investments" in a taxable or non-taxable project. More recently, Liebman and Zeckhauser (2004) and Katuscak and Feldman (2006) present suggestive evidence that individuals' labor supply responds to average income tax rates rather than marginal tax rates using variation in the child tax credit. In a separate line of research, McCaffery and Baron (2003) document that the framing and presentation of alternative tax policy choices has significant effects on individuals' rankings of hypothetical policies when surveyed. Our analysis contributes to this literature by directly testing in the field whether the simplicity of computing tax-inclusive prices affects behavioral responses to commodity taxation. An advantage of focusing specifically on commodity taxes is that they are simply proportional to prices, and hence any evidence of imperfect optimization with respect to these taxes would suggest imperfect optimization with respect to a broad class of policies.

To analyze the implications of our empirical results for tax policy, we construct a model of taxation with inattentive agents that builds on the bounded rationality literature pioneered by Simon (1955). The concept underlying models of bounded rationality is that agents face a cost of processing information – a "deliberation cost" – and therefore rationally use simplifying heuristics to solve complex problems (see e.g., Conlisk (1988), Conlisk (1996), Gabaix et. al. (2006)). This logic has been applied most widely in the macroeconomics literature. Akerlof and Yellen (1985) and Mankiw (1985) show that failing to re-optimize in response to shocks generates second-order losses to agents, but has first-order effects on the macroeconomy. More recently, Sims (2003), Reis (2006), and Mackowiac and Weiderholt (2006) develop models of boundedly rational and inattentive consumers, and show that they can explain puzzles in aggregate consumption and pricing dynamics. In related work, Mullainathan (2002) and Wilson (2003) develop bounded memory and recall models, and show that they can explain puzzles for standard economic models that assume full-optimization. Ellison and Ellison (2004) and Gabaix and Laibson (2006) study equilibrium in models where individuals face cognitive constraints and firms have technologies to obfuscate or shroud attributes to raise profits. A key result of these models is that individuals may remain uninformed about shrouded (hidden) attributes in equilibrium because no market for debiasing will emerge. Our theoretical contribution is to introduce bounded rationality and limited attention into public finance by studying their implications for the positive analysis of taxation.

In this sense, our study contributes to an emerging literature on "behavioral public finance." One strand of this literature has adopted a paternalistic approach, assuming that agents maximize a utility function that systematically differs from the planner's objective function. An early example of this approach is Feldstein's (1985) classic analysis of optimal social security with myopic agents, where the social planner has a lower discount rate than individuals. More recent examples include the analysis of cigarette consumption and addiction when preferences are time-inconsistent (Gruber and Koszegi 2001); optimal taxes on sin goods (O'Donoghue and Rabin 2006); and optimal retirement savings policies for hyperbolic agents (Amador et. al. 2006). An alternative approach – the one we adopt here – is to assume instead that the individual and social planner agree on the objective function to be optimized, but that the individual faces certain cognitive constraints in achieving his true optimum when faced with a complex tax system. This approach is less developed in the existing literature. Sheshinski (2002) provides a parsimonious model of bounded rationality and shows that even small departures from full rationality may make it desirable for a benevolent social planner to restrict choices. Bernheim and Rangel (2007) take a more agnostic approach, and propose a method for constructing bounds on welfare gains based purely on observed choices even when there is no underlying utility representation available for those choices. Our theoretical analysis can be viewed as a special case of Bernheim and Rangel's approach, where we assume that the choices in the situation where tax-inclusive prices are salient are relevant for welfare analysis.

Finally, the idea that individuals focus on salient features of tax systems also has political economy implications for how governments set taxes. For example, a politician who wants to maximize his chance of re-election may try to create a wedge between the burden perceived by taxpayers and the actual burden (Krishna and Slemrod 2003). The empirical relevance of this idea is explored by Finkelstein (2006), who finds that state toll authorities raise tolls more frequently after introducing electronic toll collection systems, which make tolls less salient to drivers.

3 Empirical Framework

We begin by presenting an organizing framework for our empirical analysis using a simple model of consumption behavior in which some agents are inattentive to tax-inclusive prices.[3] Consider a static model where an agent with wealth Z has an additively separable quasilinear utility function over two goods, x and y, of the following form:

$$U(x,y) = a\frac{x^{1-b}}{1-b} + y$$

where $b > 0$ determines the price elasticity of good x. Normalize the price of y to 1, and let p denote the price of x. Assume that y is untaxed and x is subject to an ad valorem sales tax t_s. Hence, the total price of x is given by $p_t = p(1+t_s)$. The tax t_s is not included in the posted price that consumers see when deciding how much of x to purchase. Since consumers must compute the tax-inclusive price p_t but can observe the pre-tax price p without any computation, we will say that the tax t_s is less "salient" than the pre-tax price p.

We model the relationship between salience and behavioral responses by considering a setting with two types of agents who differ in their attention to tax-inclusive prices. The first type is a fully-optimizing consumer who uses the full tax-inclusive price when making his consumption decision, as in the neoclassical model. This type's choice of x is given by

$$x^*(p,t) = (\frac{p(1+t)}{a})^{-1/b}$$

The second type is a consumer who is inattentive, and focuses solely on the pre-tax price p when making his decision. He sets consumption of x as

$$\widehat{x}(p,t) = (\frac{p}{a})^{-1/b}$$

Let θ denote the fraction of agents who optimize relative to the true tax-inclusive price.

[3] In this section, we simply assume that some agents are inattentive, without modelling the source of this inattention. In section 7, we show that the inattentiveness assumed here can be derived as a rational consequence of cognitive constraints.

Then aggregate demand for x in an economy with a unit mass of agents is given by

$$\begin{aligned} x(p,t_s,\theta) &= \theta x^* + (1-\theta)\widehat{x} = (1-\theta)(\frac{p}{a})^{-1/b} + \theta(\frac{p(1+t_s)}{a})^{-1/b} \\ &= (\frac{p}{a})^{-1/b}[1-\theta+\theta(1+t_s)^{-1/b}] \end{aligned}$$

Recognizing that t_s is small, we simplify this expression using the first-order Taylor approximation $z^\theta \approx 1-\theta+\theta z$ for z around 1 to obtain

$$x(p,t,\theta) = (\frac{p}{a})^{-1/b}(1+t)^{-\theta/b}.$$

Taking logs of this expression yields the demand specification that underlies our empirical analysis:

$$\log x(p,t,\theta) = \alpha + \beta \log p + \theta\beta \log(1+t) \tag{1}$$

where $\alpha = \frac{1}{b}\log a$ and $\beta = 1\frac{1}{b}$. The parameter of interest is θ – the fraction of individuals in the population who take the sales tax into account when making consumption decisions. The null hypothesis in canonical models of taxation is that $\theta = 1$: all agents optimize relative to tax-inclusive prices. The primary objective of our empirical analysis is to test this hypothesis, and to provide an estimate of the value of θ associated with the sales tax levied on a particular subset of products. We use two independent empirical strategies to achieve this objective.

Strategy 1: Manipulate Tax Salience. Our first approach to estimating θ is to make the sales tax as salient as the pre-tax price by posting the tax-inclusive price on the shelf along. When tax-inclusive prices are posted, all individuals presumably optimize relative to the tax-inclusive price (i.e., $\theta = 1$). Hence, the effect of posting tax-inclusive prices on demand is given by

$$\log x(p,t,1) - \log x(p,t,\theta) = (1-\theta)\beta \log(1+t)$$

Since the price elasticity of demand is

$$\varepsilon_{D,p} = \frac{\partial \log x}{\partial \log p} = \beta$$

it follows that

$$(1 - \theta) = v/\varepsilon_{D,p} \tag{2}$$

where $v = \frac{\log x(p,t,1) - \log x(p,t,\theta)}{\log(1+t)}$ denotes the normalized "tax visibility" effect. The v parameter can be interpreted as the change in demand caused by making a 1% sales tax as salient as the price. The intuition underlying (2) is straightforward: the effect of posting tax-inclusive prices on demand relative to the effect of a price increase of corresponding size on demand identifies the fraction of individuals who ignore the sales tax. Hence, by combining estimates of the effect of posting tax-inclusive prices with estimates of the price elasticity of demand, one can infer θ. If all consumers normally optimize relative to the sales tax even when it is not as salient as the price ($\theta = 1$), posting the tax-inclusive price has no effect on demand ($v = 0$), since it is redundant information.

Strategy 2: Manipulate Tax Rate. An alternative approach to estimating θ is to exploit variation in t_s and compare the price elasticity of demand with the tax elasticity of demand. In particular,

$$\theta = \frac{\partial \log x}{\partial \log(1+t)}/\beta = \frac{\partial \log x}{\partial \log(1+t)}/\frac{\partial \log x}{\partial \log p}$$

Under the null hypothesis of full optimization, prices and taxes – which differ in their salience – should affect demand in the same way.

In the next section, we implement strategy 1 using a field experiment at a grocery store. In section 4, we implement strategy 2 using observational data on alcohol consumption.

4 Evidence from an Experiment at a Grocery Store

4.1 Research Design

We conducted an experiment showing tax-inclusive prices at a large grocery store in a suburb in Northern California. The store belongs to a grocery chain which has nearly 2,000 stores in the U.S. Within the store, approximately 30% of the products on the shelves are subject to the local sales tax rate of 7.375%. When applicable, the sales tax (rounded to the nearest cent) is added at the register. Price tags on the shelf display only pre-tax prices, as in the upper half of the tag shown in Exhibit 1.

We estimate the effect of posting tax-inclusive prices on demand using a quasi-experimental differences-in-differences research design. We use this design because direct randomization of tax-inclusive prices was infeasible given limitations in the scope and duration of the experiment. In particular, the grocery chain's managers expected that showing tax-inclusive prices would reduce sales. In order to limit revenue losses, we were required to restrict the intervention to three categories that were not "sales leading" categories, and limit the duration of the intervention to three weeks.[4] The three product groups were chosen in collaboration with the managers based on this requirement and two additional criteria: (1) having relatively high prices, so that the dollar amount of the sales tax is non-trivial; and (2) belonging to what the store terms "impulse purchase categories" – goods that exhibit high price elasticities – so that the demand response to the intervention would be detectable. This led us to run the experiment on three product groups – cosmetics, hair care accessories, and deodorants – over a three week period.[5]

To estimate the effect of the intervention, we compare sales in the "treatment" group of products whose tags were modified with three control groups that serve as counterfactuals. We define the treatment group as products that belong to the cosmetics, hair care accessories, or deodorants product groups in the treatment store during the three week treatment period. The first control group is a set of control products in the same aisles as the treatment products, for which we did not change tags. These products include similar (taxable) toiletries such as toothpaste, skin care, and shaving products; see Appendix Table 1 for the full list. The second control group is a pair of control stores in nearby cities whose customers have similar demographic characteristics to the treatment store. These control stores were

[4] Our initial request was to show tax-inclusive prices for all taxable products in the store.

[5] In principle, the treatment of showing tax-inclusive price tags could have been randomized at the individual product level. However, the concern that such an intervention could be confusing and potentially deceptive (e.g. suggesting that one lipstick is taxed and another is not) dissuaded us from pursuing this strategy. We therefore tagged complete product groups, so that any plausible substitute for a treated product would also be treated.

chosen based on a minimum distance criterion using characteristics listed in Table 1, which include variables such as the size of the store and the mean income of the city where the store is located. The third control group consists of sales in the treatment store in the months prior to the experiment.

Using these control groups, we implement a standard difference-in-difference methodology, testing whether sales of the treated products fell during the intervention relative to control products and control stores. As in other difference-in-difference analyses, the identification assumption underlying our estimate is a "common trends" condition (Meyer 1995), which in this case requires that sales would have evolved identically in the treatment and control groups absent the intervention. We discuss and evaluate this assumption below in the context of our empirical estimates.

Experiment Implementation. We posted tax-inclusive prices for products in the treatment group over a three week period, beginning February 22, 2006 and ending on March 15, 2006. Exhibit 1 illustrates how price tags were altered. The original tags, which show pre-tax prices, were left untouched on the shelf. A tag showing the tax-inclusive price was attached directly below this tag for each product. The added tag stated "Total Price: $\$p$ + Sales Tax = $\$p_t$," where p denotes the pre-tax price (repeating the information in the original tag) and p_t denotes the tax-inclusive price. The original pre-tax price was repeated on the new tag to avoid the impression that the price of the product had been increased. For the same reason, the fonts used for p, p_t, and the words "Sales Tax" exactly matched the font for the original price on the shelf. The tags were printed using a template and card stock supplied by the store (often used for sales or other additional information on a product) in order to match the color scheme and layout familiar to customers.

The store changes product prices on Wednesday nights and leaves the prices fixed (with rare exceptions) for the following one week period, termed a "promotional week." To synchronize our intervention with this pricing cycle, a team of researchers and research assistants printed tags every Wednesday night and attached them to each of the 1000 products. Occasionally, some prices are changed within a promotional week. To ensure that such changes did not lead to inaccuracies in (or removal of) the tax-inclusive price tags, research assistants made continual visits to the store to update the labels and ensure that the labels remained in good condition throughout the experiment. The tags were changed between 11 pm and 2 am, which are low-traffic times at the store.

4.2 Data and Summary Statistics

We use scanner data from the treatment store and the two control stores provided by the grocery chain. The data spans week 1 of 2005 to week 15 of 2006. Data on individual products are observed by "promotional week" – weeks beginning and ending on Wednesdays, in correspondence with the pricing cycle. The dataset includes unique product identifiers (UPC and category codes), the regular product price, the sale price (if any), and the number of units sold.

Summary measures of store characteristics are displayed in Table 1. The top panel presents store characteristics. Column (1) presents the statistics for the treatment store, and columns (2) and (3) for each control store. The three stores are large (roughly 37,000 sq. feet) and have been open for about 15 years. Panel B presents characteristics for cities where each store is located using data from the 2000 Census. The cities in which these stores are located are higher income than the U.S. average: the median household income is around $55,000, compared to $42,000 for the nation as a whole.

Table 2 presents category level summary statistics, broken down by treatment and control product groups within each store. The treatment groups consists of all products in the 3 broad groups described above. Within these 3 groups, there are 15 product "categories" (e.g. lipsticks, eye cosmetics, roll-on deodorants, body spray deodorants). The treatment categories were in two adjacent aisles, and together take up space equivalent to roughly half an aisle in the store. Average revenue from the treatment products as a whole is approximately $1,300 per week. The 96 control categories consist of other products sold in the aisles where the experimental products are sold (e.g. toothpaste, skin care products), whose tags were unchanged during the intervention period.

The first row Table 2 reports average weekly revenue in the treatment group and control group categories for each of the three stores. Average revenue in the treatment group is lower partly because the average price of treatment products is about $4.50 while products in the control groups sell for around $6.25 on average. The average product price, revenue per category and per product, and the number of items purchased each week is similar across stores.

For most of our analysis, we analyze the data at the category-by-store level (so that there are 111 observations per store per week), summing quantity sold and revenue over the individual products within categories in each store. We use the category-level approach because a large number of products do not sell at all in a given week. Because scanner data includes only transactions, unsold products do not appear in the product-level dataset. Since we do not we do not have data on the set of products that are on the shelf in each week, we cannot products that are on the shelf but did not sell from products that simply are not on shelves that week. By analyzing the data at the category level, we largely circumvent this problem because there are relatively few category-weeks with missing data (6% of all observations). Since all the categories always existed in all stores throughout the sample period, we are fairly confident that these observations are true zeros. As a robustness check, we have replicated our analysis at the product level, interpolating between observations to fill in zeros as discussed in the Appendix. We find that this product-level analysis yields very similar results to those reported below.

4.3 Results

Comparison of Means. We begin our analysis with a simple comparison of means between the treatment group and the three dimensions of counterfactuals: time, stores, and categories. Table 3 shows a cross-tabulation of the mean quantity sold per week per category for various time periods and product groups. The upper panel of the table shows data for the treatment store. The data is divided into four cells by time (pre-experiment vs. the three-week intervention period) and by product group (treated categories vs. control categories in the same store). Each cell shows the mean quantity sold for the group labeled on the axes, along with the standard error and the number of observations. All standard errors reported in this and subsequent tables in this section are clustered by week to correct for serial correlation of errors across products.

The mean quantity sold in the treatment categories fell by an average of 1.12 units per week during the experimental period relative to the pre-period baseline. Meanwhile, quantity sold in the control categories within the treatment store went up by 0.84 units. Hence, sales fell in the treatment categories relative to the control categories by 1.96 units on average, with a standard error of 0.64. This change of $DD_{TS} = -1.96$ units is the "within treatment store" DD estimate of the impact of posting tax-inclusive prices. The identification assumption necessary for consistency of DD_{TS} as an estimate of the effect of showing tax-inclusive prices is that the time trend in sales of the treatment products and control products would have been similar absent the intervention.

One natural way of evaluating the validity of this identification assumption is to compare the change in sales of treatment and control products in the control stores, where no intervention took place. The lower panel of Table 3 presents such a comparison by showing mean sales for the same sets of products and time periods in the two control stores. In the control stores, sales of control products rose by 0.18 units relative to baseline during the period of the intervention, while sales of treatment products rose by 0.23 units. Hence, sales of treatment products increased by a (statistically insignificant) $DD_{CS} = 0.04$ units relative to sales of control products within the control stores. The fact that DD_{CS} is not significantly different from zero suggests that sales of the treatment and control products would in fact have evolved similarly in the treatment store had the intervention not taken place. This "placebo test" therefore supports the validity of the within treatment store DD_{TS} estimator.

Putting together the upper and lower panels of Table 3, one can construct a "triple difference" (DDD) estimate of the effect of the intervention, as in Gruber (1994). This estimate is $DDD = DD_{TS} - DD_{CS} = -2.01$. This estimate is statistically significant with $p < 0.01$, indicating that the null hypothesis of full-optimization ($\theta = 1$) is rejected. Note that both within-store and within-product time trends are differenced out in the DDD. The DDD estimate is therefore immune to *both* store-specific shocks – such as a transitory increase in customer traffic – and product-specific shocks – such as fluctuations in demand for certain goods. Hence, the identification assumption for consistency of the DDD estimate is relatively weak: it requires that there was no contemporaneous shock during our experimental intervention that differentially affected sales of the treatment products in the treatment store (relative to both other products in the same store and the same products in other stores). In

The DDD estimate is therefore immune to *both* store-specific shocks – such as a transitory increase in customer traffic – and product-specific shocks – such as fluctuations in demand for certain goods. Hence, the identification assumption for consistency of the DDD estimate is relatively weak: it requires that there was no contemporaneous shock during our experimental intervention that differentially affected sales of the treatment products in the treatment store (relative to both other products in the same store and the same products in other stores). In view of the planned, exogenous nature of the intervention, we believe that this condition is likely to be satisfied, and hence that the DDD provides a consistent estimate of the treatment effect.

To gauge the magnitude of the estimated effect, we use the framework developed in section 3. The mean quantity sold per category in the sample is 28.2 units. The estimate of -2.01 therefore implies that quantity sold fell by 7.2 percent. Given the sales tax rate of 7.375 percent, the normalized tax visibility effect is approximately $v = 1$. As we discuss below, the price elasticity of demand at the category level $\varepsilon_{d,p}$ is between 1 and 1.5. Since $\theta = \frac{v}{\varepsilon_{D,p}}$, we infer that $\theta < \frac{1}{3}$, i.e. fewer than 1/3 of individuals appear to take account of the sales tax when purchasing products in the treatment group. Note that we cannot reject the hypothesis that $\theta = 0$ given the standard error associated with our estimate of θ. Hence, the data is consistent with the case where *none* of the individuals base their purchasing decisions on the tax-inclusive price when purchasing small products such as cosmetics.

Regression Estimates. We evaluate the robustness of the DDD estimate by estimating a series of regression models with different covariate sets and sample specifications in Tables 4 and 5. Let the outcome of interest (e.g. quantity, log quantity, revenue) in store s in category c in week t be denoted by y_{sct}. Let the variables *treatstore*, *treatcat*, and *treattime* be indicators for whether where the observation is for the experimental store, categories and time, respectively. Let X denote a vector of additional covariates. To identify the effect of the experimental intervention on the outcome of interest, we estimate variants of the following linear model, which generalizes the strategy used in Table 3:

$$
\begin{aligned}
y_{sct} \;=\; & \alpha + \beta_1 treattime + \beta_2 treatstore + \beta_3 treatcat + \gamma_1 treattime \times treatcat \\
& + \gamma_2 treattime \times treatstore + \gamma_3 treatstore \times treatcat \\
& + \delta treattime \times treatcategory \times treatstore + \rho X + \varepsilon_{sct} \qquad (4)
\end{aligned}
$$

In this specification, the β_i coefficients capture changes in sales over time (β_1), time-invariant difference between the experimental store and control stores (β_2), and time-invariant differences between the treated categories and control categories (β_3). The second-level interactions control for changes in sales in the treatment categories over time (γ_1), changes in sales in the treatment store over time (γ_2), and time-invariant characteristics of the treatment category in the treatment store (γ_3). Finally, the third-level interaction (δ) captures the treatment effect of the experiment, and corresponds to exactly to the DDD estimate when no additional controls are included.

As a reference, specification 1 of Table 4 replicates the DDD estimate in Table 3 by estimating (4) for quantity sold without any additional controls.[7] Specification 2 replicates 1, controlling for the mean price of the products in each category using a quadratic specification. The estimate on the treatment coefficient is essentially unchanged with the price control, which is unsurprising given that there were no atypical price changes during our intervention period. We return to the interpretation of the estimated price effects below. In specification 3, we examine the effect of the intervention on weekly revenue (price×quantity) per category. Consistent with the evidence from the quantity analysis, we find that the experiment led to a significant reduction in revenue from the treatment products relative to the controls.

In specifications 4 and 5, we estimate analogous models in logs instead of levels. An advantage of the logs specification is that it is a better model for comparisons across categories with different baseline quantities, given that shifts are likely to be equi-proportional (i.e. elasticities of demand are roughly constant). A disadvantage of the logs specification is

[7]Including a full set of fixed effects for time, stores, and products (e.g. week dummies, store dummies, category dummies and their interactions) yields exactly the same estimate of δ. This is because there is no variation in the third-level interaction across products, stores, or time once we condition on the variables included in (4).

that it forces us to omit observations that have zero quantity sold (which account for roughly 6% of the sample at the category-week level). Consistent with the levels models, the logs specifications imply an estimated reduction in quantity sold of 8% and revenue of 11%.

Both the levels and logs specifications suggest that revenue per category fell more than quantity sold per category.[8] We explore this issue further in specification 6, by estimating the effect of the intervention on the average price of the purchased products within a category (i.e. revenue divided by quantity in each category). While imprecisely estimated, the coefficient estimate implies that the average price of items purchased fell by about $0.13 during the treatment period (roughly 3 percent), consistent with the gap between the revenue and quantity estimates in the earlier regressions. One interpretation of this result is that individuals who were considering buying a more expensive product were more likely to decide not to buy anything at all because the tax levied on more expensive products is larger in dollar terms. Another interpretation is that individuals substituted toward cheaper products within the treatment categories. Unfortunately, we cannot distinguish between these alternative hypotheses about the effect of posting tax-inclusive prices on demand substitution patterns within treated categories.[9]

Placebo Tests and Robustness Checks. As noted by Bertrand et. al. (2003), a serious concern in DD analysis is that serial correlation can induce trends that lead to overrejection of the null hypothesis of no effect. To address this concern, we first check for unusual patterns in demand in the weeks immediately before and after the experiment. We replicate specification 1 in Table 4, and include indicator variables for the three week period before the intervention began ($beforetreat$) and the three week period after the intervention ended ($aftertreat$). We also include second- and third-level interactions of $beforetreat$ and $aftertreat$ with the $treatcat$ and $treatstore$ variables, as for the $treattime$ variable in (4).

Column 1 of Table 5 reports estimates of the third-level interactions (e.g. $beforetreat \times treatstore \times treatcat$) for the periods before, during, and after the experiment. Consistent

[8] To see this for the levels specifications, note that the average price of the products in the dataset (weighted by quantity sold) is $5.49. If quantity sold of all products within each category fell equally, one would expect a revenue loss of only $11 per category based on the estimated quantity reduction of 2 units.

[9] Identifying the extent of within-category substitution would require an intervention that affects a subset of products within a category and examines the resulting shifts in demand.

with the results in Table 4, quantity sold in the treatment group is estimated to have fallen by approximately $\delta = 2$ units during the intervention. The corresponding "placebo" estimate for the period before the treatment is close to zero, suggesting that the fall in demand began during the actual intervention. The estimated effect for the period after the treatment is also close to zero. One explanation of this result is that individuals focus again on the pre-tax price once the tags are removed. Another explanation is that the set of individuals who shop for these durable goods varies substantially across weeks, so customers in the weeks after the experiment were effectively untreated. While we cannot distinguish between these alternative explanations, in either case the return of demand to pre-experiment levels supports the claim that the intervention (and not a contemporaneous trend) caused a significant reduction in demand.

Building on the logic of specification 1, we implement a non-parametric permutation test of the hypothesis that $\delta = 0$ that directly addresses concerns about serial correlation and the potential bias of t-tests. Let $t = 1, ..., T$ index the weeks for which sales data are observed. Consider the following estimating equation, in which the *treattime* variable is replaced with $time_t$, an indicator variable for an arbitrary three week interval $\{t, t+1, t+2\}$ during the sample frame:

$$\begin{aligned} y_{sct} &= \alpha + \beta_1^t time_t + \beta_2^t treatstore + \beta_3^t treatcat \\ &+ \gamma_1^t time_t \times treatcat + \gamma_2^t time_t \times treatstore + \gamma_3^t treatstore \times treatcat \\ &+ \delta_t time_t \times treatcategory \times treatstore + \varepsilon_{sct} \end{aligned} \quad (5)$$

We estimate this model for all t such that the $time_t$ variable does not overlap with the actual three-week treatment period (i.e. t such that $treattime \times time_t = 0$ for all observations). The estimated $\{\widehat{\delta_t}\}$ values yield an empirical distribution of "placebo effects" in the sample. Let G represent the cdf for this distribution. The probability of observing a DDD estimate as low or lower than the actual δ estimated in (4) is given by $G(\delta)$. This statistic represents a p-value for the hypothesis that $\delta = 0$, based on a non-parametric permutation test over the weeks in the sample. The intuition underlying this test is straightforward: if the experiment had a significant effect on demand, we would expect the estimated coefficient to be in the lower tail of estimated effects when we replicate the analysis for hypothetical "placebo" weeks.[10] Since this permutation test does not make parametric assumptions about the unobserved error structure, it does not suffer from the overrejection bias in the standard t-test (Bertrand et. al. 2003).

To illustrate this method, Figure 1a plots the empirical cdf G when the dependent variable is weekly revenue per category. The vertical line denotes the treatment effect estimate of $\delta = -\$12$. In this case, $G(\delta) < 0.03$, indicating that the hypothesis that the experiment had no effect is rejected at conventional significance levels. One can analogously implement placebo tests across categories instead of time, by permuting the *treatcat* variable across sets of 14 other categories chosen from the set of control categories, while keeping the *treattime* variable fixed. Figure 1b plots the analogous empirical cdf G for category permutations and again shows that $G(\delta) < 0.05$.

Combining the two dimensions, we implement a 2-way permutation test by estimating the model for various combinations of placebo treatment periods and placebo treatment categories. Since there are a large number of such combinations, we implement a randomization inference procedure, choosing 100 random subsets of 15 control categories for each week in the sample (thereby obtaining 6,100 $\widehat{\delta}_t$ values). We then compute the $G(\delta)$ value for each specification in Table 4 using the resulting empirical cdf of placebo estimates. In all cases, the null hypothesis that $\delta = 0$ is rejected by both the t-test and the non-parametric permutation test.

As an alternative method of probing the robustness of our identification strategy, we consider subsets of our large set of "controls" across time, categories, and stores. In columns 2-4 of Table 5, we report three difference-in-difference estimates, exploiting each pair of these counterfactuals separately. In column 1, we restrict the sample to the treatment products, and compare across stores and time. In column 2, we restrict the sample to the treatment store, and compare across products and time. In column 3, we restrict the sample to the treatment time period, and compare cross-sectionally across stores and

[10] This test can be viewed as an extension of Fisher's (1921) "exact test" for an association between two binary variables. See Rosenbaum (1986) for more on permutation tests.

categories. Reassuringly, all three DD estimates are roughly similar to the baseline DDD estimate reported in Tables 3 and 4. Other changes in the control set – such as restricting the control time period to the three months immediately before the intervention or limiting the control categories to nearby products or products in other aisles – also do not affect the estimates significantly (not reported).

Supplementary Tests. Some studies in the marketing literature (e.g., Anderson and Simester 2003) find that demand drops discontinuously when prices cross integer thresholds (such as $3.99 vs. $4.01), and that retailers respond to this by setting prices that end in '9' to maximize profits. Indeed, the retailer we study sets most products' *pre-tax* prices just below the integer threshold – an observation that in itself supports our claim that individuals focus on the pre-tax rather than the tax-inclusive price, since the tax-inclusive price is usually above the integer threshold. In this vein, it is interesting to ask whether demand for the products whose price crossed the integer threshold once taxes were included (e.g. $3.99 + Sales Tax = $4.28) fell more than demand for products whose price did not cross the integer threshold. We estimated a model analogous to (4) at the product level, including an interaction of the treatment variable with a dummy for the product price crossing the integer threshold. We find little systematic evidence that demand fell more for the products that crossed the threshold, though the interaction effect is imprecisely estimated given the small sample.

We also tested whether the intervention in the treatment categories had "spillover" effects onto the nearby control categories. In particular, if showing tax-inclusive prices reduces demand simply because individuals learn that these products are taxed, demand of nearby similar products might also fall. We find no evidence of such a spillover effect: when we estimate (4) with separate indicators for "adjacent" vs. "non-adjacent" control categories, we find no significant differences in demand during the treatment period across these two types of control categories. This suggests that the effects of the intervention were confined strictly to the products for which tax-inclusive prices were posted, a result that is useful in narrowing the class of models that fit the data.

5 Evidence from Observational Data on Alcohol Sales

5.1 Research Design

We turn now to our second empirical test of whether tax salience affects behavioral responses to taxation: comparing the effect of increases in prices and taxes on demand. We implement this strategy by focusing on alcohol consumption, exploiting the fact that alcohol is subject to two state-level taxes in the U.S.: (1) an *excise* tax that is levied at the wholesale level and thus included in the price posted on the shelf (or on a restaurant menu) and (2) a *sales* tax, which applies to alcohol (except in Vermont and Kansas) but is added at the register. Hence, the excise tax (t^E) is as salient as the posted price (and effectively serves as an instrument for the posted price). The sales tax (t^S) is less salient because it is not included in the posted price in any state.

Our research design takes state-level legislated changes in the sales and excise tax rates as exogenous, and examines the effects of these reforms on alcohol consumption. Replacing p with $(1 + t^E)$ in equation (1), we obtain the following specification for aggregate alcohol demand as a function of the excise tax, sales tax, and the fraction of individuals who pay attention to the sales tax (θ).

$$\log x(t^E, t^S, \theta) = \alpha + \beta \log(1 + t^E) + \theta\beta \log(1 + t^S) \tag{6}$$

Since both the tax rates and alcohol consumption are highly autocorrelated series, we estimate this model in first-differences. Letting t index time (years) and j index states, define the difference operator $\Delta x = x_{jt} - x_{j,t-1}$. Introducing a set of other demand-shifters (covariates) X and an error term ε_{jt} to capture idiosyncratic state-specific demand shocks, we obtain the following estimating equation by first-differencing (6):

$$\Delta \log x_{jt} = \alpha_0 + \beta\Delta \log(1 + t^E_{jt}) + \beta\theta\Delta \log(1 + t^S_{jt}) + X_{jt}\rho + \varepsilon_{jt} \tag{7}$$

We estimate (7) using OLS and test the hypothesis that the estimated gross-of-excise-tax and gross-of-sales-tax elasticities are equal, as would be predicted if $\theta = 1$. This empirical strategy complements the experimental intervention by offering evidence on the importance of salience over a longer horizon.

An important simplifying assumption we made in deriving (6) is that both the excise tax and sales tax apply only to alcohol (and not the composite commodity y that represents all other consumption). In reality, the sales tax applies to a broader set of goods than alcohol: based on statistics on sales tax revenues and tax rates, approximately 40% of consumption is subject to sales taxation in the average state. Hence, a 1% increase in t^S changes the relative price of x and y less than a 1% increase in t^E. After presenting our baseline findings, we present a calibration argument which suggests that the degree of bias from this issue is unlikely to explain the magnitude of the estimated difference between the two elasticities.

5.2 Data and Summary Statistics

For simplicity, we focus exclusively on beer consumption, which accounts for the largest share of alcohol consumption. Data on aggregate annual beer consumption by state are available from the National Institute of Alcohol Abuse and Alcoholism (2006) from 1970-2003. These data are compiled from administrative state tax records, which contain information on total gallons of beer sold by wholesalers, because this measure determines tax liabilities (see Nephew et. al. 2004 and Lakins et. al. 2004 for details on data construction). Note that these data are more precise than comparable data from surveys of alcohol consumption because they reflect total consumption in each state rather than a sample of the population.

State excise tax rates on beer are obtained from the Brewer's Almanac (various years), published annually by the Beer Institute. State sales taxes are obtained from the World Tax Database (2006) at the University of Michigan. The state sales tax is an ad valorem tax (proportional to price), while the excise tax is a specific tax (specified as cents per case of beer). We convert the excise tax rate into percentage terms by dividing the beer excise tax amount by the average cost of a case of beer in the United States, as calculated by the Bureau of Labor Statistics (BLS).[11] We also obtain annual data on state-level aggregate covariates such as the state unemployment rate and state per capita income from the BLS

[11] Because Hawaii and Alaska have higher price levels than the continental United States we follow Census Bureau practice and adjust their price levels up by 15 and 25 percent, respectively, when calculating the percentage excise tax rates for those states. None of our results are affected by this adjustment, nor by excluding these states entirely.

and BEA.

Table 6 provides summary statistics for this dataset. Between 1970 and 2003, the average cost of a case of beer (twenty-four 12 oz. cans) was $14.05 in real 2000 dollars. Mean per capita consumption of beer during this period was 23 gallons per year, equivalent to roughly 240 cans. The (unweighted) mean excise tax over state-year pairs is $0.51 per case, equivalent to 5.7 percent of the average price. The mean sales tax applied to alcohol is 4.2 percent.

The excise tax rate varies significantly more than the sales tax rate both across states and over time: the standard deviation of excise tax rates is 3 times as the standard deviation of sales tax rates. In certain states – e.g. Alabama, Georgia, North Carolina, and South Carolina – the effective tax rate has exceeded 25 percent in some years. Excise taxes have fallen as a percentage of price over time because the dollar value of the tax has generally not kept up with inflation. In contrast, sales tax rates have increased secularly over time.

5.3 Results

We begin with a simple graphical analysis to illustrate the relationship between alcohol consumption and taxes in Figures 2a and 2b. These figures plot changes in log beer consumption per capita against log changes in the gross-of-excise-tax price $\Delta \log(1 + t^E)$ and the gross-of-sales-tax price $\Delta \log(1 + t^S)$. To make the range of changes in the excise tax comparable to the smaller range of changes in the sales tax, we restrict the range of changes in alcohol tax rates to ± 2 percentage points. Without this restriction, results are similar and the effect of the excise tax on beer consumption is even more precisely estimated.

Figure 2a shows that there is a strong negative relationship between changes in the beer excise tax and beer consumption. Figure 2b shows that the relationship between beer consumption and sales taxes is considerably flatter. These figures suggest that excise taxes and sales taxes have differential effects on beer consumption.

To quantify the magnitude of the differences in the excise and sales tax elasticities, we estimate variants of the model in (7). In column 1 of Table 7, we estimate the model including only year fixed effects. A 1% increase in the gross-of-excise-tax price is estimated to reduce beer consumption by 1.17% (i.e. $\varepsilon_{x,1+t^E} = -1.17$). In contrast, a 1% increase in the gross-of-sales-tax price is estimated to reduce beer consumption by only 0.36% (i.e. $\varepsilon_{x,1+t^E} = -0.36$). The null hypothesis that the excise and sales tax elasticities are equal is rejected with $p = 0.05$. In columns 2 and 3, we investigate the robustness of this result to the inclusion of state-level controls such as population growth, income growth, and changes in the unemployment rate. The difference between the effects of the excise and sales taxes remains large, and tests of equality of the tax elasticities are rejected. Finally, in column 4, we take differences over three years instead of one year and re-estimate the baseline model. Results are broadly similar, suggesting that even in the long run, an increase in the excise tax rate has a large negative effect on alcohol consumption, while a similar increase in the sales tax does not. Hence, consistent with the evidence from the experimental intervention, this analysis suggests that θ is close to zero: most individuals appear to focus on posted prices rather than full tax-inclusive prices when making consumption decisions.

6 Information vs. Tax Salience: Survey Evidence

The evidence documented thus far indicates that behavioral responses to commodity taxation depend substantially on whether taxes are included in posted prices. There are two potential explanations for this finding. One is that customers are uninformed about the sales tax rate or the set of goods subject to the sales tax. In this case, showing the tax-inclusive price tags may have provided new information about tax rates, leading to a reduction in demand. An alternative explanation is that individuals normally do not compute the tax-inclusive price when shopping, and focus instead on the pre-tax price, which is more salient because it does not entail any computation. Distinguishing between the information and salience mechanisms is useful in developing a model that matches the evidence.

There is some suggestive evidence pointing toward the salience mechanism in the preceding empirical analysis. The fact that the experimental intervention had no detectable "spillover" effects on the taxable categories adjacent to the treatment group suggests that individuals did not simply learn that these types of goods were subject to sales tax. Similarly, the return of demand to pre-experiment levels after the intervention ended suggests that there were no persistent learning effects. In the alcohol consumption analysis, the persistence of the difference between the excise and sales tax elasticities over longer horizons also points toward the importance of salience. Individuals continue to respond less to the sales tax even after they have had considerable time (e.g. 2 or 3 years) to acquire new information.

To test between the information and salience hypotheses more directly, we surveyed 91 customers entering the treatment store in August 2006 about their knowledge of sales taxes. Survey respondents were offered small in-kind incentives such as candy bars and sodas to spend a few minutes filling out the survey, which is displayed in Exhibit 2. After collecting basic demographic information, the survey asked individuals to report whether each of eight goods (e.g. milk, cookies, beer) were subject to sales tax or not. A number of individuals remarked while filling out the survey that they did not think about taxes while shopping, and therefore were hesitant to report which goods were taxed. These individuals were encouraged to mark their best guess, in order to avoid nonresponse bias and maximize data on tax perceptions. To assess whether knowledge of taxes is correlated with experience, we also asked whether individuals had purchased each of these goods recently. Finally, we asked three separate questions about knowledge of tax rates – the local sales tax, the state income tax, and the federal estate tax.

The results of the survey are summarized in Figure 3. Knowledge about sales taxes is generally quite high. The median respondent answered 7 out of 8 of the questions about taxable status of the goods correctly. The general pattern that people appear to know is that "food is not taxed, inedible items and 'sin' goods are taxed." For example, more than 80 percent knew that milk is not taxed and that toothpaste is taxed. More than 90 percent answered correctly that beer and cigarettes are taxed. Exceptions to this general heuristic – soda and cookies – led the most errors. In California, carbonated beverages are subject to sales tax, while cookies (junk food) are not. These two goods accounted for the largest share of mistakes: 25% answered incorrectly that Coca Cola is untaxed, while 35% answered incorrectly that cookies are taxed. Knowledge of the sales tax rate is also quite good. Almost 80 percent reported the sales tax rate to within 0.5 percentage points of the true rate, and 15 percent answered exactly 7.375 percent.

We also explored whether knowledge about sales taxes varies by demographic groups.

Knowledge of taxes – measured as fraction of items whose tax status was identified correctly or deviation in reported sales tax rate from the true rate – is high across all levels of education, among both men and women, and among both single and married individuals. Age and the number of years lived in California are also uncorrelated with knowledge of taxes. Individuals who answered the income and estate tax questions correctly were no more likely to get the sales tax questions correct. Multivariate regressions indicate that these factors do not jointly predict tax knowledge either.

Only 8% of individuals answered the estate tax question correctly (<2%), consistent with the results of other surveys. On the income tax question, many respondents had trouble distinguishing the California state income tax from the federal income tax, and reported rates that are more consistent with federal tax rates. Knowledge of sales taxes may be greater than knowledge of income or estate tax rates because consumers see the sales tax rate repeatedly (e.g., on receipts), but only see the income and estate tax rates occasionally (if at all).

In summary, most individuals are well informed about commodity tax rates when their attention is drawn to the subject. This finding, coupled with the evidence that behavioral responses to taxation are larger when taxes are included in posted prices, implies that many individuals choose not to compute tax-inclusive prices when making consumption decisions. As a result, tax salience affects behavioral responses to taxation.

7 A Model of Bounded Rationality and Taxation

The preceding evidence indicates that the canonical model used to analyze tax policies in public finance – in which agents take all taxes into account when making decisions – fails to describe observed behavior. What alternative model does match the finding that salience affects behavioral responses to taxation? The remainder of the paper focuses on constructing a model consistent with the findings above, while providing a tractable framework in which to analyze standard topics such as incidence and efficiency.

We model limited attention to taxation using a model of "boundedly rational" agents, who face small costs of cognition in choosing optimal behavior. Bounded rationality is an attractive model of failures to optimize in public finance applications because it has a clearly defined measure of welfare. In this model, social welfare is simply total utility less cognitive costs. Since there are no differences between an agent's preferences and the preferences that enter the social welfare function, one can use revealed preference to calculate the efficiency and welfare consequences of policies. This property turns out to be useful in deriving formulas for efficiency and incidence of taxes using this behavioral model.

7.1 Model Setup

Consider a static model where an agent has an additively separable quasiconcave utility function $u(x) + v(y)$ over two goods, x and y. Normalize prices to 1. Let t denote the legislated (true) tax rate on good x and assume that good y is not taxed. Let Z denote the agent's total wealth, which we take to be fixed initially and later allow to depend on labor supply.

To model bounded rationality, assume that the agent is imperfectly informed about the tax rate on good x (and knows with certainty that good y is not taxed). The agent has a prior distribution on the tax rate $F(t)$, which has an upper bound $t \leq \bar{t}$. That is, the agent knows with certainty that the true tax lies below $\bar{t}$. The agent can compute the exact value of the tax on good x by paying a fixed cognitive cost c. If he chooses not to compute the true tax, the agent uses the heuristic of assuming that the the tax rate is $t^p < \bar{t}$, which we refer to as the "perceived tax," when making consumption decisions.

Quasilinear Utility. It is instructive to begin with a case without income effects. Assume that utility is quasilinear in the untaxed good y, i.e. $v(y) = y$. Then utility is given by

$$u(x) + y.$$

Assume $u'(x) > 0$, $u''(x) < 0$, $\lim_{x \to 0} u'(x) = \infty$, and $u'(Z) < 1$ to guarantee an interior optimum at any $t \geq 0$.

The agent makes two choices: whether to learn the true tax rate t and how to allocate consumption given his knowledge of the tax rate. This problem can be divided into a two-stage maximization: (1) choose an optimal bundle for any given perceived tax rate t^p; (2)

decide whether to spend c on computing the true tax rate.

A key difficulty in constructing a model where agents misperceive true prices is that the consumption choices must nevertheless satisfy the true budget constraint in order to be feasible:

$$x(1+t)+y=Z$$

Thus, one must specify how the agent chooses x and y to maximize his utility with a possibly misperceived tax rate t^p while satisfying the true budget constraint. A natural assumption in the case where good x is small relative to the overall budget is that the agent chooses x first, given his true perceived tax, and then spends his true residual wealth on y:

$$y=Z-(1+t)x$$

The issue of how the budget constraint is satisfied is particularly important when utility is not quasilinear, and we defer detailed discussion of this assumption to that case. Under the assumption that x is chosen first, at any given perceived tax rate t^p, the agent sets consumption of x as

$$x^p = \arg\max u(x) + Z - (1+t^p)x.$$

The optimal x^p satisfies

$$u\prime(x^p) = 1 + t_p$$

and hence the agent consumes

$$y = Z - (1+t)x^p$$

Characterization of the Cognitive Decision. Now consider the decision of whether to pay the deliberative cognition cost and compute the true tax on x. A partial characterization of this decision can be obtained by asking the question, "How much does the agent's utility rise (measured using a money metric) if he computes the true tax rate in the state where $t = \bar{t}$, the maximum possible value?" If this value is below the cost of cognition c, then the agent will choose not to compute the true tax irrespective of the shape of his prior distribution $F(t)$. By answering this question, we can identify a sufficient condition on the range of tax rates such that the agent will choose not to compute the exact tax rate for a given cost c.

Let $x^*(t)$ and $y^*(t)$ denote the agent's optimal choices of x and y when he knows that the true tax rate is t. Note that $u\prime(x^*(t)) = 1+t$. Then the agent's utility gain from computing the tax rate in the state where the true tax is $\bar{t}$ is

$$\begin{aligned} G(\bar{t}) &= u(x^*(\bar{t})) + Z - (1+\bar{t})x^*(\bar{t}) - [u(x^p) + Z - (1+\bar{t})x^p] \\ &= u(x^*(\bar{t})) - u(x^p) + (1+\bar{t})(x^p - x^*(\bar{t})). \end{aligned}$$

Note that $G(\bar{t})$ is a money metric since utility is quasi-linear in y. Taking a second-order Taylor approximation of $u(x)$ around x^* and using the first order condition for x^*, we obtain:

$$\begin{aligned} G(\bar{t}) &\simeq u(x^*) - [u(x^*) + u'(x^*)(x^p - x^*(\bar{t})) + \frac{1}{2}u''(x^*)(x^p - x^*(\bar{t}))^2 + (1+t)(x^p - x^*(\bar{t})) \\ &= -\frac{1}{2}u''(x^*)(x^*(\bar{t}) - x^p)^2 \end{aligned}$$

Finally, use the linear approximation $x^*(\bar{t}) - x^p = \frac{\partial x}{\partial t_p}(\bar{t} - t_p)$ to obtain:

$$G(\bar{t}) \simeq -\frac{1}{2}\frac{\partial x}{\partial t_p}(\bar{t} - t_p)^2 = \frac{1}{2}\varepsilon_{x,p}\frac{x}{1+t_p}T^2 \tag{8}$$

where $\varepsilon_{x,p} = -\frac{\partial x}{\partial t_p}\frac{1+t_p}{x}$ denotes the price elasticity of x and $T = \bar{t} - t_p$ denotes the maximum potential deviation in the tax rate from the perceived tax rate. This expression shows that the gain from computing the exact tax rate in the state where t is at its maximum varies with the square of maximum possible error in the tax rate. Hence, the welfare costs from failing to compute the correct tax rate when optimizing consumption are second-order. Given a cognitive cost c, the agent will therefore rationally choose not to compute the exact tax rate if

$$\frac{1}{2}\varepsilon_{x,p}\frac{x}{1+t_p}T^2 < c \Longrightarrow T < [2\frac{c(1+t_p)}{x\varepsilon_{x,p}}]^{1/2}.$$

The threshold T is a lower bound for the range of taxes which the agent will rationally ignore. This threshold is increasing in c, as one would expect, reflecting the intuition that higher cognitive costs make it rational to ignore a broader range of taxes. Next, consider the minimum width of the range of ignored taxes relative to the cost of cognition:

$$\frac{T}{c} = [2\frac{(1+t_p)}{x\varepsilon_{x,p}}]^{1/2} \times c^{-1/2} \tag{9}$$

This expression shows that as cognitive costs become small, the range of taxes that are ignored grows small at a slower (square root) rate. Hence

$$\lim_{c->0} \frac{T}{c} = \infty,$$

showing that for infinitesimally small cognitive costs, the range of taxes that are rationally ignored remains non-negligible. Mathematically, the source of this result is the envelope condition that arises from agent optimization, which guarantees that small changes in behavior (as would be induced by learning about small deviations in tax rates) have negligible effects on utility. The envelope condition causes the first-order (u') terms to drop out in $G(\bar{t})$, leading to the result that the gain from computing tax rates is bounded below by a second-order function of the maximum deviation in the tax rate. Figure 4 illustrates the result geometrically. In this figure, the individual's welfare loss from failing to optimize relative to the true tax is given by the lost consumer surplus, triangle A. The size of this triangle is given by $\frac{1}{2}(\bar{t}-t_p)(|\frac{\partial x}{\partial t_p}|(\bar{t}-t_p))$, which is precisely the expression for $G(\bar{t})$. As the maximum tax rate $\bar{t}$ approaches t, the size of this triangle diminishes at a second-order rate because both its height and width diminish linearly.

The economic intuition for the result is that there is little to be gained from adjustments following small changes in prices when one is already at an optimum to begin with. Hence, an agent who has small cognitive costs will not pay attention to the details of tax policies that he thinks are likely to induce at most small changes in true prices when choosing a consumption bundle. This result parallels Akerlof and Yellen's (1986) well-known result that near-rational firms will ignore monetary shocks, leading to sticky prices.

The practical implication of this result is that small cognitive costs can lead to substantial ignorance of taxes. We quantify what "small" and "substantial" mean by calculating the minimum range of taxes that are ignored for various levels of the cognitive cost c using the formula in (9). Table 1 presents the results of this exercise for various values of the demand elasticity $\varepsilon_{x,p}$ and expenditure on the good, x. The range of taxes that are ignored with small cognitive costs is large. For example, Table 1 shows that tax changes of +/-10 percentage points on a commodity on which the agent spends $\$10,000$ and has a demand elasticity of -1 are ignored if the cognitive cost $c > \$50$.

General Case. The basic result established above also applies in the general case when utility is not quasilinear. In this case, the agent's utility is

$$u(x) + v(y)$$

His perceived budget constraint is

$$x(1 + t_p) + y = Z$$

His actual budget constraint, which must be satisfied, is

$$x(1 + t) + y = Z$$

If the agent responds only to the perceived tax rate at the margin, he will set

$$u'(x) = (1 + t_p)v'(y). \quad (10)$$

This first order condition determines the consumption of x relative to y but not the *level* of consumption of x unless utility is quasilinear in y, in which case the marginal dollar is always allocated to y. Therefore, in choosing the level of x when utility is not quasilinear, the agent needs to know his total net-of-tax income. The key difficulty in solving this

There are two ways to "close the model" given this problem, which can be thought of as variations in the order in which consumption of the two goods is chosen: (1) Choose x first. The agent chooses x^p based on (10) given total income Z, and finances his spending on y using what's left: $y = Z - x^p(1 + t)$. (2) Choose y first. The agent chooses y^p based on (10) given total income Z, and finances his spending on x using what's left: $x = \frac{Z - y^p}{1+t}$. We focus in this paper on the solution where the agent chooses x first; the key qualitative results hold in the second case as well.[12]

It is straightforward to establish that second order costs of cognition will again generate a first-order band of taxes that are rationally ignored, as in the quasilinear case. Let $x^*(t)$ and $y^*(t)$ denote the optimal consumption allocation given a true tax rate t. Then the gain in utility from computing the true tax in the state where $t = \bar{t}$ is:

$$G(\bar{t}) = u(x^*(\bar{t})) + v(Z - (1+t)x^*(\bar{t})) - [u(x^p) + v(Z - (1+t)x^p)]$$

Second-order Taylor expansions and algebra along the lines above yields the following expression:

$$\begin{aligned} G(\bar{t}) &= -\frac{1}{2}(\bar{t} - t^p)^2 \frac{x\varepsilon_{x,p}}{1+t^p}\lambda \\ \text{where } \lambda &= v'(y)\frac{x}{y}\frac{\varepsilon_{x,p}}{\varepsilon_{y,p}} \end{aligned}$$

The expression for $G(\bar{t})$ in the general case differs from that in the quasilinear case in (8) by the factor λ. The factor λ corresponds to the income effect of a tax-change, which leads to a shift in demand for y relative to demand for x beyond the pure price effect. This income effect does not emerge in the quasilinear case, leading to $\lambda = 1$. Nevertheless, even when utility is not quasilinear, $G(\bar{t})$ remains a quadratic function of T, implying that second-order cognitive costs will still generate a first-order range of inattention to taxes.

We conclude that bounded rationality can explain the empirical evidence on tax salience documented in sections 3 and 4. The welfare gains from optimizing relative to true taxes on cosmetics or alcohol are likely to be very small, particularly if the individual knows that the tax rate on such goods is at most $\bar{t} = 10\%$. Therefore, individuals with limited attention spans or cognitive costs may rationally use the heuristic of focusing on the "salient" posted price rather than computing the tax-inclusive price for each product.

[12] The model can be easily extended to make the choice of the decision rule endogenous by allowing the agent to calculate the expected utility of each rule, taking into account his uncertainty about the tax rate. The agent then follows the rule that yields higher expected utility, which ultimately results in the behavior characterized here. In the quasilinear case, it is easy to establish that choosing the good with diminishing marginal utility first (x) is optimal using Jensen's inequality. See Reis (2006) for a related analysis of the choice between a consumption and savings rule in a lifecycle savings model.

7.2 Individual Welfare vs. Social Welfare

A potential concern with the bounded rationality framework is that if agents experience second-order welfare losses from ignoring taxes, the effects of these taxes on revenue and social welfare may also be second-order. Put differently, can bounded rationality explain inattention to taxes only when they have negligible effects on the government's objective? If this were the case, the set of policies for which bounded rationality is a useful description of behavior would inherently be unimportant from a policy perspective.

In this subsection, we address this issue by computing the change in revenue and social welfare from a tax increase that is rationally ignored by a boundedly rational agent. Our main result is that tax changes will in general have first-order effects on revenue and social welfare even though they have second-order effects on the agent's utility. We begin with a graphical illustration of the intuition for this result, and then establish it formally. For simplicity, we focus on the case of quasilinear utility throughout this section; we derive formulas for the welfare costs of taxation with a general utility in section 6.

Graphical Intuition. Consider a good that is initially subject to a tax t_0 that is included in the posted price (and hence fully perceived by agents – $t_p = t_0$). Let us compare the consequences of adding a tax Δt as a further increment to the posted price or as a separate tax paid separately (such as a sales tax). To do so, we analyze the effects of such a tax increase on individual welfare, tax revenue, and social welfare. Figure 4 presents this analysis for the case of quasilinear utility, where demand for the taxed good (x) depends only on its tax-inclusive price, and there are no income effects.

First consider the case where the tax increase is not salient (e.g., not shown in the posted price). The loss in individual welfare from failing to re-optimize in response to the tax increase is shown by triangle A, whose area equals $\Delta U = -\frac{1}{2}\frac{\partial x}{\partial t}(\Delta t)^2$. As discussed above, ΔU is a second-order function of the change in the tax, Δt. Next, consider the change in tax revenue from the tax increase. The change in revenue is given by the rectangle B, whose area equals $\Delta R = x(t_0)\Delta t$. Since ΔR is a linear (first-order) function of Δt, even though the change in the tax may cause a trivial welfare gain from re-optimization, it can generate a large amount of revenue for the government. Finally, consider the change in social welfare from this tax. Since there is no change in consumption of the good when the agent ignores the tax change, the change in social welfare $\Delta W = 0$, because the un-salient tax is equivalent to a lump-sum tax when utility is quasilinear (see section 6 for further details, including a discussion of why this equivalence result fails when utility is not quasilinear).

Now compare these results to the case where the tax is salient (e.g., included in the posted price). Here, the agent pays no cognitive cost to respond to the tax change, and thus re-optimizes his consumption choice. The change in revenue from the salient tax increase is $\Delta R_s = x(t_0)\Delta t - t_1 \frac{\partial x}{\partial t}\Delta t$, where the latter term reflects the loss in revenue from the agent's behavioral response. The change in social welfare if the individual were to respond to the tax is given by the "Harberger trapezoid" with area

$$\Delta W_s = -\frac{1}{2}\frac{\partial x}{\partial t}(\Delta t)^2 + t_0 \frac{\partial x}{\partial t}(\Delta t).$$

Note that the area of this trapezoid varies at a first-order level with Δt. This analysis shows that whether a tax increase is made salient or not has a first-order effect on social welfare and government revenue, even though the welfare costs to the agent of ignoring the tax change are second-order. Intuitively, the difference arises because the agent imposes a fiscal externality on the government when he reduces consumption of the good in response to the tax. This loss in revenue constitutes a first-order loss in social welfare, but does not affect the agent's private welfare because he does not receive that surplus.

Formal Derivation. We now derive the formulas above using a quadratic approximation to the utility function $u(x)$. The expression for ΔU was already derived in the previous section. Tax revenue with a tax rate of t is given by $R(t) = tx(t)$. When the agent has a cognitive cost $c > \Delta U$, behavior is unchanged and the change in revenue is simply

$$\Delta R = \Delta t x(t_0)$$

When $c < \Delta U$ (or when the tax is salient to begin with), the agent will re-optimize consumption. In this case,

$$\begin{aligned}\Delta R &= t_1 x_1 - t_0 x_0 = \Delta t x_0 - t_1 \frac{\partial x}{\partial t}\Delta t \\ &= \Delta t x_0 - \frac{t_1}{1+t_0}\varepsilon_{x,p} x_0 \Delta t\end{aligned}$$

Social welfare is defined in the quasilinear case as

$$\begin{aligned}W(t) &= u(x(t)) + y(t) + R(t) \\ &= u(x(t)) + Z - (1+t)x(t) + tx(t) \\ &= u(x(t)) - x(t) + Z\end{aligned}$$

Hence, the deadweight loss from raising tax from t_0 to t_1 is

$$\Delta W = [u(x(t_1) - x(t_1)] - [u(x(t_0) - x(t_0)]$$

If the agent does not perceive the tax increase, $x(t_1) = x(t_0)$ and $\Delta W = 0$. If the agent does re-optimize in response to the tax change, approximations analogous to those used above give

$$\begin{aligned}\Delta W &= t_0 \frac{\partial x}{\partial t}\Delta t + \frac{1}{2}\frac{\partial x}{\partial t}(\Delta t)^2 \\ &= \frac{t_0}{1+t_0}\varepsilon_{x,p} x_0 \Delta t + \frac{1}{2}\frac{\varepsilon_{x,p}}{1+t_0} x_0 (\Delta t)^2\end{aligned}$$

Calibrations. To illustrate the quantitative importance of the differences between individual welfare, social welfare, and government revenue, we report calibrations similar to those in Table 5. We consider a good on which the agent spends $x = \$10,000$, and compute ΔU, ΔR, ΔR_s, and ΔW_s for a range of tax increases, starting from an initial rate of $t_0 = 10\%$. For example, a tax increase of 10% raises revenue by 10% and (if imposed saliently) reduces social welfare by 2.4%; but the loss in welfare from failing to optimize with respect to this tax change is only $\Delta U = 0.4\%$.

The policy relevance of this result can be seen with the following example. Suppose boundedly rational agents perceive payroll taxes (e.g. SS, DI, UI) as a pure tax on the margin because the tax-benefit linkage is opaque. By reforming policy so that this link is more transparent, the government can reduce boundedly rational agents' perceived tax rates. This reform will only have small benefits for agents in terms of improved welfare due to better optimization (which is why they ignore the details of these policies in the first place). However, the social surplus gained from making these policies more transparent could be large. Lowering agents' perception of taxes raises labor supply and raises government revenue through the income tax. This fiscal externality leads to an increase in social surplus: for example, if the government has a fixed revenue requirement, it can lower tax rates, thus raising individuals' welfare.

The broader point of this analysis is that tax policies that are "important" from the government's perspective may nevertheless create small utility costs from failures to optimize at the individual level. Hence, bounded rationality and salience effects could be relevant in the analysis of many of the tax policies that have received attention in the academic and policy debate.

8 Tax Heuristics and Stylized Facts

In addition to matching the empirical evidence on salience, the bounded rationality framework makes sharp predictions about the types of heuristics that agents will use when faced with complex tax policies. These heuristics can be identified by asking, "How much would an individual pay to learn about each aspect of the tax code (marginal rates, average rates, kink points, etc.)?" By conducting this calculation for each feature of the tax code, one can at least partially characterize how an agent with a cognitive cost c will perceive the tax schedule. This heuristic can then be used to predict behavioral responses to tax policies more precisely.

In this section, we illustrate this idea using three examples: average vs. marginal tax rates, cliffs vs kink points, and behavioral responses among the rich vs. poor. We then discuss how the implied heuristics match stylized facts about behavioral responses to taxation.

Average vs. Marginal Rates. Liebman (1998) and Liebman and Zeckhauser (2005) point out that agents tend to be much more aware of and responsive to average tax rates relative to marginal rates.[13] In contrast, much of the theoretical and empirical public finance literature has focused on analyzing marginal rates under the presumption that these rates are the key determinants of behavior for an unboundedly rational agent. Here, we evaluate which rate is most relevant to a boundedly rational agent. In particular, we compare the welfare gains from optimizing relative to the marginal tax rate on labor income and the average tax rate on labor income.

We find that the welfare gains from computing and responding to the average rates exceeds the welfare gains from responding to the marginal rate significantly. Mistakes in the average rate can lead to errors in budgeting of consumption goods, with potentially greater costs to utility than errors in marginal labor supply decisions. To see this point, consider a utility function of the following form:

$$u(f,h,l) = (h^{-\rho} + f^{-\rho})^{-\frac{1}{\rho}} - \frac{1}{1+1/\sigma} l^{1+1/\sigma}$$

where h and f are untaxed consumption goods and l denotes labor supply. The agent's true budget constraint is

$$f + h = Z = y(1-t_y) + w(1-t_m)l$$

where Z denotes total income, y denotes unearned income (e.g. income from spousal labor supply, capital income, or wage income from another job), w denotes the marginal wage rate, t_y is the tax on unearned income, and t_m is the marginal tax rate on labor income.

The agent's optimization problem can be solved using two-stage budgeting: first choosing the optimal consumption bundle for a given level of income Z, and then choosing l to maximize total utility:

$$\begin{aligned} &\max v(y((1-t_y) + w(1-t_m)l) - \frac{1}{\sigma} l^{\sigma} \\ &\text{s.t. } v(Z) = \max_h (h^{-\rho} + (Z-h)^{-\rho})^{-\frac{1}{\rho}} \end{aligned}$$

[13] The demand for information about average rates is indirectly evident in wage and tax return statements, which typically show average tax rates and net-of-tax income, but typically do not show marginal rates.

Given the functional form of utility, $v(Z)$ is homogeneous of degree 1 in Z. Hence, indirect utility in the second stage maximization problem over l is effectively quasilinear. Thus, the optimal choice of l satisfies

$$l^* = k(w(1 - t_m))^\sigma$$

where $k = 2^{-(1+\frac{1}{\rho})}$. Hence, under this utility function, the agent's optimal labor supply choice depends only on the marginal tax rate and not the average rate.

To simplify, first consider the case of Cobb-Douglas utility (where $\rho = 0$). In this case, the utility function reduces to

$$u(f, h, l) = h^{\frac{1}{2}} f^{\frac{1}{2}} - \frac{1}{1 + 1/\sigma} l^{1+1/\sigma}$$

Let t_y^p and t_m^p denote the agents perceived tax rates, and assume without loss of generality that consumption of good h is chosen first. Then the agent's consumption and labor supply allocation is

$$\begin{aligned} l^p &= [w(1 - t_m^p)]^\sigma \\ h^p &= \frac{1}{2}[y((1 - t_y^p) + w(1 - t_m^p)l^p] \\ f^p &= y((1 - t_y) + w(1 - t_m)l^p - h^p \end{aligned}$$

In Table 8, we present calibrations showing the value of providing the agent with two types of information: (1) information on the marginal rate, t_m, and (2) information on total net-of-tax income, $Z = y((1 - t_y) + w(1 - t_m)l^p$. Note that (2) is equivalent to providing information about the average tax rate at the boundedly-rational level of labor supply. The calibrations show that providing information about net-of-tax income is far more valuable than information about the marginal rate. The intuition underlying this result is straightforward. Knowledge about the average tax rate is valuable in budgeting consumption: if the agent misestimates the average rate by +/-10%, he misallocates a large amount of money from f to h. In contrast, if the elasticity of earned income with respect to t_m is not too large, the marginal welfare cost of under or over-supplying labor because of the misperceived wage is much smaller. More concretely, if agents underestimate their average tax rate, they may substantially over-spend on housing relative to food, with a sharp utility cost. If they underestimate their marginal tax rate, they may work somewhat more relative to their optimum, with lower welfare costs. This example illustrates that boundedly rational agents are more likely to know and respond to average rates than marginal rates.

To highlight the key role of the budgeting distortion in this result, consider instead the case where h and f are perfect substitutes: $\rho = -1$. In this case, the utility does not depend on how Z is allocated between f and h, and hence $v(Z) = Z$. In this case, the agent would be unwilling to pay anything for information about t_y, since it affects neither his labor supply decision nor his consumption allocation decision. In contrast, the agent would be willing to pay for information about t_m, as in the two-good case analyzed above. Hence, insofar as agents have diminishing marginal utility over goods and therefore make budgeting decisions – an intuitively plausible condition – average rates are likely to be particularly important. Heightened awareness and responsiveness to average rates relative to marginal rates may therefore be consistent with rational behavior by agents who face cognitive costs and face important budgeting decisions.

9 Conclusion

The broad objective of this paper has been to incorporate insights from the literature on psychology and economics into public finance to better understand the consequences of tax policies. Evidence from a field experiment and observational data both show that behavioral responses to taxation of commodities differ significantly based on whether the tax is included in posted prices. Since individuals appear to be well informed about sales taxes when their attention is drawn to the topic, these findings indicate that tax salience has a substantial impact on behavior. This result contradicts the assumption of full optimization against tax schedules in existing public finance models. As a tractable alternative, we constructed a simple model of boundedly rational agents, and showed that small costs of cognition can explain our empirical findings as well as a number of other stylized facts. Somewhat surprisingly, small cognitive costs can affect the efficiency and welfare consequences of a broad range of large-scale tax policies. The model yields simple Harberger-type formulas for incidence and efficiency costs of taxation that can be easily adapted to other applications.

We view our empirical and theoretical analysis as a first step in analyzing tax policy in an environment that departs from the traditional unbounded rationality framework. This basic approach could be generalized and refined in a number of dimensions. Empirically, it would be interesting to revisit studies that have estimated behavioral responses to taxation and calculate the utility cost of failing to optimize against the tax changes used for identification. This analysis could shed light on which of the tax reforms used in the literature are most likely to overcome limited attention constraints and identify the underlying price elasticities relevant for efficiency and welfare analysis. Theoretically, there are a number of positive and normative topics that could be analyzed in richer models of bounded rationality and inattention. For example, the formation of tax perceptions may be much more important than the particular marginal rates that an individual faces. Such analysis could shed further light on issues such as consumption taxation (where taxes are likely to be included in posted prices) and the value of tax simplification, topics that have attracted attention in the recent policy debate on tax reform.

References

[1] Amador, M., I. Werning, and G. Angeletos (2006) "Commitment vs. Flexibility" *Econometrica* 74(2): 365-396.

[2] Anderson, E. and D. Simester (2003). "Effects of $9 Price Endings on Retail Sales: Evidence from Field Experiments." *Quantitative Marketing and Economics*, 1, 93-110, 2003.

[3] Akerlof, G. and J. Yellen (1985) "Can Small Deviations from Rationality Make Significant Differences to Economic Equilibria" *American Economic Review* 74 (4):708-720.

[4] Barber, B., T. Odean, and L. Zheng (2005) "Out of Sight, Out of Mind: The Effects of Expenses on Mutual Fund Flows" *Journal of Business.*

[5] Barber, B. and T. Odean (2005) "All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors" Working Paper.

[6] Barlow, T. and M.S. Wogalter (1993) "Alcoholic Beverage Warnings in Magazine and Television Advertisements" *Journal of Consumer Research* 20: 147-156.

[7] Bernheim, D. and Rangel, A. (2007) "Beyond Revealed Preference: Toward Choice-Theoretic Foundations for Behavioral Welfare Economics," Stanford University Working Paper.

[8] Brewer's Alamanac, Beer Institute (various years). Available online at http://www.beerinstitute.org/statistics.asp?bid=200

[9] Brown, C.V. (1968) "Misconceptions About Income Tax and Incentives" *Scottish Journal of Political Economy*, 1-12.

[10] Busse, M., Silva-Risso, J. and F. Zettlemeyer (2006) "$1000 Cash Back: The Pass-Through of Auto Manufacturer Promotions", forthcoming *American Economic Review.*

[11] Chetty, R. (2006) "A General Formula for the Optimal Level of Social Insurance," *Journal of Public Economics* 90: 1879-1901.

[12] Conlisk, J. (1988) "Optimization Cost," *Journal of Economics Behavior and Organization* 9(3): 213–28.

[13] Conlisk, J. (1996) "Why Bounded Rationality" Journal of Economic Literature 34: 669-700.

[14] Davis, S. and Henrekson, M. (2006) "Economic Performance and Work Activity in Sweden after the Crisis of the Early 1990s," *NBER Working Paper No. 12768.*

[15] de Bartolome, C., (1995) "Which Tax Rate Do People Use: Average or Marginal?," *Journal of Public Economics*, 56: 79-96.

[16] DellaVigna, S. and J. Pollet (2006) "Attention, Demographics, and the Stock Market" UC Berkeley Working Paper.

[17] Diamond, P. and E. Sheshinski (1995) "Economic Aspects of Optimal Disability Benefits" *Journal of Public Economics* 57(1): 1-23.

[18] Ellison, G. and S. Ellison (2004) "Search, Obfuscation, and Price Elasticities on the Internet" *NBER Working Paper No. 10570.*

[19] Feldman, N. and Katuscak, P. (2005) "Should the Average Tax Rate be Marginalized?" Charles University, Center for Economic Research and Graduate Education.

[20] Feldstein, M. (1985) "The Optimal Level of Social Security Benefits" *The Quarterly Journal of Economics*, 100(2): 303-20.

[21] Finklestein, A. (2006) "E-Z Tax: Tax Salience and Tax Rates" MIT Working Paper.

[22] Fujii, E. and Hawley, C. (1988) "On the Accuracy of Tax Perceptions" *The Review of Economics and Statistics* 70(2): 344-347.

[23] Gabaix, X. and D. Laibson (2006) "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets" *Quarterly Journal of Economics* 121 (2): 505-540.

[24] Gabaix, X., Laibson, D., Moloche G. and S. Weinberg (2006) "Information Acquisition: Experimental Analysis of a Boundedly Rational Model", *American Economic Review*

[25] Golosov, M., Kocherlakota, N. and Aleh Tsyvinski (2003) "Optimal Indirect and Capital Taxation" *Review of Economic Studies* 70(3): 569-587.

[26] Goolsbee, A. (2000) "What Happens When You Tax the Rich? Evidence from Executive Compensation" *Journal of Political Economy* 108(2): 352-378.

[27] Gruber, J. and B. Koszegi (2001) "Is Addiction "Rational"? Theory and Evidence" *Quarterly Journal of Economics* 116 (4): 1261-1305.

[28] Gruber, J. and E. Saez (2002) "The Elasticity of Taxable Income: Evidence and Implications" *Journal of Public Economics* 84: 1-32.

[29] Hausman, J. (1979) "Individual Discount Rates and the Purchase and Utilization of Energy-using Durables" *Bell Journal of Economics* 10: 33-54.

[30] Hausman, J. and P. Joskow (1982) "Evaluating the Costs and Benefits of Appliance Efficiency Standards" *American Economic Review* 72: 220-225.

[31] Hossain, T. and J. Morgan (2005) "...Plus Shipping and Handling: Revenue (Non) Equivalence in Field Experiments on eBay" *Advances in Economic Analysis & Policy.*

[32] Judd, K. (1985) "Redistributive Taxation in a Simple Perfect Foresight Model" *Journal of Public Economics* 28: 59-83.

[33] Krishna, A. and J. Slemrod (2003) "Behavioral Public Finance: Tax Design as Price Presentation" International Tax and Public Finance 10: 189-203.

[34] Lakins, N.E., Williams, G.D., Yi, H., and Smothers, B.A., 2004. Surveillance Report #66: Apparent Per Capita Alcohol Consumption: National, State, and Regional Trends, 1977-2002. Bethesda, MD: NIAAA, Alcohol Epidemiologic Data System.

[35] Liebman, J. (1998) "The Impact of the Earned Income Tax Credit on Incentives and the Income Distribution" In *Tax Policy and the Economy* 12. James Poterba, ed. Cambridge: MIT Press.

[36] Liebman, J. and R. Zeckhauser, “Schmeduling.” Harvard KSG Working Paper, 2005.

[37] Mackowiak, B. and M. Wiederholt (2006) “Optimal Sticky Prices under Rational Inattention.” Northwestern University Working Paper.

[38] Mankiw, G. M. (1985) "Small Menu Costs and Large Business Cycles: A Macroeconomic Model" *Quarterly Journal of Economics* 100 (2): 529-538.

[39] McCaffery, E.J. and J. Baron (2004) "Heuristics and Biases in Thinking about Tax" *Center for the Study of Law and Politics Working paper No. 21.*

[40] McCaffery, E.J. and J. Slemrod (2004) "Toward an Agenda for Behavioral Public Finance" *Center for the Study of Law and Politics Working paper No. 36.*

[41] Morwitz, V.G., Greenleaf, E.A. and E.J. Johnson (1998) “Divide and Prosper: Consumers’ Reactions to Partitioned Prices” *Journal of Marketing Research* 35: 453-463.

[42] Mullainathan, S. (2002) "A Memory-Based Model of Bounded Rationality" *Quarterly Journal of Economics* 117 (3): 735-774.

[43] Mullainathan S., Schwartzstein J. and A. Shleifer (2006) "Coarse Thinking and Persuasion" Working Paper.

[44] National Institute on Alcohol Abuse and Alcoholism (2006), “Per Capita Alcohol Consumption.” Data available at http://www.niaaa.nih.gov/Resources/DatabaseResources/QuickFacts/AlcoholSales

[45] Nephew, T.M., Yi, H., Williams, G.D., Stinson, F.S., and Dufour, M.C., 2004. U.S. Alcohol Epidemiologic Data Reference Manual, Vol. 1, 4th ed. U.S. Apparent Consumption of Alcoholic Beverages Based on State Sales, Taxation, or Receipt Data. Bethesda, MD: NIAAA, Alcohol Epidemiologic Data System. NIH Publication No. 04-5563.

[46] O’Donoghue, T. and M. Rabin (2006) “Optimal Sin Taxes” *Journal of Public Economics*

[47] Pope, D. (2006) "Reacting to Rankings: Evidence from America’s Best Hospitals and Colleges" UC Berkeley Working Paper.

[48] Prescott, E. (2004) "Why do Americans Work so Much More than Europeans?" *NBER Working Paper No. 10316.*

[49] Reis, R. (2006) "Inattentive Consumers." *Journal of Monetary Economics*, 53(8): 1761-1800

[50] Saez, E. (2002) "Do Taxpayers Bunch at Kink Points" *NBER Working Paper No.* 7366.

[51] Saez, E. (2003) "The Effect of Marginal Tax Rates on Income: a Panel Study of 'Brack Creep'" Journal of Public Economics 87: 1231-1258.

[52] Saez, E. (2004) "Reported Incomes and Marginal Tax Rates, 1960-2000: Evidence and Policy Implications" *NBER Working Paper No.* 10273.

[53] Sheshinski, E. (2002) "Bounded Rationality and Socially Optimal Limits on Choice in A Self-Selection Model" The Hebrew University of Jerusalem Working Paper.

[54] Shimer, R. and I. Werning (2005) "Liquidity and Insurance for the Unemployed" *NBER Working Paper No.* 11689.

[55] Shin, Jeong-Shik (1985) "Perception of Price When Information is Costly: Evidence from Residential Electricity Demand" *Review of Economics and Statistics*, 67: 591-598.

[56] Simon, H. (1955) "A Behavioral Model of Rational Choice" *Quarterly Journal of Economics* 69 (1): 99-118.

[57] Sims, C. (2003) "Rational Inattention." Princeton University mimeo.

[58] Wilson, A. (2003) "Bounded Memory and Biases in Information Processing" Harvard Univ. mimeo

[59] World Tax Database (2006), University of Michigan Business School. Available at http://www.bus.umich.edu/OTPR/otpr/introduction.htm

TABLE 1
Descriptive Statistics: Grocery Stores

	Treatment Store	Control Store #1	Control Store #2
A. Store Characteristics			
Mean Weekly Revenue ($)	307,297	268,193	375,114
Total Floor Space (sq ft)	41,609	34,187	37,251
Store Opening Year	1992	1992	1990
Number of Product Categories	111	110	112
B. City Characteristics (in 1999)			
Population	88,625	96,178	90,532
Median Age (years)	33.9	31.1	32.3
Median Household Income ($)	57,667	51,151	60,359
Mean Household Size	2.8	2.9	3.1
Percent bachelor's degree or higher	19.4	20.4	18.2
Percent Married	60.2	56.9	58.1
Percent White	72.1	56.2	65.3
Distance to Treatment Store (miles)		7.7	27.4

NOTES -- Data on store characteristics obtained from grocery chain. Weekly revenue statistics based on sales in calendar year 2005. Data for city characteristics are obtained from the U.S. Bureau of the Census, Census 2000. Control stores were chosen using a least-squares minimum-distance criterion based on this set of variables.

TABLE 2
Descriptive Statistics by Product Groups

	Treatment Store		Control Store #1		Control Store #2		Total
	Treatment Products	Control Products	Treatment Products	Control Products	Treatment Products	Control Products	All Stores and Products
A. Category Level Statistics:							
Weekly revenue per category	$84.80 (83.0)	$134.63 (169.6)	$86.60 (83.0)	$142.59 (187.5)	$104.70 (100.9)	$161.83 (224.2)	$139.24 (186.0)
Weekly quantity sold per category	21.73 (23.9)	26.35 (37.0)	23.10 (24.4)	28.15 (41.4)	26.69 (29.5)	32.15 (51.5)	28.23 (42.2)
Number of categories	15	96	15	96	15	96	111
B. Product Level Statistics							
Pre-tax product regular price	4.46 (1.8)	6.26 (4.3)	4.37 (1.6)	6.30 (4.4)	4.64 (1.8)	6.32 (4.1)	6.05 (4.1)
Pre-tax product regular price (weighted by quantity sold)	4.27 (4.7)	5.61 (3.9)	4.16 (1.6)	5.58 (3.9)	4.38 (1.7)	5.60 (3.7)	5.45 (3.7)
Weekly quantity sold per product (conditional >0)	1.47 (0.9)	1.82 (1.6)	1.58 (1.0)	1.95 (1.8)	1.63 (1.1)	2.01 (2.0)	1.88 (1.7)
weekly quantity sold > 0 (indicator for positive sales)	0.31 (0.5)	0.41 (0.5)	0.37 (0.5)	0.44 (0.5)	0.33 (0.5)	0.45 (0.5)	0.42 (0.5)
Average number of products sold per week	223 (21)	1391 (51)	205 (18)	1389 (59)	245 (20)	1548 (58)	1670 (112)

NOTES--Standard deviations in parentheses. Statistics are based on sales between 2005 week 1 and 2006 week 15. Data source is scanner data obtained from grocery chain. "Treatment products" are the set of products for which tax-inclusive prices were shown in the experimental period; "control products" are unaffected products located near the treatment products. See Appendix Table 1 for list of treatment and control categories. Product price reflects actual price paid, including any discount if product is on sale (not including the sales tax). Because scanner data includes only records of items sold in each week, we impute prices for items that were not sold using prices from preceding and subsequent weeks. In addition, we impute zero quantity and revenue for 7 percent of category-weeks during which no sales were recorded. See appendix for details on this imputation procedure.

TABLE 3

DDD Analysis of Means: Weekly Quantity by Category

	TREATMENT STORE		
Period	Control Categories	Treated Categories	Difference
Baseline (2005:1- 2006:6)	26.20 (0.22) [5568]	21.81 (0.34) [812]	-4.39 (0.41) [6380]
Experiment (2006: 8- 2006:10)	27.04 (0.86) [288]	20.69 (0.95) [42]	-6.35 (0.60) [330]
Difference over time	0.84 (0.74) [6200]	-1.12 (0.86) [854]	**DD_{TS} = -1.96** (0.64) [6710]
	CONTROL STORES		
Period	Control Categories	Treated Categories	Difference
Baseline (2005:1- 2006:6)	30.10 (0.24) [11136]	25.05 (0.28) [1624]	-5.05 (0.31) [12760]
Experiment (2006: 8- 2006:10)	30.28 (0.71) [576]	25.28 (0.98) [84]	-5.01 (1.03) [660]
Difference over time	0.18 (0.63) [11712]	0.23 (0.86) [1708]	**DD_{CS} = 0.04** (0.90) [13420]
		DDD Estimate	***-2.01*** (0.58) [20130]

a. Each cell shows mean number of units sold by category and promotional week, for various groups.

b. A promotional week is a standard calendar week, but which begins on a Wednesday instead of a Monday, and ends the following Wednesday.

c. The Experimental period spans promo week 8 in 2006 to promo week 10 in 2006. The Baseline period spans promo week 1 in 2005 to promo week 6 in 2006.

d. Sales at the two control grocery stores were combined to produce the control store group. For a classification of treatment categories, see Table 1. For a classification of control categories, see Table 2.

e. Standard errors in parentheses (clustered by week), number of observations in brackets.

TABLE 4

Effect of the Experiment on Sales: Regression Estimates

Dependent Variable:	Quantity per category	Quantity per category	Revenue per category	Log quantity per category	Log revenue per category	Price paid per product
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	**-2.00**	**-1.98**	**-12.20**	**-0.08**	**-0.11**	**-0.13**
	(0.59)* **	**(0.59)*** **	**(4.71)** **	**(0.04)** **	**(0.05)** **	**-(0.090)**
Average Price		-7.13	-10.09	-1.4	-0.37	
		(0.08)***	(0.34)***	(0.01)***	(0.01)***	
Average Price Squared		0.14	0.15			
		(0.00)***	(0.01)***			
Implied Price Elasticity		-1.34	-2.08	-1.40		
Sample size	20,252	20,252	20,252	19,002	19,002	304,860

Standard errors in all specifications are clustered on promo week.
All columns report estimates of the linear regression model as specified in text.
Quantity and revenue reflect total sales of products within a given category in a given promotional week in a given store.
Average Price is an average of the prices of the goods for sale in each category.

TABLE 5

Experiment Estimates: Restricted Samples

	Full Sample	Restricted Samples		
Dependent variable: Quantity Per Category	Treat. Categories	Treat. Categories	Treat. Store	Treat. Time
	(1)	(2)	(3)	(4)
Treatment	-2.05	-1.40	-2.19	-1.49
	(0.59)***	(0.31)***	(0.62)***	(0.63)
Before Treatment	0.04			
	(0.96)			
After Treatment	0.15			
	(0.71)			
N	2,1580	2,684	6,771	996

Dependent variable in all specifications is quantity sold per category per week.Standard errors in all specifications are clustered on promo week. Specification 1 includes "placebo" treatment variables (and their interactions) for the 3 week period before the experiment and the 3 week period after the experiment. Specifications 2-4 report DD estimates. Specification 2 restricts the sample to treatment categories only. The "Treatment" variable is defined as the interaction between the treatment store dummy and treatment time dummy. Specification 3 restricts to the sample to treatment store only. The "Treatment" variable is defined as the interaction between the treatment category dummy and the treatment time dummy. Specification 4 restricts to the sample to treatment period only. The "Treatment" variable is defined as the interaction between the treatment category dummy and the treatment store dummy.

Table 6

Summary Statistics for Alcohol Excise Taxes, Sales Taxes and Alcohol Consumption

Beer Excise Tax (Cents/Case)	0.50
	0.46
Beer Excise Tax (Percent)	5.5
	5.4
General Sales Tax (Percent)	4.2
	1.8
State Beer Consumption (Gallons)	106916
	118759
State Per-Capita Beer Consumption (Cans/Pop.)	243.1
	46.1
N	1734

Means; standard deviations in italics.
Source: Brewer's Almanac 2005; National Institute on Alcohol Abuse and Alcoholism;

Table 7

Effect of Excise and Sales Taxes on Beer Consumption

Dependent Variable:	Change in Log(per capita beer consumption)			
	1-year differences			3-year differences
ΔLog(1+Excise Tax Rate)	**-1.17** (0.25)***	**-1.15** (0.25)***	**-1.15** (0.25)***	**-1.17** (0.41)***
ΔLog(1+Sales Tax Rate)	**-0.36** (0.30)	**-0.36** (0.30)	**-0.12** (0.30)	**0.03** (0.31)
ΔLog(Population)		0.10 (0.07)	-0.16 (0.09)*	
ΔLog(Income per Capita)			0.22 (0.05)***	
ΔLog(Unemployment Rate)			-0.01 (0.01)	
Year Dummies	Yes	Yes	Yes	Yes
F-Test for Equality of Tax Variables (Prob>F)	0.05	0.05	0.01	0.02
Observations	984	984	937	924

Notes: Dependent variable is the first-differenced log of per capita beer consumption (from National Institute on Alcohol Abuse and Alcoholism). Source of tax data is Brewer's Almanac (2005) and the Univ. of Michigan World Tax Database.

Standard errors in parentheses * significant at 10%; ** significant at 5%; *** significant at 1%

TABLE 8

Calibration: Welfare Cost of Ignoring Taxes

Implied Welfare Loss from Failure to Optimize ($G):

	x_0 =1000		x_0 =10,000	
	$\varepsilon_{x,p}$ =0.5	$\varepsilon_{x,p}$ =1	$\varepsilon_{x,p}$ =0.5	$\varepsilon_{x,p}$ =1
T				
0.05	0.63	1.25	6.25	12.50
0.1	2.50	5	25	50
0.2	10	20	100	200
0.3	22.50	45	225	450
0.4	40	80	400	800

The calibrations assume that t_p=0.

TABLE 9

Calibration: Individual Welfare and Social Welfare

Individual vs. Social Welfare Cost of Ignoring a Tax

Δt	*Welf loss to indiv.* ΔU	ΔR	ΔR_s	ΔW_s
0.01	0.004	1	0.792	0.204
0.05	0.1	5	3.8	1.1
0.1	0.4	10	7.2	2.4
0.2	1.6	20	12.8	5.6
0.3	3.6	30	16.8	9.6

The calibrations assume that x_0=100, $\varepsilon_{x,p}$=1 and t_0=0.25

TABLE 10

Average vs. Marginal Tax Rate Calibration Results

	Alpha (h share) = 0.5		Alpha (h share) = 0.7	
Actual Tax Rate	WTP for Avg Rate	WTP for Marg Rate	WTP for Avg Rate	WTP for Marg Rate
0.05	52.7	25	112.5	26.5
0.1	222.9	100	497.9	105.8
0.15	533.6	225	1262.9	238.1
0.2	1016.1	400	2601.4	423.3
0.25	1715.7	625	4941.5	661.4
0.3	2701.8	900	9708.5	952.4
0.35	4091.1	1225	inf	1296.4
0.4	6111.5	1600	inf	1693.2

The calibrations assume that t_p=0.
WTP is willingness to pay.
Cobb-Douglas utility with housing share alpha.
Gross unearned income y = 20000, gross labor income chosen at t_p = 0 is 20000.
Table lists dollar values of welfare cost from failing to optimize relative to true average and marginal rates as

Figure 1a
Placebo Estimates: Sets of Control Products

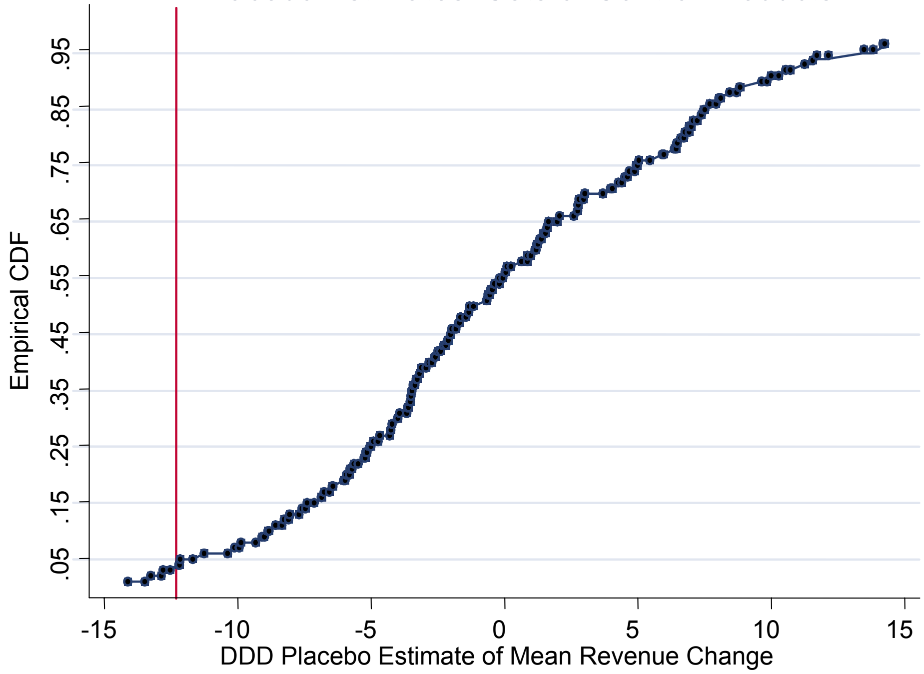


Figure 1b
Placebo Estimates: Other Time Periods

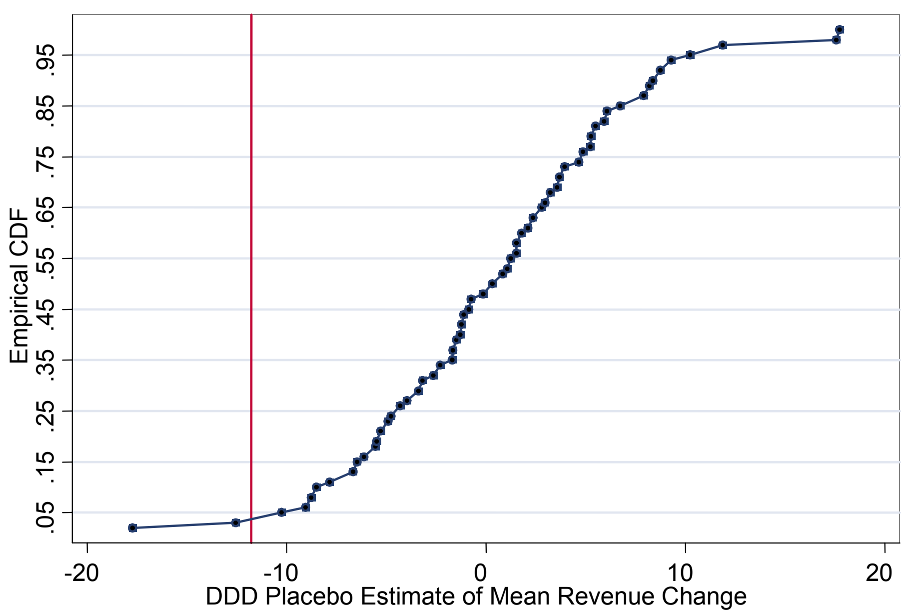


Figure 2a

Effect of Excise Tax Increases on Beer Consumption

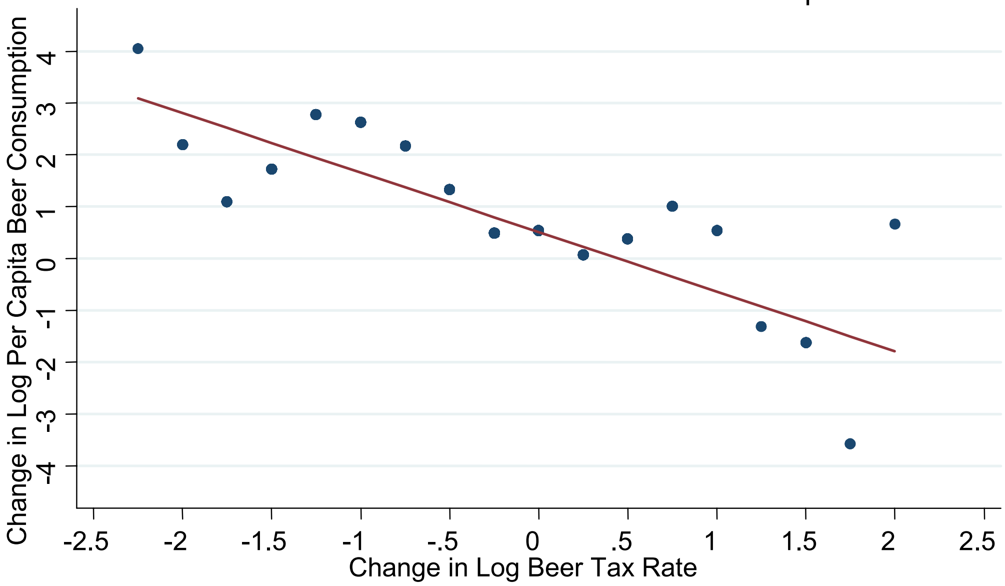


Figure 2b

Effect of Sales Tax Increases on Beer Consumption

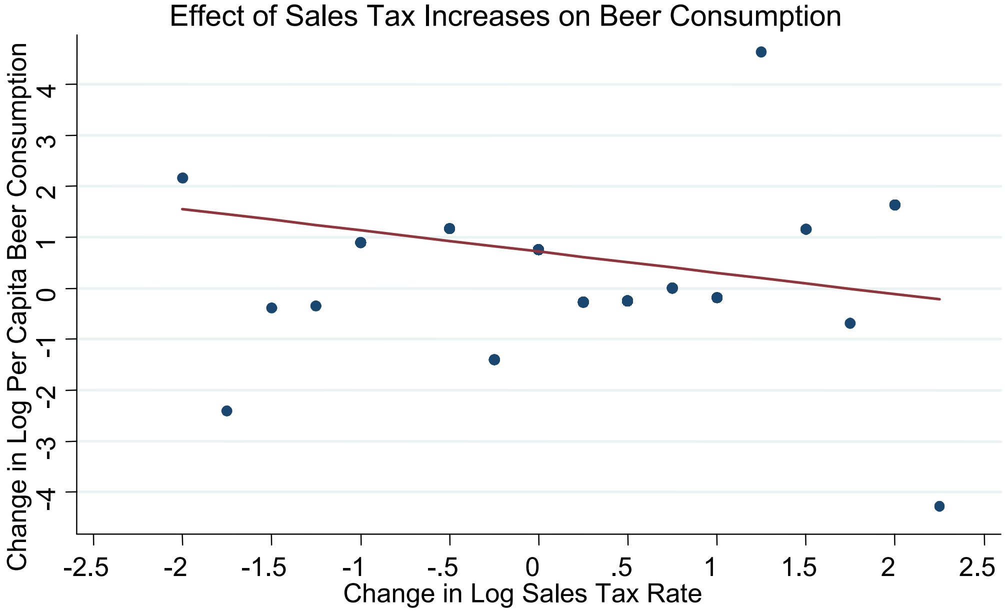


Figure 3a
Number of Correctly Reported Taxed Items on Survey

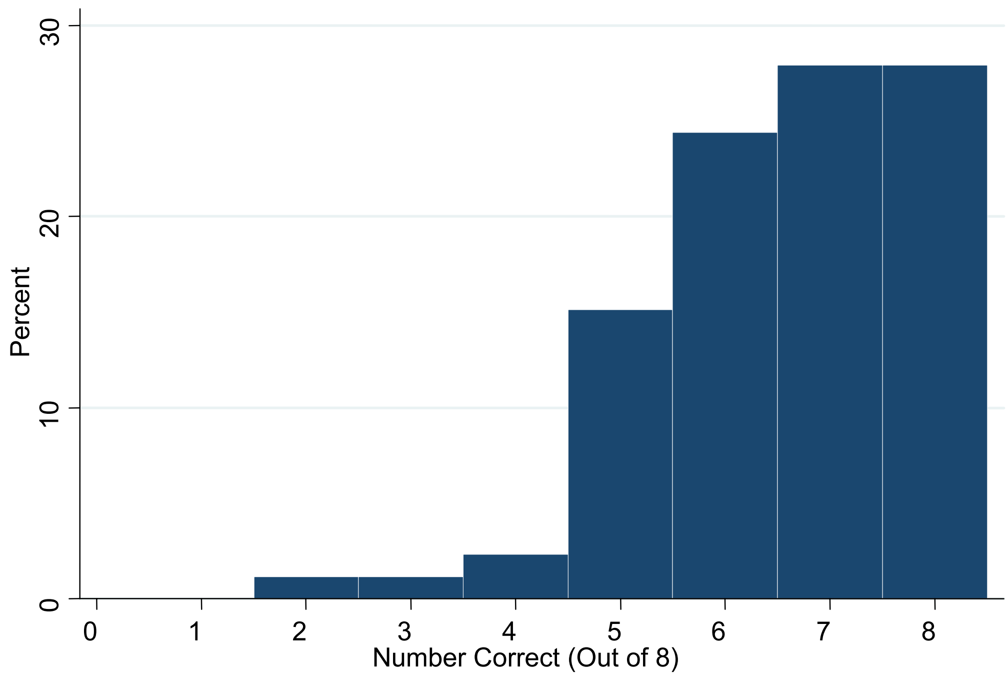
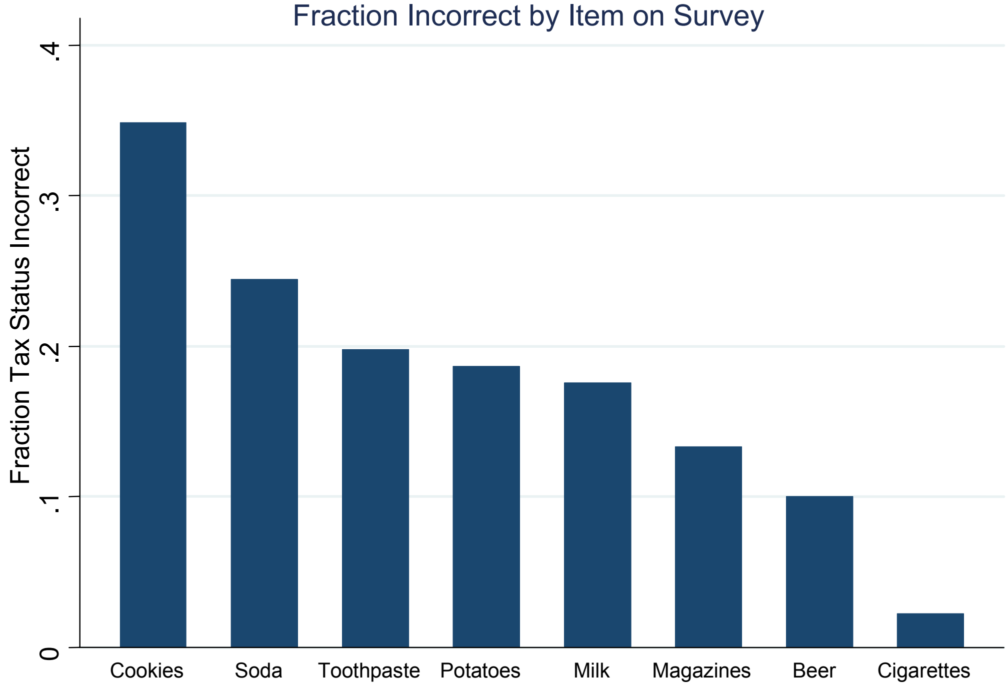


Figure 3b:
Fraction Incorrect by Item on Survey



Bounded Rationality and Efficiency Costs of Taxation

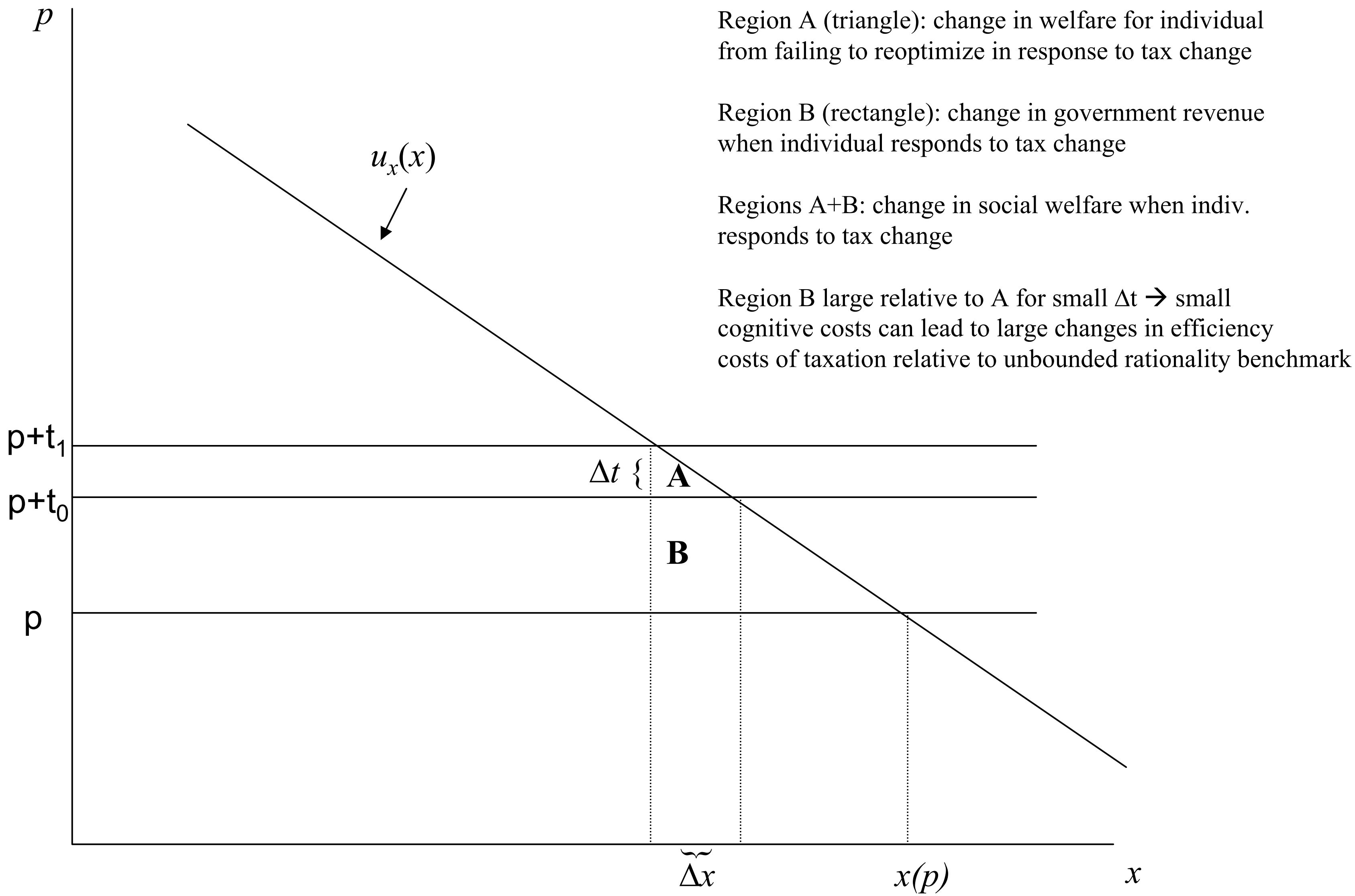


EXHIBIT 1: TAX-INCLUSIVE PRICE TAGS

EXHIBIT 2: TAX SURVEY

University of California, Berkeley
Department of Economics

This survey is part of a project about taxes being conducted by researchers at UC Berkeley. Your identity will be kept strictly confidential and will not be used in the research. If you have any questions about your rights or treatment as a participant in this research project, please contact UC-Berkeley's Committee for Protection of Human Subjects at (510) 642-7461, or e-mail: subjects@berkeley.edu.

Gender:	**Age**:	**Marital Status**:	**Education**:	**Years You Have Lived in California:**
☐ Male ☐ Female		☐ Married ☐ Unmarried	☐ High School ☐ College Degree ☐ Graduate Degree	

Is tax added **at the register** (in addition to the price posted on the shelf) for each of the following items?				Have you purchased these items within the last month?			
milk	Y N	toothpaste	Y N	milk	Y N	toothpaste	Y N
magazines	Y N	soda	Y N	magazines	Y N	soda	Y N
beer	Y N	cookies	Y N	beer	Y N	cookies	Y N
potatoes	Y N	cigarettes	Y N	potatoes	Y N	cigarettes	Y N

What is the sales tax rate in Vacaville? ____________%

What is the California **state** income tax rate in the highest tax bracket? ______________ %

What percentage of families in the US do you think pay the federal estate tax when someone dies?

< 2%   2-10%   10-25%   25-50%   > 50%

Thank you for your time!