Salience and Taxation

Raj Chetty
UC-Berkeley

Adam Looney
Federal Reserve Board

Kory Kroft
UC-Berkeley

April 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 grocery store where we posted tax-inclusive prices for 750 products subject to sales tax for a three week period. We find that posting tax-inclusive prices reduced demand by roughly 7% among the treated products relative to control products and nearby control stores. Second, 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). Both sets of results indicate that tax salience affects behavioral responses. We propose a simple bounded rationality model to explain why salience matters, and show that it matches our evidence as well as several additional stylized facts. In the model, agents incur second-order (small) utility losses from ignoring some taxes, even though these taxes have first-order (large) effects on social welfare and revenue. Using this framework, we derive formulas for the efficiency cost and incidence of commodity taxes when agents do not optimize fully.

*Preliminary draft. We are very grateful to Sofia Berto Villas-Boas, Reed Johnson, and Steve Flaming for help in implementing the experiment. Thanks to George Akerlof, David Ahn, Alan Auerbach, Kitt Carpenter, Stefano Della Vigna, Amy Finkelstein, Michael Greenstone, Jon Gruber, Shachar Kariv, Botond Koszegi, James Poterba, Matthew Rabin, Emmanuel Saez, Jesse Shapiro, Andrei Shleifer, 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. Authors' e-mails: chetty@econ.berkeley.edu, adam.looney@frb.gov, kroft@econ.berkeley.edu.

1 Introduction

A central assumption in public finance is that agents optimize fully with respect to 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 respond to price changes. Models of optimal income taxation assume that agents choose labor supply and consumption optimally irrespective of the complexity of the tax schedule they face (e.g. Mirrlees 1971, Atkinson and Stiglitz 1976). Similar assumptions are implicit in positive analyses of taxation and empirical studies of behavioral responses to taxation. In practice, 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 and are often not directly displayed in posted prices (sales taxes, hotel city taxes, vehicle excise fees). Classic results on tax incidence and efficiency costs (e.g. Harberger 1964) rely on full optimization against such tax policies.

In this paper, we investigate whether individuals optimize fully with respect to taxes by analyzing the effect of "salience" on behavioral responses to commodity taxation. We define the "salience" of a tax in terms of the simplicity of calculating the gross-of-tax price of a good.[1] In the empirical component of the paper, we test whether a commodity tax has a larger effect on demand if it is included in the posted price that customers see when shopping (and hence is more salient). We focus on this particular question because it is a fairly stringent test of the full-optimization benchmark: if individuals optimize imperfectly even with respect to linear commodity taxes, similar issues may arise in the analysis of a broad set of policies. We use two complementary empirical strategies: (1) an experiment in a grocery store and (2) an observational study of the effect of alcohol taxes on alcohol consumption.

The experiment was implemented in collaboration with a major grocery chain at a large store over a three-week period in early 2006. In this store, prices posted on the shelf exclude sales tax of 7.375%. If the good is subject to sales tax, it is added to the bill only at the

[1] To be precise, we say that a tax policy t_a is more "salient" than a tax policy t_b if calculating the gross-of-tax-t_a price of a good requires less computation than calculating the gross-of-tax-t_b price.

register, as in most other retail outlets in the United States.[2] Our intervention was to post tags showing the tax-inclusive price below the original pre-tax price tag for all products in three taxable groups (cosmetics, hair care accessories, and deodorants), thereby increasing the salience of the sales tax. We analyze the effect of this intervention using a quasi-experimental differences-in-differences research design. Using scanner data, we find that quantity sold and total revenue in the treated group of products fell by about 7% during the intervention relative to two "control groups" – other products in the same aisle of the treatment store that were not tagged and two stores in the same chain in nearby cities. The null hypothesis that posting tax-inclusive prices has no effect on demand is rejected with $p < 0.05$ using both t-tests and non-parametric permutation tests. To interpret the magnitude of this effect, we compare the estimate with the price elasticity of demand for the these categories, which is in the range of 1 to 1.5. Hence, showing the tax-inclusive price reduced demand by nearly the same amount as a 7.375% price increase. This finding suggests that the vast majority of customers normally do not take the sales tax on these products into account.

A concern with the experiment is that posting 750 new tags may have reduced demand because of a "Hawthorne effect" or a short-run violation of familiar norms. This issue motivates our second empirical strategy, which compares the effect of price changes with tax changes using observational data over a longer horizon. To implement this test, we focus on alcohol consumption, because alcohol 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 changes in these two tax rates between 1970 and 2003 coupled with annual data on aggregate alcohol consumption by state, we find that increases in the excise tax reduce alcohol consumption an order of magnitude more than similar increases in the sales tax. This difference in elasticities persists over relatively long horizons (e.g. 2 or 3 years). A simple calibration suggests that the magnitude of the difference in the elasticity estimates is unlikely to be explained by the fact that the sales tax

[2] The sales tax affects relative prices because it does not apply to all goods. Approximately 40% of expenditure is subject to sales tax in the United States. Since food is typically exempt, the fraction of items subject to sales tax in grocery stores is much lower.

applies to a broader base, especially since food and non-alcoholic beverages are exempt from sales tax in most states.

Both strands of evidence indicate that behavioral responses to 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 which goods are subject to sales tax. An alternative hypothesis is that salience matters: the customers know what is taxed, but choose to focus on the posted price because computing the tax-inclusive price for each good entails 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.

This empirical result motivates the second portion of the paper, which focuses on developing a theoretical model that can match the evidence on the importance of salience while providing a tractable framework for analyzing issues such as the welfare consequences of taxation. We propose a simple bounded rationality model in which agents face a small cognitive cost of computing tax-inclusive prices (as in Simon 1955, Akerlof and Yellen 1985) to explain why salience matters. In particular, we show that second-order (small) cognitive costs can lead agents to ignore a first-order (large) range of taxes, and focus instead on the salient pre-tax price. For instance, when utility is quasilinear, the cost of ignoring a 10 percent tax on an item on which the agent spends $10,000 is only $50. 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). Viewed from this perspective, it is perhaps unsurprising that individuals with limited time or attention choose not to compute tax-inclusive prices for small goods such as cosmetics and alcohol.

In addition to matching our empirical evidence, the model also makes predictions about the circumstances in which individuals are more likely to pay attention to taxes. The fraction of agents who compute tax-inclusive prices is endogenously determined by the tax rate and other factors such as the price elasticity of demand. The heuristic behaviors predicted by the model help explain some stylized facts in the literature that pose problems for existing models.

An attractive feature of the model for public finance is that it offers a simple framework for welfare analysis when agents do not optimize perfectly. Bounded-rationality and salience can be important in the analysis of many large-scale tax policies from a social perspective. Even though individuals incur second-order utility losses from ignoring certain taxes, these taxes can nevertheless have first-order (large) impacts on social welfare and revenue. A 10% tax increase raises 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.

To quantify the effects of taxes on welfare, we derive empirically implementable Harberger-type formulas for the incidence and efficiency costs of taxation. The deadweight loss of taxation is determined by two additional factors beyond the standard compensated elasticity of demand when agents are boundedly rational: (1) the magnitude of distortionary income effects (allocation errors) that arise because agents do not optimize relative to true tax-inclusive prices and (2) the effect of the tax rate on the fraction of individuals who pay attention to taxes. The incidence of taxes on consumers depends on the fraction of agents who compute tax inclusive prices and the "fundamental" price elasticity of demand rather than the tax elasticity. Because of these factors, the efficiency cost and incidence of tax policies can differ substantially from predictions based on estimates of compensated price elasticities using existing formulas. For example, tax increases can have a substantial efficiency cost even when individual behavior does not change, by distorting consumption allocations for inattentive individuals. Another implication is that incidence will in general depend on whether the tax is levied on consumers or firms, violating the classic tax neutrality result in competitive markets.

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 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 empirical 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.

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 (2007), 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

[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.

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 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.

Suppose the economy has 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 maximizes $U(x,y)$ and chooses

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

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

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

Let θ denote the fraction of agents who optimize relative to the true tax-inclusive price. Then aggregate demand for x in an economy with a unit mass of agents is given by

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

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

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

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

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

where $\alpha = -\frac{1}{b} \log a$ and $\beta = -\frac{1}{b}$. 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 for certain goods in the U.S.[4] 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

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

Defining the price elasticity of demand as $\varepsilon_{D,p} = -\frac{\partial \log x}{\partial \log p} = \beta$, it follows that

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

where $\rho = -\frac{\log x(p,t,1) - \log x(p,t,\theta)}{\log(1+t)}$ denotes the normalized "tax visibility" effect. The parameter

[4]In practice, there could be other "types" in the population who use different heuristics, e.g. adding 10% to all posted prices to account for taxes. In this case, our estimate of θ cannot be interpreted as the fraction of full optimizers in the population. Nevertheless, our hypothesis tests remain qualitatively valid: an estimated $\theta < 1$ constitutes a rejection of a model where all individuals optimize fully.

ρ can be interpreted as the (absolute value of) 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. 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 ($\rho = 0$), 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,

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

Under the null hypothesis of full optimization, prices and taxes – which differ in their salience – should affect demand in the same way: $\varepsilon_{x,1+t^S} = \varepsilon_{x,p} \Leftrightarrow \theta = 1$.

In the next section, we implement strategy 1 using a field experiment at a grocery store. In section 5, 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 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.[5] 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.[6]

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. 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 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 products sold 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),

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

[6] 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 direct substitute for a treated product would also be treated.

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 beginning on 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.[7]

The store changes product prices on Wednesday nights and leaves the prices fixed (with rare exceptions) for the following week, 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 750 products. 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

[7] An important concern with this experiment is that the tags themselves may have created confusion, thereby reducing demand for reasons unrelated to the information that was provided. While we cannot rule out such a "Hawthorne effect," we show below that most individuals know the parameters of the sales tax almost exactly, a finding which suggests that most customers were likely to have understood the tags.

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 and product level summary statistics, broken down by treatment and control product groups within each store. The treatment product group consists of all products in cosmetics, hair care accessories, and deodorants . Within these 3 broad groups, there are 13 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. The 95 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 upper panel of Table 2 reports category-level statistics in the treatment group and control group categories for each of the three stores. Average weekly revenue per category from the treatment products is approximately $100, while the average quantity sold per category is approximately 25 units. The treatment products as a whole account for approximately $1,300 of revenue per week. Average revenue per category in the control group is higher, partly because the average price of the control category products is higher and partly because the volume of sales in those categories is somewhat higher. The average revenue per category and the number of items purchased each week is roughly similar within categories across stores. The lower panel of Table 2 reports some product-level statistics. The average price of products that sold in the treatment group is $4.27. Among products that sold in a given week, the average number of units sold is 1.47.

For most of our analysis, we analyze the data at the category-by-store level (so that there are 13+95=108 observations per store per week), summing quantity sold and revenue over the individual products within categories in each store. We use the category-level data because of a missing data problem at the product level. In particular, we cannot distinguish products that were on the shelf but did not sell from products that were not stocked in a given week. The scanner data includes only transaction information, and we do not have data on the set of products that were on the shelf in each week. If we impute quantity as zero for items that sold in both an earlier week and subsequent week, we find that only 31% of products sell in a given week (see the Appendix for details on the interpolation procedure). By analyzing the data at the category level, we largely circumvent this problem because there are relatively few category-weeks with missing data (4.7% 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 using the interpolation procedure. 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 cross-tabluation of mean quantity sold in Table 3. 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 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 adjust for serial correlation of errors across products.

The mean quantity sold in the treatment categories fell by an average of 1.30 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 2.14 units on average, with a standard error of 0.68. This change of $DD_{TS} = -2.14$ 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 treatment products increased by a (statistically insignificant) $DD_{CS} = 0.06$ units relative to sales of control products. 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 $DDD = DD_{TS} - DD_{CS} = -2.20$. This estimate is statistically significant with $p < 0.01$, rejecting full-optimization ($\theta = 1$). 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 only of the treatment products in the treatment store. 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 29.01 units. The estimate of -2.20 therefore implies that quantity sold fell by 7.58 percent. Given the sales tax rate of 7.375 percent, the normalized tax visibility effect is approximately $\rho = 1$. As we discuss below, estimates of the price elasticity of demand at the category level (i.e., the effect of a 1% increase in the prices of *all* goods within a category) range from roughly $\varepsilon_{d,p} = 1$ to 1.5. Since $1 - \theta = \frac{\rho}{\varepsilon_{D,p}}$, we infer that $\theta < \frac{1}{3}$. Note that we cannot reject the hypothesis that $\theta = 0$ given the standard error on the estimate of θ. Hence, the data are consistent with the hypothesis that *none* of the customers normally base their decisions on the tax-inclusive price in these product groups.

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 the observation is in the experimental store, categories, and time, respectively. Let X denote a vector of additional covariates. We estimate variants of the following linear model, which generalizes the DDD strategy used in Table 3:

$$\begin{aligned} y_{sct} &= \alpha + \beta_1 treattime + \beta_2 treatstore + \beta_3 treatcat + \gamma_1 treattime \times treatcat \\ &\quad + \gamma_2 treattime \times treatstore + \gamma_3 treatstore \times treatcat \\ &\quad + \delta treattime \times treatcategory \times treatstore + \rho X + \varepsilon_{sct} \end{aligned} \tag{3}$$

In this specification, the β 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 equals 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 (3) for quantity sold without any additional controls.[8] 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

[8] 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 (3).

3, we examine the effect of the intervention on weekly revenue (price×quantity) per category. Consistent with the evidence from the quantity analysis, the experiment led to a significant reduction in revenue from the treatment products relative to the control groups.

In specifications 4 and 5, we estimate analogous models in logs instead of levels. In these specifications, we weight each observation y_{sct} by $\overline{y}_{sc}$, the mean revenue by store by category, placing greater weight on the larger categories as in the levels regressions. An advantage of the logs specification is that it may be a better model for comparisons across categories with different baseline quantities, given that shifts may be equi-proportional. A disadvantage of the logs specification is that it forces us to omit observations that have zero quantity sold. Despite these differences, the logs specifications imply an estimated reduction consistent with the levels models: a decline in quantity sold of 8.5% and revenue of 10.8%.

Our regression specifications yield estimates of the category-level price elasticity of demand $\varepsilon_{d,p}$ by exploiting the variation in average category-level prices across weeks and categories within the grocery stores. In the levels model, the estimated price elasticity is $\varepsilon_{d,p} = 1.39$ at the sample mean price of $5.45. The log-linear model yields a similar estimate. These estimates are broadly consistent with those obtained by Hoch et. al. (1995), who estimate a full product-level demand system and obtain category-level price elasticities of 1 to 1.5 using scanner data from the same grocery chain.[9]

Both the levels and logs specifications suggest that revenue per category fell more than quantity sold per category.[10] 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 sold in each category). While imprecisely estimated, the coefficient estimate implies that the average price of items purchased fell by about $0.13 (3 percent) during the treatment period, consistent with the gap between the revenue and quantity estimates in the earlier regressions. One interpretation of this result is that individuals in the market for a more expensive product were more likely to buy nothing at

[9] This similarity of estimates is reassuring because the informal approach of aggregating over the categories and regressing mean quantity on mean price need not in general produce a consistent estimate of the category-level price elasticity.

[10] To see this for the levels specifications, note that the average price of the products in the dataset (weighted by quantity sold) is $5.45. If quantity sold of all products within each category fell equally, one would expect a revenue loss of only $12 per category based on the estimated quantity reduction of 2.2 units.

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.[11]

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 (3). 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 with the results in Table 4, quantity sold in the treatment group is estimated to have fallen by approximately $\delta =$ 2.2 units during the intervention. The corresponding "placebo" estimates for the periods before and after the treatment are close to zero.[12] These results indicate that the fall in demand coincides precisely with the intervention period, supporting the identification strategy.

Building on the logic of this specification check, 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:

$$\begin{aligned} y_{sct} &= \alpha + \beta_1^t time_t + \beta_2^t treatstore + \beta_3^t treatcat \\ &\quad + \gamma_1^t time_t \times treatcat + \gamma_2^t time_t \times treatstore + \gamma_3^t treatstore \times treatcat \qquad (4) \\ &\quad + \delta_t time_t \times treatcategory \times treatstore + \varepsilon_{sct} \end{aligned}$$

[11] 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.

[12] We discuss why demand returns to pre-experiment levels after the tags were removed in section 6.

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 statistic $G(\delta)$ represents a p-value for the hypothesis that $\delta = 0$, based on a non-parametric permutation test over the weeks in the sample. Intuitively, 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.[13] 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 $\delta = -\$13.1$. Since $G(\delta) < 0.05$, 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 13 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 13 control categories for each week in the sample (thereby obtaining roughly 6,000 $\widehat{\delta}_t$ 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

[13] 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.

2 and 3 of Table 5, we report difference-in-difference estimates, exploiting pairs of these counterfactuals separately. In column 2, we restrict the sample to the treatment product categories, and compare across time and stores. In column 3, we restrict the sample to the treatment store, and compare across time and categories. Reassuringly, both DD estimates are similar to the DDD estimates 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 (3) 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 (3) 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 most states: (1) an *excise* tax that is levied at the wholesale level and thus is included in the price posted on the shelf (or on a restaurant menu) and (2) a *sales* tax, which applies to alcohol but is added at the register (except in Hawaii, which we exclude). Hence, the excise tax (t^E) is more salient than the sales tax (t^S).

Our research design takes state-level changes in the sales and excise tax rates as exogenous, and examines the effects of these changes 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).

$$\log x(t^E, t^S, \theta) = \alpha + \beta \log(1 + t^E) + \theta\beta \log(1 + t^S) \tag{5}$$

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 X and an error term ε_{jt} to capture idiosyncratic state-specific demand shocks, we obtain the following estimating equation by first-differencing (5):

$$\Delta \log x_{jt} = \alpha_0 + \beta\Delta \log(1 + t^E_{jt}) + \beta\theta\Delta \log(1 + t^S_{jt}) + X_{jt}\rho + \varepsilon_{jt} \tag{6}$$

We estimate (6) 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 made in deriving (5) 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 percent of consumption is subject to sales taxation on average.[14] 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 some additional evidence and calibrations which suggest that the degree of bias from this issue is unlikely to explain the the estimated difference between the two elasticities.

5.2 Data and Summary Statistics

For simplicity, we focus 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 in a sample of the population.

We obtain data on state excise tax rates on beer from the Brewer's Almanac (various years), the Tax Foundation's State Tax Collections and Rates (various years), and the State Tax Handbook. The excise rate includes local excise taxes that are applied state-wide, any taxes levied at the wholesale level, and the federal excise tax. State sales taxes are obtained from the World Tax Database (2006) at the University of Michigan and are checked for accuracy against the Census Bureau's Annual Survey of State and Local Government Finances. We supplement this data on state-level sales taxes with data on average local

[14] In 2004, sales tax revenues were about 2.1 percent of personal consumption expenditures (PCE) while the average income weighted state tax rate was 5.3 percent. Hence the tax base is approximately 40 percent of PCE.

sales tax rates, imputed from data on local revenues from the Census Bureau's Survey of State and Local Government Finances and a tax base defined to be state revenues divided by the state rate. In the four states that apply a different sales tax rate to alcohol than to other products, we use the alcohol sales tax rate.

Since our estimation strategy relies on the timing and magnitude of the tax changes, we first evaluate the quality of the data by regressing the change in the log of state tax revenues on the change in the log of the sales tax rate, controlling for state income. The coefficient estimate on the sales tax rate is 0.76 (s.e. 0.03). Given that states sometimes change the tax base while changing rates, this estimate suggests that our sales tax rate variable is reasonably precise. A state-by-state analysis of changes in rates and changes in revenues also yields similarly high correlations, with the exception of West Virginia, where the data on tax rates and revenues are negatively correlated. In view of this apparent measurement problem, we exclude West Virginia from our analysis, though including it does not affect our conclusions.

The state sales tax is an ad valorem tax (proportional to price), while the excise tax is typically a specific tax (specified as dollars per unit of beer). We convert the excise tax rate into percentage units by dividing the beer excise tax per case by the average cost of a case of beer in the United States in the corresponding year, as measured by the Producer Price Index for Malt Beverages.[15]

Finally, we use data on two sets of covariates to mitigate concerns about the endogeneity of tax reforms. First, one may be concerned that the business cycle affects tax revenues and therefore tax rates given that states must balance their budgets. To separate the causal effect of tax changes from contemporaneous changes in economic conditions, we include flexible controls for the state unemployment rate and state per capita income (from BLS and BEA). Second, excise tax increases are sometimes associated with other alcohol regulations, particularly efforts to reduce drunk driving or underage drinking. To separate the causal effect of tax changes from contemporaneous changes in regulations, we obtained state-by-year

[15] We normalize by the average price in the nation because the price in each state is endogenous to its own tax rates. Since Alaska has a higher price level than the continental United States, we follow Census Bureau practice and adjust its price level up by 25 percent when calculating the percentage excise tax rate. None of our results are affected by this adjustment, or by excluding Alaska entirely.

measures of the legal drinking age, the legal blood alcohol content (BAC) limit, implementation of stricter drunk driving regulations for youths, and introduction of administrative license revocation laws.[16]

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 $0.49 per case, while the Federal excise tax averaged $.90 over the time period. The average state excise tax is 5.6 percent of the average price. The mean sales tax applied to alcohol is 4.3 percent. Changes in alcohol regulations are relatively infrequent in our sample period: states changed their alcohol control policies in about 12 percent of years.

It is important to note that 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. Among states in the so-called "Bible Belt" (Alabama, Georgia, Louisiana, Mississippi, North Carolina, and South Carolina), the combined (federal+state) excise tax rate has exceeded 30 percent in several years. The nominal value of the excise tax is updated infrequently, so excise tax rates have fallen as a percentage of price over time. In contrast, sales tax rates have increased secularly over time. Since identification from these secular changes may be contaminated by aggregate time trends in alcohol consumption, we include year fixed effects in all specifications, effectively identifying the model from differential changes in tax rates within years across states.

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

[16]Thanks to Christopher Carpenter for sharing this data.

in alcohol tax rates to $\pm.02$ log points. Without this restriction, results are similar and the effect of the excise tax on beer consumption is more precisely estimated.

To construct Figure 2a, we first remove year effects by computing residuals of $\Delta \log(x)$ and $\Delta \log(1+t^E)$ in regressions on year dummies. These residuals correspond to the change in log consumption and the log tax rate in a given state relative to the average changes that year for the entire country. We then divide the interval from $[-0.02, 0.02]$ into equal-sized bins, and compute means of the residual change in log beer consumption and the residual of $\Delta \log(1+t^E)$ within each bin. Finally, we plot the means of the residual changes in beer consumption against the means of $\Delta \log(1+t^E)$, superimposing a best-fit line as a visual aid.[17]

To quantify the magnitude of the difference in the excise and sales tax elasticities, Table 7 presents estimates of the model in (6) from a set of alternative samples and specifications. In this and all subsequent tables, we adjust for potential serial correlation in errors by clustering the standard errors by state. Column 1 reports estimates of a baseline model that includes only year fixed effects and log state population as covariates. In this specification, a 1 percent increase in the gross-of-excise-tax price is estimated to reduce beer consumption by 1.06 percent (i.e. $\varepsilon_{x,1+t^E} = 1.06$).[18] In contrast, a 1 percent increase in the gross-of-sales-tax price is estimated to reduce beer consumption by 0.20 percent (i.e. $\varepsilon_{x,1+t^S} = 0.20$). The null hypothesis that the excise and sales tax elasticities are equal is rejected with $p = 0.02$.

Columns 2-4 evaluate the robustness of these estimates to controlling for factors that are likely to be correlated with the tax changes. In column 2, we control for changes in alcohol regulations. In particular, we include an indicator variable for a shift toward stricter regulations in any of our four measures (legal drinking age, drunk driving regulations for youths, changes in the legal blood alcohol content limit, and administrative license revocation laws). Controlling for changes in alcohol regulations does not affect the coefficient estimate on the excise tax rate significantly. This is because these regulation changes in are estimated to have modest effects on alcohol consumption: alcohol consumption falls by 0.5%

[17] The best-fit line is weighted to account for the fact that some bins contain more observations than others.

[18] This elasticity estimate should not be confused with the elasticity of beer consumption with respect to the excise tax rate (ε_{x,t^E}) that is often reported in empirical studies of beer demand. The latter elasticity is much smaller because of the difference in units ($\log(1+t^E) \approx t^E$ vs. $\log(t^E)$).

on average when one of these regulations are tightened. In other specifications (not shown), we introduce each of the alcohol regulation variables separately, and find similar results.

In column 3, we control for the state-level business cycle by including state per capita income and the state unemployment rate as covariates. Introducing these controls reduces the estimated sales tax coefficient, a finding that is consistent with the fact that sales taxes are sometimes raised during budgetary shortfalls that occur in recessions. Since alcohol is a normal good (as indicated by the positive coefficient on per capita income and negative coefficient on unemployment rate), failing to control for the business cycle biases the correlation between alcohol consumption and sales tax changes upward in magnitude. Hence, the endogeneity of sales tax rates appears to, if anything, work *against* rejecting the null hypothesis that $\varepsilon_{x,1+t^E} = \varepsilon_{x,1+t^S}$.

Finally, one may be concerned that budgets and tax policies adjust to revenue shortfalls with a lag. We evaluate this concern by estimating specifications with lags and leads of economic indicators. Estimates from one representative specification, which includes the lagged unemployment rate and lagged per capita income, are reported in column 4. The coefficient on the excise tax rate falls to -0.81, but remains significantly different from the sales tax coefficient.

Robustness Checks. In Table 7b, we assess the robustness of the results to additional changes in specification. First, note that the sales tax variable used in the previous table excluded changes in local taxes. If localities lower taxes to offset increases in state rates, changes in the state sales tax may overstate the true change in the combined tax rate. Column 1 in table 7b reports results incorporating changes in local sales taxes, imputed from data on local tax revenues. In this specification, the excise tax coefficient is unchanged, while the estimate of the sales tax coefficient is positive and statistically insignificant. The hypothesis that the two coefficients are equal is rejected at conventional significance levels.

One concern in our identification strategy for the excise tax effect is that trends in tax rates may be correlated with changes in cultural norms, which directly influence alcohol consumption. For example, rising acceptance of alcohol consumption in historically conservative states (the "Bible Belt states") may have led to both a reduction in the excise tax as a percentage of price and an increase in alcohol consumption. To assess whether this channel leads to significant bias, we control for region-specific trends in column 2 of Table 7b by including region dummies. The coefficient on the excise rate falls modestly, but remains statistically significantly different from the coefficient on the sales tax, suggesting that our results are not due to long-term region-specific trends.

As an alternative approach to disentangling the effect of trends, we isolate the effect of explicit changes in legislated excise tax rates. There are two sources of variation identifying the excise tax coefficient. The first is policy changes in the nominal tax rate. The second is the erosion of the nominal value of the tax by inflation, which creates differential changes in excise tax rates across states because they have different initial tax rates.[19] We believe that both sources of variation constitute plausibly exogenous changes in tax rates. To test whether the two sources of variation yield similar results, we isolate the effect of the policy changes using an instrumental variables strategy. We fix the price of a case of beer at its sample average and compute the implied ad valorem tax as the legislated nominal excise tax divided by this *time-invariant* price. The only variation in this simulated tax rate is due to changes in law. Using this simulated tax rate to instrument for the actual rate, we replicate the baseline specification. The results are presented in column 3. The point estimates of both tax elasticities are similar to those in previous specifications, but standard errors rise as expected since part of the variation in excise tax rates has been excluded.

Thus far, our analysis has focused on changes in tax rates and alcohol consumption at an annual frequency. One potential explanation of the difference between the sales and excise tax effects is learning: people might immediately perceive excise taxes, but learn about changes in the sales tax over the next few years. In this case, the actual effects of the two taxes on alcohol demand might be similar in the long-run steady state despite the preceding results. To test for such learning effects and estimate longer-run elasticities, we estimated specifications including lags and leads of the tax variables and differences over longer horizons (e.g. two or three year changes, as in Gruber and Saez (2002)). This analysis reveals no evidence of an increase in the sales tax elasticity over time. For example, Column 4 of Table

[19] To clarify why inflation generates identifying variation, consider the following example. Suppose the pre-tax price of beer is $1 and state A has a nominal alcohol tax of 50 cents, while state B has no excise tax. If prices of all goods double, the gross-of-tax price of beer relative to other goods falls by 17% in state A but is unchanged in state B.

7b shows the effect of sales and excise tax changes on consumption over a three-year horizon. The estimates indicate that even in the long run, an increase in the excise tax rate has a large negative effect on alcohol consumption, whereas a similar increase in the sales tax does not.

We have fit a wide variety of other specifications to further probe the robustness of the results in Table 7. Estimating the model in levels with state fixed-effects (instead of first-differences) yields results similar to those reported above, as does estimating a linear model instead of a log-linear model. Excluding states with unusually high excise tax rates (i.e., the "Bible Belt" states) or observations in the upper or lower 5% of the distribution of the changes in tax rates also does not affect the results. In addition, we have estimated models of the price elasticity of demand using data on the average price of beer by year by state from the ACCRA survey available for a subset of years. When price is instrumented using the excise tax rate, the estimated price elasticity of demand is approximately 0.9, and the hypothesis that the price elasticity of beer consumption equals the sales tax elasticity is rejected with $p < 0.05$.

Relative Price Changes and Excise vs. Sales Taxes. As noted above, an important concern with our analysis is that the sales tax applies to 40% of consumption goods, and therefore leads to a smaller change in the relative price of alcohol than a change in the excise tax rate. As a result, the estimated sales tax elasticity may understate the effect of a change in a tax that is added at the register but applies only to alcohol. We evaluate the magnitude of this bias in two ways.

First, we estimate the model using only the thirty states that fully exempt all food items from the sales tax. In these states, changes in the sales tax have a larger effect on the relative price of alcohol (x) and all other goods (y). In addition, changes in the sales tax always affect the relative price of alcohol and food (and non-alcoholic beverages), which may be the most plausible substitute for alcohol. Column 5 of Table 7b shows that the coefficient estimates in this subsample are very similar to the results in the full sample. The stability of the sales tax elasticity as the base of the sales tax is narrowed suggests that the large difference between the effects of the excise and sales tax on alcohol consumption is unlikely to be fully explained by the difference in tax bases.

As an alternative approach, we use our two good model to calibrate the effect of a 1% increase in a (hypothetical) tax t^A that applies solely to alcohol and is excluded from the posted price. Treating y as a composite commodity of which 40 percent is subject to sales tax, observe that a 1% increase in the gross-of-sales-tax price $(1+t^S)$ un-saliently increases $\frac{p_x}{p_y}$ by approximately 0.6%. It follows that the effect of a 1% increase in the un-salient tax t^A that applies solely to alcohol is given by $\varepsilon_{1,1+t^A} = \frac{1}{0.6}\varepsilon_{1,1+t^S} = \frac{5}{3}\varepsilon_{1,1+t^S}$. Scaling up the largest estimated response to the sales tax in Table 7 of -0.20 by $\frac{5}{3}$ yields an estimate of $\varepsilon_{1,1+t^A} = -0.33$, which remains substantially below the excise tax elasticity estimates. Hence, this simple calibration also suggests that the difference in tax bases is unlikely to explain the estimated difference between the excise and sales tax effects.

In summary, averaging across the coefficient estimates in Tables 7a and 7b, the mean estimate of the gross-of-excise-tax price elasticity is 0.96. The mean estimate of the gross-of-sales-tax price elasticity is 0.033. Scaling up the sales tax coefficient by $\frac{5}{3}$, we obtain an implied elasticity of 0.06 for a tax that is applied solely to alcohol at the register. Combining these estimates yields a point estimate of $\theta = 0.06$. Note that we cannot reject the hypothesis that $\theta = 0$ – i.e., all individuals are inattentive to the tax added at the register – given the standard error on this point estimate. We conclude that most individuals focus on posted prices rather than full tax-inclusive prices when making consumption decisions about small goods, even in the long run.

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 most individuals 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. In this section, we attempt to distinguish between these two competing mechanisms in order to model behavior more precisely.

A few pieces of evidence in our preceding empirical analysis point toward the salience mechanism. First, 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. Second, one interpretation of the return of demand to pre-experiment levels after the intervention ended is that there were no persistent learning effects: individuals began to focus again on the pre-tax price once the tags are removed. In this case, however, we cannot rule out another plausible explanation: the set of individuals who shop for these durable goods may vary substantially across weeks, so customers in the weeks after the experiment may effectively have been untreated. Finally, in the alcohol consumption analysis, we find that the difference between the excise and sales tax elasticities persists over longer horizons. 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. This finding also points toward the importance of tax salience.

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. Among individuals who got 7 out of the 8 questions right, Coca Cola and cookies accounted for more than half the mistakes. Knowledge of the sales tax rate is also quite good: 75 percent reported the sales tax rate within 0.5 percentage points of the true rate, and 97 percent reported a rate between 6.75% and 8.75%. The modal answer (given by 15 percent of those surveyed) was 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. Coupled with the evidence that behavioral responses to taxation are larger when taxes are included in posted prices, this finding implies that many individuals choose not to compute tax-inclusive prices when making consumption decisions.

7 A Model of Bounded Rationality and Taxation

In the remainder of the paper, we focus on constructing a model that fits the evidence documented above while providing a tractable framework to analyze questions such as tax incidence and efficiency. We begin by characterizing individual behavior, and then turn to implications of the model for social welfare.

Model Setup. Consider the same environment as in section 3: an agent with wealth Z choosing consumption of two goods, x and y. Good x is subject to a sales tax (not included in the posted price) at rate t, while good y is untaxed. Choose units so that the pre-tax prices of x and y are both 1. In section 3, we assumed that a fraction $1-\theta$ of individuals choose not to compute tax-inclusive prices when making consumption decisions. Our objective here is to provide "micro foundations" for this assumption by constructing a model that generates a low value of θ and thereby matches the empirical evidence documented above.

We depart from the neoclassical model of consumer choice by assuming that the agent must pay a cost c to calculate the tax-inclusive price of good x, $p_t = 1 + t$. This cost could reflect a cognitive cost of deliberation (as in Conlisk 1996) or an opportunity cost of time. The agent can alternatively choose to make his consumption decision based solely on the pre-tax price, which is posted on the shelf and hence costless to compute (i.e., perfectly salient).

The agent makes two choices: whether to compute the tax-inclusive price p_t and how to allocate consumption given his perceived prices. This problem can be divided into a two-stage maximization: (1) choose an optimal bundle for any given perceived price p; (2) decide whether to spend c on computing p_t.

Consumption Decision with Quasilinear Utility. It is instructive to begin with a case without income effects by assuming that utility is quasilinear in y, as in Section 3. Let utility over x be given by a function $u(x)$, so that total 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 $t \geq 0$.

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 price 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 further discussion of this assumption to that case. Following the derivation in section 3, if the agent does compute the tax-inclusive price, he chooses a bundle (x^p, y^p) that satisfies

$$\begin{aligned} u\prime(x^p) &= 1+t \\ y^p &= Z - (1+t)x^p \end{aligned}$$

By contrast, if the agent does compute the tax-inclusive price, he chooses a bundle (x^*, y^*) that satisfies

$$\begin{aligned} u\prime(x^*) &= 1+t \\ y^* &= Z - (1+t)x^* \end{aligned}$$

Characterization of the Cognitive Decision. Now consider the decision of whether to pay the cost c and compute p_t. To characterize the agent's cognitive decision, we ask the question, "How much does the agent's utility rise (measured using a money metric) if he computes the tax-inclusive price p_t?" If this value is below the cost c, we infer that the agent will choose not to compute p_t.[20]

[20] A difficulty with this characterization is that if the agent literally followed this strategy when deciding whether to compute p_t, he would have to calculate the tax-inclusive price in order to know the utility gain from making this calculation. Thus the cost of solving the cognitive problem would be higher than the cost

The agent's utility gain from computing the tax-inclusive price is

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

Note that $G(t)$ 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^* gives:

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

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

$$G(t) \simeq -\frac{1}{2}\frac{\partial x}{\partial t}t^2 = \frac{1}{2}\varepsilon_{x,p}xt^2 \tag{7}$$

where $\varepsilon_{x,p} = -\frac{\partial x}{\partial t}\frac{1}{x}$ denotes the price elasticity of x. Hence, the utility cost from failing to compute the tax-inclusive price when optimizing consumption is a second-order function of the tax rate. It is optimal not to compute p_t if $G(t) < c$, i.e. if $t < T$ where $T = [\frac{2c}{x\varepsilon_{x,p}}]^{1/2}$. The threshold T is a bound for the range of taxes which the agent will rationally ignore. This threshold is increasing in c, as one would expect; when $c = 0$, the model collapses back to the neoclassical case where agents always calculate tax-inclusive prices. Now consider the minimum width of the range of ignored taxes *relative* to the cost of cognition:

$$\frac{T}{c} = [\frac{2}{x\varepsilon_{x,p}c}]^{1/2} \tag{8}$$

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

of fully optimizing consumption. This issue is an example of the generic problem of "regression" in bounded rationality models (Conlisk 1996). In this particular case, one might imagine that the agent solves the problem of whether to compute tax-inclusive prices for a particular class of goods (e.g. items in a grocery store) once, and then applies that rule whenever he considers buying those products. In this repeated decision setting, solving the cognitive problem once may be less expensive than computing the tax-inclusive price each time.

of this section:

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

This result shows 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 computing tax-inclusive prices) have negligible effects on utility. The envelope condition causes the first-order (u') terms to drop out in $G(t)$, leading to the result that the gain from computing tax rates is a second-order function of the tax rate. Figure 4 illustrates the result geometrically. In this figure, the individual's welfare loss from failing to optimize relative to p_t is given by the lost consumer surplus, triangle A. The size of this triangle is given by $\frac{1}{2}(t)|\frac{\partial x}{\partial t}|(t)$, which is precisely the expression for $G(t)$. As the tax rate t approaches 0, the size of this triangle diminishes at a second-order rate because both its height and width diminish linearly.

The practical implication of this result is that small cognitive costs can lead to substantial inattention to taxes. We can quantify what "small" and "substantial" mean by calculating the minimum range of taxes (T) that are ignored for various levels of the cognitive cost c using the formula in (8). 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 a 10% tax on a commodity on which the agent spends $\$10,000$ and has a demand elasticity of -1 is ignored if the cognitive cost $c > \$50$.

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

General Case: Arbitrary Utility. The limiting result established above also applies in the general case when utility is not quasilinear. Suppose the agent's utility is given by

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

where $v(y)$ is increasing and strictly concave. If he does not compute the tax-inclusive price, his perceived budget constraint is

$$x + y = Z$$

His actual budget constraint, which must be satisfied, is

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

If the agent responds only to the pre-tax price of x at the margin, he sets

$$u'(x) = v'(y). \tag{9}$$

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. There are two natural 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 (9) 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 (9) given total income Z, and finances his spending on x using what's left: $x = \frac{Z-y^p}{1+t}$. We focus here on the solution where the agent chooses x first; the key qualitative results hold in the second case as well.[21]

Letting $x^*(t)$ and $y^*(t)$ denote the optimal consumption allocation given a true tax rate

[21] 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 net-of-tax income. The agent then follows the rule that yields higher expected utility, which ultimately results in the behavior characterized here. In the quasilinear case, Jensen's inequality directly implies that choosing the good with diminishing marginal utility first (x) is optimal. See Reis (2006) for a related analysis of the choice between a consumption and savings rule in a lifecycle model.

t, the gain in utility from computing p_t is:

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

This expression can be converted into a money-metric by dividing the utility gain by the marginal utility of wealth, which equals $v'(y^*(t))$ at the optimum. In particular, define

$$G(t) = \frac{\widetilde{G}(t)}{v'(y^*(t))}$$

as the amount of money an inattentive agent must be paid to bring his utility to the full-optimization level. Using a quadratic approximation for the utility function, we derive the following expression for $G(t)$ in the appendix:

$$G(t) = \frac{1}{2}t^2 x \varepsilon_{x,p}[1 + (\frac{x}{y})\gamma_y] \quad (10)$$

where $\gamma_y = \frac{v''(y^*)}{v'(y^*)}y^*$ measures the curvature of utility over y. When utility is quasilinear, $\gamma_y = 0$, and $G(t)$ reduces to (7). When $\gamma_y > 0$, i.e. marginal utility over y is also diminishing, the tax-change has an income effect that leads to a shift in demand for y relative to demand for x beyond the pure price effect. As a result, the gain in computing the tax-inclusive price is larger, because it has two components: one that reflects the substitution effect (purchase less x because its price is higher) and a second that reflects the income effect (purchase even less x because net income is now lower). Nonetheless, even when utility is not quasilinear, G remains a quadratic function of t. Hence, the result that $\lim_{c->0} \frac{T}{c} = \infty$ still holds. Second-order cognitive costs generate a first-order range of inattention to taxes irrespective of the utility function.

Aggregate Demand with Heterogeneous Agents. Having characterized behavior for a single agent in terms of his cognitive cost c, now consider an economy populated by heterogeneous boundedly rational agents. Suppose, for simplicity, that agents have identical preferences but heterogeneous cognitive costs, distributed according to a cdf $F(c)$. Then the fraction of individuals who choose to compute tax-inclusive prices is given by:

$$\theta = F(G(\cdot)) = (\frac{1}{2}t^2 x\varepsilon_{x,p}[1 + (\frac{x}{y})\gamma_y]) \tag{11}$$

Intuitively, the threshold level at which the gains from computing tax-inclusive prices are offset by the cognitive costs of doing so is given by $G(t)$. Individuals with c below this threshold compute p_t while the rest focus on the salient posted price p. Hence, aggregate demand is given by

$$x(t) = \theta x^*(t) + (1 - \theta)x^p(t) \tag{12}$$

as was assumed in the two-type framework in section 3.

Equation (12) shows that introducing costs of computing tax-inclusive prices into the neoclassical model enables us to match the qualitative features of the data. Consistent with the evidence in sections 4-6, some individuals choose not to compute tax-inclusive prices despite having the information to do so, making the behavioral response to a commodity tax depend on whether that tax is included in the posted price. The model can also explain why (1) showing tax-inclusive prices does not have "spillover effects" to the adjacent control categories or persistent effects after the tags were removed and (2) individuals remain inattentive to taxes in the long run, consistent with the evidence on alcohol consumption. The key feature of the model that enables it to match the data is that agents must pay a cost *every time* they calculate a tax-inclusive price, an assumption that we view as a plausible description of the cognitive process in practice. As a result, salience effects remain important in the analysis of tax policies in steady-state, and not just on the transition path after a tax policy change.

Perhaps more importantly, the preceding analysis shows that the model can match the quantitative features of the data – namely, that θ is low – with a plausible $F(c)$. The calibrations in Table 1 indicate that the utility gain from optimizing relative to tax-inclusive prices as opposed to pre-tax prices on goods such as cosmetics and alcohol is very small. Hence, with small cognitive or time costs, one should expect that most individuals will use the heuristic of focusing on the salient posted price rather than computing the tax-inclusive price for each product when they shop.

7.1 Additional Predictions: Tax Heuristics

A useful feature of the model is that one can make predictions about tax perceptions and behavioral responses to taxation in contexts beyond commodity taxes on small goods. In particular, the equation for θ in (11) gives a description of how tax perceptions are determined. The comparative statics of this equation shed light on the circumstances under which agents are likely to use the simple "heuristic" of focusing on the salient pre-tax price.

The first prediction is that $\frac{\partial \theta}{\partial x} > 0$: more individuals pay attention to taxes on large goods relative to small, repeated purchases. This is because the cognitive cost is fixed whereas the utility benefit of computing p_t scales up with expenditure on the good. Behavioral responses to taxation are predicted to be larger for more expensive one-time purchases (e.g. durables), even if the underlying price elasticities of demand are similar to those for smaller goods. Analogously, behavioral responses to income taxation will be larger among the rich – consistent with evidence in Feldstein (1995), Goolsbee (2000), and Saez (2004) – since they have more at stake in dollar terms (and therefore may choose to pay the fixed cost c by hiring an accountant).

Second, $\frac{\partial \theta}{\partial t} > 0$: individuals are more likely to pay attention to taxes when tax rates are high, because the utility cost of ignoring the tax grows with the square of the tax rate. By extension, in a more general dynamic setting, individuals should pay greater attention to large tax reforms than small changes in marginal rates. This point has bearing on the interpretation of empirical estimates of behavioral responses to taxation. Many recent microdata-based studies estimate elasticities by examining how individual behavior changes in a short window around a tax reform that may not be perfectly salient. If individuals incur small cognitive costs to understand the incentive effects of new tax reforms, they will rationally not pay attention to small reforms. Thus, one may underestimate the impact of taxation on behavior using studies of short-run responses to small tax changes. This reasoning could perhaps 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 around tax reforms (e.g. Gruber and Saez 2003).

Third, $\frac{\partial \theta}{\partial \varepsilon_{x,p}} > 0$: individuals are more likely to compute tax-inclusive prices for products where demand is elastic; if demand is inelastic, there is less benefit from paying attention, since one would change behavior relatively little if one were to calculate the true price. In the extreme case as $\varepsilon_{x,p}$ tends to 0, there is no benefit in calculating the tax-inclusive price at all, since one's consumption choice is essentially unaffected by price. This force magnifies the effect of elasticities on behavioral responses to taxation: higher elasticities lead to larger responses both through a direct effect of inducing larger responses and an indirect effect of raising the probability that individuals pay attention to taxes.

Fourth, $\frac{\partial \theta}{\partial \gamma_y} > 0$: individuals who have more curved utility – e.g. because of credit constraints or consumption commitments – are more likely to pay attention to taxes. Intuitively, these individuals can potentially lose a lot of utility from ignoring taxes when making decisions, and therefore are more likely to compute tax-inclusive prices and respond to changes in incentives.

A final set of predictions can be obtained in an extension of the model where individuals face non-linear taxes. In the interest of space, we discuss these predictions without providing formal derivations. One simple prediction is that individuals are more likely to pay attention to policies that induce discontinuous changes in tax burdens – e.g. due to eligibility cutoffs or tax holidays – than to marginal taxes that induce continuous changes in the tax burden. If a small change in behavior (e.g. buying a good one day later or reducing earned income by \$1 to meet a cutoff) leads to a non-negligible change in utility, the smoothness requirement for agent's objective function, which leads to the second-order expression for $G(t)$ in (11), does not hold. As a result, the utility cost of ignoring a discontinuous change in the tax burden is a linear (first-order) function of the amount of the tax. In contrast, the utility cost of ignoring a change in the slope of the tax schedule (a continuous change in the marginal) rate is second-order. Consequently, the model predicts that behavioral responses to "cliffs" in the tax schedule will be large, whereas there will be limited "bunching" at kink points, consistent with the evidence of Saez (2002) and others. A closely related prediction is that boundedly rational agents should respond more on the extensive margin than on the intensive margin to tax policies. For example, an agent's potential gain in utility from entering the labor force in response to the introduction of the Earned Income Tax Credit (EITC) could be much larger than the gain from reoptimizing hours worked because of the change in the marginal incentive. Hence, agents on the margin of entering the labor force may be more likely to pay attention to the EITC incentive than agents who are already working..

The predictions above all pertain to the simple question of when individuals will opt to use the heuristic of focusing on the pre-tax price vs. computing tax-inclusive prices. One can build on this approach and partially characterize the heuristics used in more complex environments by calculating the willingness to pay for various pieces of information about the tax code. To illustrate this approach, we consider an application involving a non-linear tax schedule.

Application: Average vs. Marginal Tax Rates. de Bartolome (1995), Liebman (1998), Liebman and Zeckhauser (2005), and others have documented that individuals tend to be more aware of and responsive to average income tax rates relative to marginal rates.[22] In contrast, much of the theoretical and empirical public finance literature has focused on analyzing marginal tax 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 utility gain from optimizing relative to the marginal tax rate on labor income and the average tax rate on labor income.

Introducing labor into the two-good consumption model analyzed above, consider a utility function of the following parametric 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.

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

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:

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

Given the functional form of utility, $U(Z)$ 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.

Now focus on the subcase 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, which we now allow to be non-zero (i.e., the agent can use a heuristic of assuming that $t_y^p=20\%$ of his income goes to the tax). 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 $U(Z) = Z$. 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 have limited attention and face important budgeting decisions.[23] More generally, this analysis suggests that legislated marginal rates – which are the focus of much of the existing theoretical and empirical literature – may be less important in determining behavior than broad tax perceptions.

[23]This result does not necessarily imply, however, that agents should use the average rate as a proxy for the marginal rate when making labor supply decisions. Evidence that agents behave in this manner suggests that there are additional factors, outside our simple model, which lead agents to confuse these two rates.

8 Efficiency Cost and Incidence of Taxation

In this section, we characterize the efficiency costs and incidence of taxes that are not perfectly salient. Auerbach (1985) and Kotlikoff and Summers (1987) present thorough analyses of excess burden and incidence in the traditional model where agents fully perceive taxes. Here, we present a parallel analysis of the excess burden and incidence of a tax that is not perfectly salient using the two-good model developed above.[24]

8.1 Efficiency Cost

Following Mohring (1971), we define the excess burden of a tax using the concept of equivalent variation.[25] To incorporate tax salience effects, it is necessary to define generalized versions of the indirect utility and expenditure functions that allow for prices and taxes to have different effects. Note that these functions will differ between agents who compute and do not compute tax-inclusive prices. We begin with a general definition of excess burden that applies to any individual, irrespective of his optimization strategy. We then use the structure of our bounded rationality model to derive expressions for aggregate deadweight loss in terms of θ,the fraction of attentive agents.

Definitions. Let $V(p,t,Z)$ denote the agent's indirect utility as a function of the posted price, the tax levied in addition (e.g., sales tax), and wealth. Let $e(p,t,V)$ denote the agent's expenditure function, which represents the wealth necessary to attain utility V given the posted price and tax. Let $x(p,t,Z)$ denote uncompensated (Marshallian) demand for x and $x_c(p,t,V)$ denote the compensated (Hicksian) demand. Fixing pre-tax prices of x and y at 1 as above, let $R(t^E,t^S,Z)=(t^E+t^S)x(1+t^E,t^S,Z)$ denote the tax revenue raised by imposing an tax t^E that is included in the posted price and a tax t^S that is not.

The excess burden of introducing a tax t that is not included in the posted price in a

[24] In the context of the present model, existing models can be interpreted as calculating excess burden of an increase in t^E. The analysis below focuses instead on the excess burden of an increase in t^S.

[25] The other commonly used approach of defining excess burden using compensating variation (Diamond and McFadden 1974) raises some conceptual problems in the case where some agents are inattentive to taxes. If the money is returned lump sum, the excess burden depends on whether agents are attentive to this transfer (yet inattentive to the proportional tax) when making consumption decisions.

previously untaxed market is given by:

$$\begin{aligned} EB_t &= Z - e(1,0,V(1,t,Z)) - R(0,t,Z) \\ &= Z - e(1,0,V(1,t,Z)) - tx(1,t,Z) \end{aligned}$$

The value EB_t can be interpreted as the amount of additional tax revenue that could be collected from the consumer while keeping his utility constant if the distortionary tax were replaced with a lump-sum tax. It is straightforward to extend this expression to the case where the initial equilibrium is already distorted by pre-existing taxes. Suppose that there are two pre-existing taxes on good x: a tax t_0^E included in the posted price and a tax t_0^S that is added later. Let $t_0 = t_0^E + t_0^S$ denote the total tax levied on good x. For concreteness, one could think of t_0^E as the excise tax and t_0^S as the sales tax, in analogy with our preceding empirical work. Letting $V_1 = V(1 + t_0^E, t_0^S + \Delta t, Z)$ denote the agent's indirect utility after the tax increase, the excess burden of a sales tax increase Δt is given by

$$EB_{\Delta t} = Z - e(1 + t_0^E, t_0^S, V_1) - [R(t_0^E, t_0^S + \Delta t, Z) - R(t_0^E, t_0^S, e(1 + t_0^E, t_0^S, V_1))]$$

which, following Auerbach (1985), can be simplified to

$$\begin{aligned} EB_{\Delta t} &= Z - e(1 + t_0^E, t_0^S, V_1) - \Delta t x(1 + t_0^E, t_0^S + \Delta t, Z) \\ &\quad + (t_0^E + t_0^S)[x_c(1 + t_0^E, t_0^S, V_1) - x(1 + t_0^E, t_0^S + \Delta t, Z)]. \end{aligned} \tag{13}$$

The last term in (13) reflects the "fiscal externality" (lost tax revenue) that individuals impose on the government when they reduce consumption of x. Our objective is to derive empirically implementable expressions for (13) and contrast them with those obtained in the traditional model.[26] We begin with a graphical derivation for the case of quasilinear utility, and then turn to the general case.

[26] The excess burden of an excise tax increase can be defined analogously. We do not analyze the excess burden of taxes included in posted prices explicitly here since the results are essentially identical to those obtained in the traditional model.

8.1.1 Graphical Intuition: Quasilinear Utility

When utility is quasilinear in y, the Marshallian and Hicksian demands coincide and are independent of wealth and utility: $x(p,t,Z) = x_c(p,t,V) \equiv x(p,t) \forall Z, V$. Hence, the excess burden of a tax increase can be characterized simply using the Marshallian demand curve and the notion of "consumer surplus." In Figure 4, we illustrate the excess burden of a tax increase for two agents, one who does not pay attention to the sales tax and another who does. A formal algebraic derivation of the formulas below, starting from the definition of excess burden in (13), is given in the Appendix.

First, consider a consumer whose cognitive cost is sufficiently high that he fails to compute the tax-inclusive price even after the sales tax rate rises. This inattentive agent's consumption decision in the initial situation with excise tax t_0^E and sales tax t_0^S is depicted by x_0^p, which is the point where $u'(x_0^p) = 1 + t_0^E$. The agent does not reoptimize in response to the sales tax increase Δt. The loss in consumer surplus from the failure to reoptimize is shown by triangle A, whose area equals

$$G(\Delta t) = -\frac{1}{2}\Delta x \Delta t = -\frac{1}{2}\frac{\partial x}{\partial t}(\Delta t)^2$$

consistent with equation (7).

Taking into account the higher price of x, the tax increase reduces overall consumer surplus by $\Delta CS = x_0^p \Delta t$, which is given by rectangle B. The change in tax revenue is also given by the rectangle B: $\Delta R^p = x_0^p \Delta t$. Total surplus is unchanged, because the lost consumer surplus is fully transferred to the government. Hence, the excess burden associated with the sales tax increase is zero:

$$EB_{\Delta t}^p = \Delta CS - \Delta R = 0.$$

Intuitively, since there is no change in consumption of x when the agent ignores the tax change, the sales tax increase is equivalent to a lump-sum tax because utility is linear in y. With quasilinear utility, both the lump-sum tax and the ignored sales tax are borne fully on consumption of good y, leading to zero excess burden. As we show below, however, excess burden is strictly positive when agents ignore tax changes if utility over y is concave.

Now consider the excess burden of the same tax increase for an agent with zero cognitive cost, so that he always computes p_t. Letting x_0^* denote this attentive agent's consumption choice in the initial situation and x_1^* the choice after the tax increase, the effect of the tax increase on tax revenue collected from such an agent is

$$\Delta R = (t_0 + \Delta t)x_1^* - t_0 x_0^* = \Delta t x_0^* + (t_0 + \Delta t)\frac{\partial x}{\partial t}\Delta t$$

where the latter term reflects the loss in revenue from the agent's behavioral response. The excess burden of the tax increase for the agent who computes the tax-inclusive price is given by the familiar "Harberger trapezoid" with area A+C:

$$\begin{aligned} EB_{\Delta t}^* &= -\frac{1}{2}\Delta x \Delta t - t_0 \Delta x = -\frac{1}{2}\frac{\partial x}{\partial t}(\Delta t)^2 - t_0\frac{\partial x}{\partial t}(\Delta t) \\ &\approx t_0 \varepsilon_{x,p} x_0 \Delta t + \frac{1}{2}\varepsilon_{x,p} x_0 (\Delta t)^2 \\ &= = t_0 \varepsilon_{x,p} x_0 \Delta t + G(\Delta t). \end{aligned}$$

This approximation shows that the excess burden of the tax is given by the sum of the fiscal externality imposed on the government when there is a pre-existing tax t_0 and the private gain to the agent of computing the tax and re-optimizing his demand. Note that the first term rises linearly with Δt, while the second term is a function of Δt^2 and hence will typically be much smaller – a point to which we return in the discussion below.

Aggregate Deadweight Loss. Thus far, we have computed the excess burden of a tax increase for individuals who by assumption either do or do not pay attention to the tax change. We now build on this analysis to characterize aggregate deadweight loss in the economy where agents have heterogeneous costs of cognition and the fraction who compute the tax-inclusive price is endogenously determined by the tax rate. Let $\theta_0 = F(\frac{1}{2}\varepsilon_{x,p}x_0(t_0^S)^2)$ denote the fraction of agents who compute the tax-inclusive price when the sales tax rate is t_0^S and $\theta_1 = F(\frac{1}{2}\varepsilon_{x,p}x_0(t_0^S + \Delta t)^2)$ denote the same after the tax increase.

To calculate aggregate deadweight loss, consider three groups of agents. First, agents with the highest cognitive costs (fraction $1 - \theta_1$) ignore both the initial tax and the tax increase, and therefore do not change their consumption of x. There is no excess burden from increasing the tax on these agents. Second, agents with the lowest cognitive costs (fraction θ_0), optimize relative to the initial sales tax and reoptimize relative to the sales tax increase. These agents contribute $\theta_0 EB^*_{\Delta t}$ to aggregate deadweight loss.

Third, agents with cognitive costs in the intermediate range (fraction $\theta_1 - \theta_0$) initially ignored the sales tax, but now compute the full tax-inclusive price. In addition, these agents now expend the cognitive cost of thinking about the tax, which further contributes to deadweight loss. To calculate the size of the cognitive cost increase, observe that the total cognitive cost incurred in the population from the computation of sales taxes is given by

$$C(t^S) = \int_0^{G(t^S)} cf(c)dc \tag{14}$$

Hence $\frac{\partial C}{\partial t^S}(t^S) = G(t^S)f(G(t^S))\frac{\partial G}{\partial t^S} = G(t^S)\frac{\partial \theta}{\partial t}$. It follows that the third group of agents contributes $(\theta_1 - \theta_0)EB^*_{t_0^S+\Delta t} + (\theta_1 - \theta_0)G(t_0^S)$ to aggregate deadweight loss. Summing the terms and simplifying, we obtain

$$\begin{aligned} DWL(\theta, \Delta t) &= \theta_0 EB^*_{\Delta t} + (\theta_1 - \theta_0)(EB^*_{t_0^S+\Delta t} + G(t_0^S)) \\ &= \theta_1 EB^*_{\Delta t} + (\theta_1 - \theta_0)(EB^*_{t_0^S} + G(t_0^S)) \\ &= \theta_1[\frac{1}{2}x\varepsilon_{x,p}(\Delta t)^2 + x\varepsilon_{x,p}t_0(\Delta t)] + \theta_1\varepsilon_{\theta,t}x\varepsilon_{x,p}t_0(\Delta t) \end{aligned} \tag{15}$$

This expression shows that an increase in t^S generates additional deadweight loss through two margins, which correspond to the two terms in (15). First, the tax increase amplifies distortionary costs for individuals who compute tax-inclusive prices through the traditional Harberger channel. This effect is attenuated by θ_1 – the fraction of individuals who compute the tax – since there is no excess burden for inattentive individuals in the quasilinear case. Second, the tax increase raises the fraction of individuals who are attentive to tax-inclusive prices. For the marginal "switchers" who begin to compute p_t after the tax change, the perceived tax increase is the full amount of the sales tax, $t_0^S + \Delta t$ (since they previously ignored the sales tax entirely), potentially leading to a substantial distortion in behavior.

The marginal switchers earn no net private benefit from computing the tax-inclusive price, since the cognitive cost G fully offsets this gain. The only consequence of their change in behavior is therefore a first-order revenue loss for the government, since these agents now consume $x^* < x^p$. The excess burden due to this channel is proportional to $\varepsilon_{\theta,t}$, which measures the sensitivity of tax perceptions in the population to the tax rate. Note that $\varepsilon_{\theta,t}(t^S) = \frac{\partial\theta}{\partial t}\frac{t^S}{\theta}G(t^S) = o((t^S)^3)$ is a cubic function of the sales tax rate. Hence, the endogeneity of tax perceptions can substantially amplify the marginal deadweight loss of a tax increase in a market where tax rates are already high, but is of less relevance when initial tax rates are low.

Discussion. One important implication of the analysis above is that taxes which generate small utility losses if ignored by individuals can nevertheless have large effects on social welfare and revenue. Bounded rationality and inattention can therefore be relevant in the analysis of many large-scale tax policies. Mathematically, the source of this result is an individual's utility loss from ignoring the tax (G) is a function of $(\Delta t)^2$, whereas government revenue (ΔR) and deadweight loss (DWL) are functions of Δt. Intuitively, reoptimizing following a tax change has small benefits for the agent because he has already chosen his consumption bundle to maximize utility. Tax revenue, however, has not been optimized in this manner, and therefore rises linearly with the tax rate. Similarly, excess burden can be substantial 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.

To illustrate the practical significance of this result, we calibrate G, ΔR, and DWL for a range of tax increases in Table 10. We contrast two polar cases: (1) $\theta = 0$: all agents have infinite cognitive costs and therefore always ignore the sales tax and (2) $\theta = 1$: all agents have zero cognitive cost and therefore always calculate tax-inclusive prices. Note that these polar cases abstract from the endogeneity of θ to the tax rate, a factor that further amplifies the difference between individual and social welfare changes. As above, consider a good on which the agent spends $x = \$10,000$ with a price elasticity of demand $\varepsilon_{x,p} = 1$. Suppose the initial excise tax (included in the posted price) is $t_0^E = 25\%$ and the initial sales tax is

$t_0^S = 0$. Now consider the effects of introducing a 10% sales tax. For the individual, as noted above, ignoring a 10% tax increase leads to a modest utility loss of $G = \$50$ under these parameters. For society, however, the effects of the tax are much larger. When $\theta = 0$, the change in revenue is $10\% \times 10,000 = \$1,000$. In contrast, when $\theta = 1$, the individuals' behavioral response of reducing demand for x reduces ΔR to \$720. When $\theta = 0$, the sales tax has no effect on social welfare in the quasilinear case. But when $\theta = 1$, the sales tax generates an efficiency cost of \$300 – 5 times larger than an individual's private utility loss would have been from ignoring the tax. These calibrations show that whether a tax increase is made salient or not can have a large effect on social welfare and government revenue, even though the welfare costs to the agent of ignoring the tax change are small.

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 many agents 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 social welfare.

Taking this logic one step further, one is tempted to infer that lowering perceived tax rates by "hiding" taxes – to the extent possible given the endogeneity of θ – will reduce efficiency costs. However, this surprising result is unique to the quasilinear case, as we show in the next section.[27]

[27] In addition, as emphasized by Becker and Mulligan (2003), "hidden" taxes may raise efficiency costs due to political economy issues, an important issue that it outside the scope of our analysis.

8.1.2 General Case

When utility is given by a general additive function $u(x) + v(y)$, excess burden for an agent with zero cognitive cost who computes tax-inclusive prices is approximately

$$EB^*_{\Delta t} \approx t_0 \varepsilon^c_{x,p} x_0 \Delta t + \frac{1}{2} \varepsilon^c_{x,p} x_0 (\Delta t)^2$$

where $\varepsilon^c_{x,p}$ denotes the compensated elasticity of demand (see e.g., Auerbach 1985).

To calculate EB^p for the inattentive agent, assume temporarily that there are no pre-existing taxes. The excess burden of introducing a sales tax t is

$$EB^p_t = Z - e(1, 0, V^p(1, t, Z)) - tx^p(1, t, Z)$$

where $V^p(1, t, Z)$ is the utility attained by the agent when he does not optimize relative to the true tax-inclusive price. Letting $V(1, t, Z)$ denote the utility attained by the agent when he does optimize relative to the tax-inclusive price, some algebra yields

$$EB^p_t = EB^*_t + [e(1, 0, V(1, t, Z))) - e(1, 0, V^p(1, t, Z))] + t[x^* - x^p]. \tag{16}$$

This is an exact expression for the excess burden, and can be computed for a given indirect utility and expenditure function. To elucidate the key determinants of EB^p_t, we use a series of approximations in the Appendix to show that

$$EB^p_t \approx \frac{1}{2} t^2 x (\frac{x}{z})(\frac{x}{y}) \varepsilon_{x,Z} \gamma_y \tag{17}$$

Equation (17) provides an elasticity-based measure of the excess burden of taxation for an agent who ignores the tax on good x. This expression shows that when utility is not quasilinear (i.e., $\gamma_y > 0$), introducing the sales tax has a strictly positive excess burden even though the agent does not reoptimize his consumption bundle. To understand this result, recall that the excess burden of a distortionary tax is determined by the extent to which the agent's consumption allocation differs from the allocation he would choose if subject to a lump sum tax of an equivalent amount. In the quasilinear case, the agent's consumption bundle when ignoring the tax coincides precisely with the bundle he would choose under lump sum taxation, because there are no income effects: if subject to lump sum taxation, the agent would optimally reduce only consumption of y (the good over which utility is linear). When utility is not quasilinear, the agent's optimal consumption bundle under lump sum taxation involves a reduction in consumption of *both* x and y. Consequently, when the agent ignores the un-salient tax, he ends up with a sub-optimal consumption bundle, one that yields lower utility than he would attain with an equivalent lump-sum tax.

A useful feature of (17) is that it is analogous to Harberger's formula for salient taxes, in that it can be applied in a wide variety of contexts using standard methods to estimate the relevant elasticities. The amount of lost surplus is determined by (1) the income elasticity $\varepsilon_{x,z}$, which determines how much the agent overconsumes x when he does not recognize the lost income due to the tax; and (2) the curvature of the utility function, which determines the utility cost of the resulting underconsumption of y. The excess burden is a second-order function of the tax-rate, owing to the usual envelope condition from agent optimization that eliminates first-order costs when there is no pre-existing sales tax.

Some examples may be helpful in understanding more concretely why the income elasticity is the key determinant of excess burden. First consider an agent choosing consumption of two goods – cars and healthcare – and suppose that cars are taxed in an un-salient manner (e.g. through opaque excise fees) that the agent is unaware of. Suppose the agent chooses his car first, overspending because he does not perceive the taxes he owes, and later needs healthcare, over which he has very curved utility (high γ_y). In this case, the "hidden" tax on cars may generate a large efficiency cost, potentially larger than the efficiency cost of a perfectly salient tax that only distorts behavior through the conventional substitution effect. In contrast, consider an agent choosing consumption of cars and housing. Suppose utility is linear over housing – the agent spends any extra money he has on a better house, leading to $\varepsilon_{x,z} = 0$ for the car. In this case, the "hidden" tax on cars has no efficiency cost, since it induces the agent to choose the allocation he would have chosen under lump sum taxation. The extent to which "hidden" taxes create a deadweight burden thus is determined by the nature of the goods upon which the impact of the tax is ultimately borne.

As a numerical illustration, consider the efficiency cost of the sales tax for an agent who ignores sales taxes when making consumption decisions. The sales tax applies to $\frac{x}{z} = 0.4$ of consumption on average in the U.S. Suppose the Engel curve for goods subject to the sales tax is linear ($\varepsilon_{x,z} = 1$) and that the coefficient of relative risk aversion over the remaining consumption bundle is $\gamma_y = 1$. Then $EB_t^p = 0.133t^2x$. The marginal excess burden of raising the sales tax as a percentage of revenue collected is given by

$$\frac{MEB_t^p}{x} = \frac{1}{x}\frac{dEB_t^p}{dt}(t = 0.1) = 0.26t.$$

Hence, under these parameter choices, raising the sales tax further from the current rate of approximately 10% would generate an efficiency loss of approximately 2.5% of the revenue raised if individuals ignore the sales tax and absorb its impact entirely on untaxed goods.[28]

Aggregate Deadweight Loss. Deadweight loss in the economy where a fraction θ of individuals compute the tax-inclusive price is given by

$$DWL(\theta, t) \approx \frac{1}{2}x_0t^2[\theta\varepsilon_{x,p}^c + (1-\theta)(\frac{x}{z})(\frac{x}{y})\varepsilon_{x,Z}\gamma_y] + C(t^S) \tag{18}$$

where $C(t^S)$ is defined as in (14). Differentiating (18) and collecting terms, we obtain the following expression for the marginal deadweight loss of raising the tax rate from an initial sales tax of t_0^S (and excise tax $t_0^E = 0$).

$$MDWL(\theta, t_0^S) = x_0t_0^S[\theta\varepsilon_{x,p}^c + (1-\theta)(\frac{x}{z})(\frac{x}{y})\varepsilon_{x,Z}\gamma_y] + \theta x_0t_0^S\varepsilon_{\theta,t}\varepsilon_{x,p} \tag{19}$$

This expression, which mirrors (15) in the quasi-linear case, shows that an increase in t again generates additional deadweight loss through two margins: (1) it amplifies distortionary costs for both attentive and inattentive individuals and (2) it raises the fraction of individuals who compute tax-inclusive prices.[29] As above, the latter effect strictly raises deadweight loss – despite reducing the excess burden due to the distortionary income effect – because the

[28] The assumption that agents choose x and then set y as the residual is not innocuous in this calculation. In a dynamic setting, it is more plausible to assume that agents bear the tax on both x and y over time. This will mitigate the excess burden, since the consumption allocation is less distorted relative to what would be chosen under lump sum taxation. However, the qualitative result that $EB_t^p > 0$ still holds.

[29] Equation (19) corresponds to a first-order approximation of the formula derived in equation (15) for the quasilinear case: $\frac{DWL(\theta,\Delta t)}{\Delta t}$ approaches $MDWL(\theta, t_0^S)$ as $\Delta t \to 0$.

added cognitive costs offset the utility benefits of computing p_t for the marginal switchers.

Discussion. While we have motivated our analysis by a model of bounded rationality, the expression for excess burden for agents who do not compute p_t in equation (17) actually does *not* rely on this particular model of behavior. This point reflects Bernheim and Rangel's (2007) insight that making welfare statements in a model where agents' choices depend on "ancillary conditions," such as the salience of taxes, does not require a positive theory of behavior. Rather, what is necessary is a means of judging which ancillary condition reveals the agent's true ranking of choices. If one directly assumes that the agent's choices in the state where taxes are perfectly salient reveal his true preferences, one obtains (17) without further assumptions about why salience matters. Hence, in the derivation of EB^p for an agent who does not compute tax-inclusive prices, the bounded rationality theory can be viewed as a simple justification for why the choices made when taxes are salient reveal true preferences. The structure of our bounded rationality model is, however, needed to obtain the expressions for aggregate deadweight loss in (18) and (19). These expressions depend on the fraction of agents who pay attention to taxes at a given tax rate. Computing this fraction and its sensitivity to tax changes requires a positive theory of tax perceptions. Our model of cognitive costs, in which some agents endogenously choose to compute p_t, provides such a theory.

In the simple model we consider here, the bounded rationality structure places an upper bound on the excess burden that can be caused by taxes that agents ignore. Since EB^p arises entirely from a private utility cost, an agent with cognitive cost c must have $EB^p < c$, else he would compute the tax-inclusive price. This would seem to imply that the "distortionary income effect" of un-salient tax policies will generally be negligible in magnitude. While this inference is correct in a literal interpretation of our model, we believe that distortionary income effects could be large in a more general model in which agents face uncertainty or unawareness regarding the benefits of paying attention to different tax policies. In such an environment, uninformed agents may ignore potentially valuable aspects of the tax code because the cost of optimizing relative to each aspect of the tax code may outweigh the expected benefit from doing so. In this situation, levying a large "hidden tax" on an inattentive agent could potentially lead to a substantial reduction in his utility.

In summary, our analysis identifies two additional factors beyond the compensated elasticity that are relevant in evaluating the efficiency cost of taxation: distortionary income effects (allocation errors) and the endogeneity of tax perceptions. Much as Harberger's analysis identified $\varepsilon^c_{x,1+t}$ as a key parameter to be estimated in subsequent empirical work, the present analysis suggests that estimating tax perceptions (θ and $\varepsilon_{\theta,t}$) and the magnitude of allocation distortions due to taxes ($\varepsilon_{x,Z}\gamma_y$) could further improve our understanding of the efficiency costs of taxation.

8.2 Incidence

How is the burden of a commodity tax shared between consumers and producers in competitive equilibrium when individuals are boundedly rational? To answer this question, we model the demand side of the economy as above, with a set of individuals with different cognitive costs. Individuals who choose to pay attention to taxes have a demand function $D(p_t) = D(p+t)$, while the rest of the individuals' demand is given by $D(p)$. We model the supply side of the market using a supply curve $S(p)$, which is a function solely of the pre-tax price. At the market-clearing pre-tax price p,

$$\theta(t)D(p+t) + (1-\theta(t))D(p) = S(p) \tag{20}$$

where $\theta(t)$ is endogenous to the tax rate as noted above.

Our objective is to calculate $\frac{dp}{dt}$ and $\frac{dp_t}{dt}$, the incidence of the tax on producers and consumers, respectively. Implicitly differentiating (20), we obtain

$$\frac{\partial D}{\partial p}[\frac{dp}{dt} + \theta] + \frac{\partial \theta}{\partial t}[D(p+t) - D(p)] = \frac{\partial S}{\partial p}\frac{dp}{dt}$$

which implies

$$\begin{aligned} \frac{dp}{dt} &= -\frac{\varepsilon_{D,p}\theta(1+\varepsilon_{\theta,t})}{\varepsilon_{S,p}+\varepsilon_{D,p}} \\ \frac{dp_t}{dt} &= 1 + \frac{dp}{dt} = \frac{\varepsilon_{S,p}+\varepsilon_{D,p}[1-\theta(1+\varepsilon_{\theta,t})]}{\varepsilon_{S,p}+\varepsilon_{D,p}} \end{aligned} \tag{21}$$

where $\varepsilon_{D,p} = -\frac{\partial D}{\partial p}\frac{p}{D}$ is the price elasticity of demand, $\varepsilon_{S,p} = \frac{\partial S}{\partial p}\frac{p}{S}$ is the price elasticity of supply, and $\varepsilon_{\theta,t} = \frac{\partial \theta}{\partial t}\frac{t}{\theta}$ is the elasticity of θ with respect to the tax rate. In the traditional model, $\theta = 1$ and $\varepsilon_{\theta,t} = 0$, and (21) reduces to the standard formula for incidence (see e.g., Kotlikoff and Summers (1987)). This formula reflects the well-known result that the more elastic factor bears less of the burden of the tax, e.g.:

$$\partial[\frac{dp_t}{dt}]/\partial\varepsilon_{D,p} < 0$$

Fixed θ. To understand the implications of (21) when agents are boundedly rational, it is helpful to start with the case where $\theta < 1$ and $\varepsilon_{\theta,t} = 0$, i.e. the fraction of inattentive individuals is fixed at some positive level. In this case, the incidence on supply reduces to

$$\frac{dp}{dt} = -\frac{\varepsilon_{D,p}\theta}{\varepsilon_{S,p} + \varepsilon_{D,p}} = -\frac{\varepsilon_{D,1+t}}{\varepsilon_{S,p} + \varepsilon_{D,p}} \tag{22}$$

where $\varepsilon_{D,1+t} = \varepsilon_{D,p}\theta$ is the elasticity of aggregate demand with respect to the gross-of-tax price. Equation (22) shows that incidence on producers is attenuated by the factor θ. Since demand is less sensitive to the tax, producers are under less pressure to reduce the pre-tax price p, and consumers bear more of the burden in equilibrium when θ is low:

$$\partial[\frac{dp_t}{dt}]/\partial\theta < 0.$$

In the extreme case where $\theta = 0$, consumers fully bear the tax, because their demand is unchanged by the tax.

One interpretation of the source of this result is that demand effectively becomes more inelastic when individuals are inattentive. Though an increase in θ and an increase in $\varepsilon_{D,p}$ both reduce incidence on demand, the magnitudes of the effects are not equivalent. This is apparent in (22), since θ attenuates only the $\varepsilon_{D,p}$ term in the numerator in the expression for $\frac{dp}{dt}$. As a result, the incidence of a tax in two markets A and B with the same gross-of-tax elasticity $\varepsilon_{D,1+t} = \theta\varepsilon_{D,p}$ but different values of θ and $\varepsilon_{D,p}$ differs. It matters whether the source of a low tax elasticity is inattention or a low price elasticity. Intuitively, when firms reduce their pre-tax prices to share the burden of the tax, demand from the inattentive consumers who focus on the pre-tax price *rises* relative to the no-tax case, since they focus solely on the pre-tax price. Because of this offsetting increase in demand, firms need to reduce p by a smaller amount in order to re-equilibriate the market. A 1% reduction in θ thus raises incidence on demand by *more* than a 1% reduction in $\varepsilon_{D,p}$, even though both changes affect the gross-of-tax elasticity $\varepsilon_{D,1+t}$ equivalently.

Distinguishing the components of $\varepsilon_{D,1+t}$ can be quantitatively important, particularly when the supply elasticity $\varepsilon_{S,p}$ is small. As $\varepsilon_{S,p}$ approaches 0, $\frac{dp}{dt}$ approaches θ, irrespective of $\varepsilon_{D,p}$. Suppliers can shift considerable incidence to consumers even when supply is inelastic if they are in a market where consumers are inattentive. As a quantitative illustration, contrast two markets, A and B, where $\varepsilon^A_{S,p} = \varepsilon^B_{S,p} = 0.1$. In market A, $\varepsilon^A_{D,p} = 0.3$ and $\theta^A = 1$; in market B, $\varepsilon^B_{D,p} = 1$ and $\theta^B = 0.3$. In both markets, $\varepsilon_{D,1+t} = 0.3$ – that is, a change in taxes would be estimated to have a modest effect on demand. However, $[\frac{dp}{dt}]^A = -0.75$ whereas $[\frac{dp}{dt}]^B = -0.27$. In market A, suppliers are forced to bear most of the incidence since demand is 3 times more elastic to price than supply. In market B, even though demand is 10 times as elastic to price than supply, producers are able to shift most of the incidence of the tax to the demand side because only 30 percent of the individuals pay attention to the tax.

An important implication of this point is that the "shortcut" of directly estimating the tax-elasticity of demand $\varepsilon_{D,1+t} = \theta\varepsilon_{D,p}$ and then applying the standard Harberger formula to calculate incidence fails. To calculate incidence when some agents are inattentive, one must estimate *both* the "fundamental" price elasticity $\varepsilon_{D,p}$ and θ – or, equivalently, both $\varepsilon_{D,1+t}$ and $\varepsilon_{D,p}$, as in our empirical analysis – and apply (22).

Endogenous θ. Now consider the case where θ varies with the tax rate. This case can be viewed as a characterization of how incidence may vary across markets (or economies) that have different tax rates in the long run, and thus have different fractions of inattentive individuals. To reduce notation, we focus on the case where utility is quasilinear; the results below hold with arbitrary utility. Recalling that $\theta = F(\frac{1}{2}\varepsilon_{x,p}xt^2)$ in the quasilinear case, observe that

$$\varepsilon_{\theta,t} = \frac{f(\bar{c})}{F(\bar{c})}\varepsilon_{D,p}Dt^2$$

and hence, using (21),

$$\frac{dp}{dt} = -\frac{\varepsilon_{D,p}(\theta + f(\bar{c})\varepsilon_{D,p}Dt^2)}{\varepsilon_{S,p} + \varepsilon_{D,p}} \tag{23}$$

This expression differs from the case where θ is fixed by the term $f(\bar{c})\varepsilon_{D,p}Dt^2 > 0$, which reflects the extent to which θ changes when the tax rate changes. A number of intuitive comparative static implications emerge from this new term. First, $\frac{dp}{dt}$ is more negative when θ is endogenous than when θ is fixed. Producers are forced to reduce prices further when θ is endogenous, because more individuals compute tax-inclusive prices in markets with high tax rates ($\frac{\partial\theta}{\partial t} > 0$). The endogeneity of tax perceptions therefore mitigates the extent to which large taxes are borne by boundedly rational consumers.

Another implication of (23) is that the elasticity of demand $\varepsilon_{D,p}$ has an amplified effect on incidence when θ is endogenous. Consumers are less likely to pay attention to taxes in markets where demand is inelastic. A lower $\varepsilon_{D,p}$ therefore leads to more incidence on consumers both through the traditional Harberger channel and through more inattention. Consumers are also more likely to bear taxes on inexpensive goods (D small), since they are less likely to compute tax-inclusive prices when buying such products.

Discussion. One direct implication of this analysis is that the classic "tax neutrality" result – that the incidence of taxes does not depend on whether the tax is levied on the consumer or producer in a competitive market – does not hold when agents are boundedly rational. If a producer wishes to pass a tax – such as the excise tax on alcohol or a tariff on an intermediate input – through to consumers, he generally must include it in the posted price of the good.[30] But this automatically leads to $\theta = 1$, limiting the ability of the producer to shift the tax by the traditional Harberger formula. In contrast, the producer may face less pressure to reduce pre-tax prices in response to a tax levied on the consumer side – such as the sales tax or after-market vehicle excise fees – since consumers with limited attention may not notice such taxes and thus could bear most of the burden in equilibrium.[31]

[30] In certain markets, producers also have access to technologies – such as add-ons or after-market fees – that allow them to charge customers in a less salient manner. In such an environment, producers may be able to shift a tax onto consumers by changing the prices of these "shrouded attributes" (Gabaix and Laibson 2007).

[31] Interestingly, Busse, Silva-Risso, and Zettlemeyer (2006) report evidence consistent with this prediction. They find that 35% of manufacturer rebates given to car dealers are passed through to the buyer, while 85% of rebates given to buyers stay with the buyer. Their interpretation of this result is that most consumers

The analysis also casts some light on the distributional incidence of commodity taxes that are levied on consumers. Consumers are especially likely to bear the incidence for small staples, such as food or clothing. Insofar as small staples constitute a larger fraction of expenditure for lower income individuals, un-salient commodity taxation (i.e., sales taxes as implemented in the U.S.) may have undesirable distributional effects. More generally, taxes that are not included in posted prices will lead to redistribution from individuals with higher cognitive costs to those with lower cognitive costs. Individuals who optimize relative to taxes will be able to avoid taxes through demand substitution more effectively than those who are inattentive. If higher income individuals have lower cognitive costs – e.g., if education lower costs of cognition – complex or opaque taxes may have undesirable effects on equity. If, in contrast, costs of computation are positively correlated with income – e.g., if the opportunity cost of the time needed to pay attention to taxes is higher for high-income people – opaque taxes may enhance redistribution.

9 Conclusion

The broad objective of this paper has been to incorporate insights from the literature on behavioral economics into public finance to better understand the consequences of tax policies. Our empirical analysis indicates 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, we conclude that tax salience has a substantial impact on behavioral responses to taxation. A simple model of boundedly rational agents can explain our empirical findings as well as other stylized facts. Somewhat surprisingly, small cognitive costs can affect the efficiency consequences of large-scale tax policies. The model yields simple Harberger-type formulas for incidence and efficiency costs of taxation that can be empirically implemented and 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

did not find out about the dealer rebates, but did know (by design) about the consumer rebates.

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 and identify the underlying price elasticities relevant for the analysis of incidence and efficiency. Theoretically, our analysis illustrates that one can make concrete statements about the welfare consequences of tax policies even in situations where agents do not adhere to the neoclassical paradigm. Our approach can be extended and refined in a number of dimensions, to better understand the formation of tax perceptions and evaluate the efficiency costs of taxation in a more general environment. Ultimately, such analysis could shed further light on a wide range of normative 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] Atkinson, A.B., Stiglitz, J.E. (1976). The design of tax structure: direct versus indirect taxation. Journal of Public Economics 6, 55–75.

[5] Auerbach, A. (1985) “The Theory of Excess Burden and Optimal Taxation.” in: A. J. Auerbach & M. Feldstein (eds.), *Handbook of Public Economics* vol. 1, Elsevier Science Publishers B. V. (North-Holland), 67- 127.

[6] 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*.

[7] 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.

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

[9] Becker, G. and C. Mulligan (2003) “Deadweight costs and the size of government.” Journal of Law and Economics, XLVI: 293 – 340.

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

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

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

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

[14] Chetty, R. “A New Method of Estimating Risk Aversion,” *American Economic Review* 96(5): 1821-1834, December 2006

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

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

[17] 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.*

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

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

[20] Diamond, P.A. and D.L. McFadden (1974). "Some uses of the expenditure function in public finance," Journal of Public Economics 3, 3-21.

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

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

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

[24] Feldstein, M. (1995) "The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1986 Tax Reform Act." *Journal of Political Economy,* 103: 551-572.

[25] Finklestein, A. (2007) "E-Z Tax: Tax Salience and Tax Rates." NBER Working Paper No. 12924.

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

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

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

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

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

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

[32] Harberger, A. "The Measurement of Waste," *American Economic Review* 54(3): 58-76.

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

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

[35] Kotlikoff, L. and L. Summers (1987) "Tax Incidence." in: A. J. Auerbach & M. Feldstein (eds.), *Handbook of Public Economics* vol. 2, Elsevier Science Publishers B. V. (North-Holland), 1043-1092.

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

[37] 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.

[38] 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.

[39] Liebman, J. and R. Zeckhauser, "Schmeduling." Harvard KSG Working Paper, 2005.

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

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

[42] 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.*

[43] 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.*

[44] Mohring, H., 1971, Alternative welfare gain and loss measures, Western Economic Journal 9, 349-368.

[45] 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.

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

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

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

[49] 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.

[50] O'Donoghue, T. and M. Rabin (2006) "Optimal Sin Taxes" *Journal of Public Economics*

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

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

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

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

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

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

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

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

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

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

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

Appendix: Proofs

Derivation of Equation (10) for $G(t)$ in General Case

Start from equation the gain in utility from computing the tax-inclusive price:

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

Using the first-order-condition that characterizes the choice of the fully-optimizing agent, $u'(x^*) = (1+t)v'(y^*)$, we obtain

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

Next, totally differentiating the agent's first-order-condition with respect to t yields

$$\begin{aligned}
u''(x^*)\frac{\partial x}{\partial t} &= v'(y^*) + (1+t)v''(y^*)\frac{\partial y}{\partial t} \\
&= v'(y^*) + (1+t)[-(1+t)\frac{\partial x}{\partial t} - x^*]v''(y^*)
\end{aligned}$$

Recognizing that $\frac{\partial y}{\partial t} = -(1+t)\frac{\partial x}{\partial t} - x^*$, it follows that

$$[u''(x^*) + (1+t)^2 v''(y^*)]\frac{\partial x}{\partial t} = v'(y^*) - (1+t)x^* v''(y^*)$$

and hence

$$\widetilde{G}(t) = \frac{1}{2}(\Delta x)^2 \frac{[v'(y^*) - (1+t)x^* v''(y^*)]}{\partial x/\partial t}.$$

To simplify this expression, we use the linear approximation $\Delta x = -\frac{\partial x}{\partial t}t$

$$\widetilde{G}(t) = \frac{1}{2}t^2 \varepsilon_{x,p} x^* v'(y^*) - \frac{1}{2}t^2 \varepsilon_{x,p}(x^*)^2 v''(y^*)$$

where $\varepsilon_{x,p} = \frac{-\partial x}{\partial t}\frac{1}{x^*}$ denotes the price elasticity of x at x^*. Defining $\gamma_y = \frac{-v''(y^*)}{v'(y^*)}y^*$ yields

$$G(t) = \frac{\widetilde{G}(t)}{v('(y)} = \frac{1}{2}t^2 \varepsilon_{x,p} x^*[1 + (\frac{x^*}{y^*})\gamma_y].$$

Derivation of Equation (24) for Excess Burden in Quasilinear Case

To derive an exact measure for the excess burden in the quasilinear case algebraically, note that the agent's indirect utility function when utility is quasilinear is given by

$$V(p,t,Z) = Z - (p+t)x(p,t) + u(x(p,t)$$

and hence the expenditure function is

$$e(p,t,V) = V + (p+t)x(p,t) - u(x(p,t))$$

Substituting these expressions into (13) with $p = 1 + t_0$ and $t = \Delta t$, it follows that

$$EB_{\Delta t} = [u(x(1+t_0,0)) - x(1+t_0,0)] - [u(x(1+t_0,\Delta t)) - x(1+t_0,\Delta t)] \qquad (24)$$

For agents who ignore taxes that are not included in posted prices, $x(1+t_0,0) = x(1+t_0,\Delta t)$ and hence $EB_{\Delta t} = 0$.

For agents who do re-optimize relative to tax-inclusive prices, (24) gives an exact measure of excess burden, which could in principle be calculated by recovering the underlying utility u from choice data, as in Hausman (1981). Using a quadratic approximation to u and letting $x_0 = x(1+t_0,0)$, we obtain

$$EB^*_{\Delta t} \approx t_0 \varepsilon_{x,p} x_0 \Delta t + \frac{1}{2}\varepsilon_{x,p} x_0 (\Delta t)^2$$

Finally, to characterize aggregate deadweight loss in the economy with heterogeneous agents, observe that deadweight loss at a sales tax rate t is given by $\theta(t)EB^*(t) + C(t)$ and hence, following the notation introduced in the text:

$$\begin{aligned} DWL(\theta,\Delta t) &= \theta(t_0^S + \Delta t)EB^*(t_0 + \Delta t) - \theta(t_0^S)EB^*(t_0) + C(t_0^S + \Delta t) - C(t_0^S) \\ &= \theta_0(EB^*(t_0+\Delta t) - EB^*(t_0)) + (\theta_1 - \theta_0)EB^*(t_0+\Delta t) + (\theta_1 - \theta_0)G(t_0^S)) \\ &= \theta_0 EB^*_{\Delta t} + (\theta_1 - \theta_0)(EB^*_{t_0^S+\Delta t} + G(t_0^S)) \end{aligned}$$

Simplifying this expression along the lines indicated in the text yields equation (15).

Derivation of Equation (17) for Excess Burden EB_t^p in General Case

Starting from equation (16), define $x^c = x^c(1,0,v(1,t,Z))$ as the Hicksian demand at the original level of utility when the tax is removed. Using the quadratic approximation $EB_t^* = -\frac{1}{2}t(x^* - x^c)$ yields

$$EB_t^p = -\frac{1}{2}t(x^* - x^c) + [e(1,0,V(1,t,Z))) - e(1,0,V^p(1,t,Z))] - t[x^p - x^*]$$

To calculate the dollar value of computing the tax-inclusive price, use a linear approximation of the expenditure function:

$$\begin{aligned} e(1,0,V(1,t,Z))) - e(1,0,V^p(1,t,Z)) &= \frac{\partial e}{\partial V}[V(1,t,Z) - V^p(1,t,Z)] \\ &= [\frac{\partial V}{\partial e}]^{-1}[V(1,t,Z) - V^p(1,t,Z)] = G(t) \end{aligned}$$

where $\frac{\partial V}{\partial e} = v'(y^*)$ is the marginal utility of wealth, i.e. the effect of an extra dollar of expenditure on utility. Hence

$$EB_t^p = -\frac{1}{2}t(x^* - x^c) + G(t) - t[x^p - x^*]$$

Plugging in the expression for $G(t)$ in (10) yields

$$EB_t^p = -\frac{1}{2}t^2 x^* \varepsilon_{x,p}(1 + \frac{x^*}{y^*}\gamma_y) - \frac{1}{2}t(x^* - x^c) - t[x^p - x^*]$$

Next, use the approximations $\Delta x^c = -\frac{\partial x^c}{\partial p}t$ and $[x^p - x^*] = -\frac{\partial x}{\partial p}t$ to obtain

$$EB_t^p = -\frac{1}{2}t^2 x^* \varepsilon_{x,p}(1 + \frac{x^*}{y^*}\gamma_y) - \frac{1}{2}t^2 x^* \varepsilon_{x,p}^c + t^2 x^* \varepsilon_{x,p}^c \tag{25}$$

The key step in simplifying this expression is to recognize that, following Chetty (2006):

$$\gamma_y = \frac{-y}{z}\frac{\varepsilon_{x,Z}}{\varepsilon_{x,p}^c} \tag{26}$$

To derive (26), implicitly differentiate the agent's first-order-condition for x to calculate $\frac{\partial x}{\partial p}$ and $\frac{\partial x}{\partial Z}$:

$$\begin{aligned} \frac{\partial x}{\partial p} &= \frac{v'(y) - xpv''(y)}{u''(x) + p^2 v''(y)} \\ \frac{\partial x}{\partial z} &= \frac{pv''(y)}{u''(x) + p^2 v''(y)} \end{aligned}$$

Using the Slutsky equation $\frac{\partial x^c}{\partial p} = \frac{\partial x}{\partial p} + x\frac{\partial x}{\partial z}$, we obtain

$$\frac{\partial x/\partial z}{\partial x^c/\partial p} = \frac{pv''(y)}{v'(y)}$$

Finally, defining $\varepsilon_{x,p}^c = -\frac{\partial x^c}{\partial p}\frac{p}{x^c}$ and $\varepsilon_{x,Z} = \frac{\partial x}{\partial Z}\frac{Z}{x}$ and rearranging yields (26). Plugging in the expression for γ_y into (25) and simplifying, we obtain (17).

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	$97.85	$136.05	$93.26	$144.09	$120.81	$165.24	$143.10
per category	(81.9)	(169.9)	(82.7)	(187.9)	(99.1)	(225.3)	(187.1)
Weekly quantity sold	25.08	26.63	24.88	28.45	30.80	32.83	29.01
per category	(24.1)	(38.1)	(24.5)	(41.5)	(29.7)	(51.9)	(42.5)
Number of categories	13	95	13	95	13	95	108
B. Product Level Statistics							
Pre-tax product regular price	4.46	6.26	4.37	6.30	4.64	6.32	6.05
	(1.8)	(4.3)	(1.6)	(4.4)	(1.8)	(4.1)	(4.1)
Pre-tax product regular price	4.27	5.61	4.16	5.58	4.38	5.60	5.45
(weighted by quantity sold)	(4.7)	(3.9)	(1.6)	(3.9)	(1.7)	(3.7)	(3.7)
Weekly quantity sold	1.47	1.82	1.58	1.95	1.63	2.01	1.88
per product (conditional >0)	(0.9)	(1.6)	(1.0)	(1.8)	(1.1)	(2.0)	(1.7)
weekly quantity sold > 0	0.31	0.41	0.37	0.44	0.33	0.45	0.42
(indicator for positive sales)	(0.5)	(0.5)	(0.5)	(0.5)	(0.5)	(0.5)	(0.5)
Average number of products	223	1391	205	1389	245	1548	1670
sold per week	(21)	(51)	(18)	(59)	(20)	(58)	(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)	26.48 (0.22) [5568]	25.17 (0.34) [812]	-1.31 (0.41) [6380]
Experiment (2006: 8- 2006:10)	27.32 (0.86) [288]	23.87 (0.95) [42]	-3.45 (0.60) [330]
Difference over time	0.84 (0.74) [6200]	-1.30 (0.86) [854]	**DD_{TS} = -2.14** (0.64) [6710]
	CONTROL STORES		
Period	Control Categories	Treated Categories	Difference
Baseline (2005:1- 2006:6)	30.57 (0.24) [11136]	27.94 (0.28) [1624]	-2.63 (0.31) [12760]
Experiment (2006: 8- 2006:10)	30.76 (0.71) [576]	28.19 (0.98) [84]	-2.57 (1.03) [660]
Difference over time	0.19 (0.63) [11712]	0.25 (0.86) [1708]	**DD_{CS} = 0.06** (0.90) [13420]
		DDD Estimate	***-2.20*** (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 Posting Tax-Inclusive Prices: 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	**-2.20**	**-2.20**	**-13.10**	**-0.09**	**-0.11**	**-0.132**
	(0.59)***	(0.59)***	(4.88)***	(0.03)**	(0.05)**	(0.089)
Average Price		-7.35	-10.68	-1.40	-0.37	
		(0.08)***	(0.34)***	(0.01)***	(0.01)***	
Average Price Squared		0.14	0.16			
		(0.00)***	(0.01)***			
Implied Price Elasticity		-1.39		-1.40		
Sample size	19,764	19,764	19,764	18,827	18,827	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

Effect of Posting Tax-Inclusive Prices: Robustness Checks

Dependent variable: Quantity Per Category

	Full Sample (1)	Treat. Categories (2)	Treat. Store (3)
Treatment	**-2.27**	**-1.57**	**-2.44**
	(0.60)***	(0.33)***	(0.71)***
Before Treatment	-0.17		
	(1.07)		
After Treatment	0.22		
	(0.78)		
N	21,060	2,379	6,588

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.

Table 6

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

State Beer Excise Tax ($/case)	0.49 (0.47)
Federal Beer Excise Tax ($/case)	0.90 (0.03)
State Excise Tax (Percent)	5.6 (6.3)
General Sales Tax (Percent)	4.3 (1.9)
State Beer Consumption (Gallons)	110,003 (120144)
State Per-Capita Beer Consumption (Cans/Pop.)	243.2 (46.1)
State Drinking Age is 21	0.65 (0.48)
State has Drunk Driving Standard	0.66 (0.47)
Any Alcohol Control Policy Change	0.12 (0.32)
N	1,666

Means; standard deviations in parenthesis. State drinking age is 21 is an indicator for state has raise the legal drinking age to 21; drunk driving standard indicates state has set a threshold blood alcohol content level above which one is automatically guilty of drunk driving; any alcohol control policy change is a dummy variable equal to one in any year where a state has raised the drinking age, implemented a per se drunk driving standard, implemented an administrative license revocation law, or a zero tolerance youth drunk driving law. Sources: Brewer's Almanac 2005; National Institute on Alcohol Abuse and Alcoholism; Univ. of Michigan World Tax Database.

Table 7a

Effect of Excise and Sales Taxes on Beer Consumption

Dependent Variable: Change in Log(per capita beer consumption)

	Baseline (1)	Alc Regulations (2)	Bus Cycle (3)	Full Controls (4)
ΔLog(1+Excise Tax Rate)	**-1.06**	**-1.01**	**-1.03**	**-0.81**
	(0.19)***	(0.19)***	(0.19)***	(0.20)***
ΔLog(1+Sales Tax Rate)	**-0.20**	**-0.12**	**-0.03**	**-0.07**
	(0.30)	(0.30)	(0.30)	(0.30)
ΔLog(Population)	0.05	-0.15	-0.27	-0.26
	(0.06)	(0.08)**	(0.08)***	(0.08)***
ΔLog(Income per Capita)			0.22	0.20
			(0.05)***	(0.05)***
Change in Alcohol Policy		-0.005	-0.005	-0.005
		(0.002)*	(0.003)*	(0.002)*
ΔLog(Unemployment Rate)			-0.01	-0.01
			(0.01)*	(0.01)*
Lag Economic Controls				Yes
Year Dummies	Yes	Yes	Yes	Yes
F-Test for Equality of Tax Variables (Prob>F)	0.02	0.01	0.01	0.04
Observations	1,607	1,607	1,487	1,487

Notes: Dependent variable is the first-differenced log of per capita beer consumption. Change in alcohol policy is a dummy variable indicating that the state implemented an alcohol regulation (higher drinking age, per se drunk driving standards, administrative license revocation laws, or zero tolerance youth drunk driving laws). Sample includes all states except HI and WV.

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

Table 7b
Effect of Excise and Sales Taxes on Beer Consumption

Dependent Variable: Change in Log(per capita beer consumption)

	Include Local Sales Taxes	Region Trends	IV for Excise w/ Policy	3-year differences	Food Exempt
	(1)	(2)	(3)	(4)	(5)
ΔLog(1+Excise Tax Rate)	**-1.08**	**-0.86**	**-0.75**	**-0.96**	**-1.07**
	(0.25)***	(0.20)***	(0.27)***	(0.31)***	(0.26)***
ΔLog(1+Sales Tax Rate)	**0.26**	**-0.06**	**-0.10**	**0.19**	**-0.17**
	(0.33)	(0.30)	(0.30)	(0.35)	(0.30)
ΔLog(Population)	-0.30	-0.28	-0.16	-1.40	-0.16
	(0.09)***	(0.10)***	(0.08)**	(0.39)***	(0.09)*
ΔLog(Income per Capita)	0.19	0.22	0.18	0.09	0.22
	(0.05)***	(0.05)***	(0.04)***	(0.07)	(0.05)***
Year Dummies	Yes	Yes	Yes	Yes	Yes
F-Test for Equality of Tax Variables (Prob>F)	0.00	0.03	0.11	0.02	0.03
Sample Size	1,104	1,607	1,607	1,389	937

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

TABLE 9

Average vs. Marginal Tax Rate: Calibration Results

	α (h share) = 0.5		a (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 and Cobb-Douglas utility with housing share α.
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.

TABLE 10

Individual vs. Social Welfare with Quasilinear Utility

Δt	G	$\Delta R(\theta=0)$	$\Delta R(\theta=1)$	$DWL(\theta=1)$
0.01	0.5	100	74	25.5
0.05	12.5	500	350	137.5
0.1	50	1000	650	300
0.2	200	2000	1100	700
0.3	450	3000	1350	1200

NOTE--These calibrations assume that x_0=10,000, $\varepsilon_{x,p}$=1 and t_0=0.25

APPENDIX TABLE 1

Category Classification in Grocery Store Data

Categories	Group Description	Category Description	Mean Weekly Revenue
Treatment			
5101	Deodorant	Aerosols	82.40
5103	Deodorant	Body Sprays	55.22
5105	Deodorant	Roll-ons	44.12
5110	Deodorant	Clear Solids	323.38
5115	Deodorant	Clear Soft	35.13
5120	Deodorant	Clear	123.48
5125	Deodorant	Visible Sticks	75.57
5245	Hair Care	Accessories	189.47
5501	Cosmetics	Facial	84.20
5505	Cosmetics	Eye	195.00
5510	Cosmetics	Nail	73.38
5515	Cosmetics	Lipstick	48.39
5520	Cosmetics	Accessories	19.37
Control			
5005	Oral Hygiene	At Home Whitening	107.24
5010	Oral Hygiene	Manual Toothbrush	340.57
5012	Oral Hygiene	Power Toothbrush	120.89
5015	Oral Hygiene	Oral Rinse/Mouthwash	314.75
5020	Oral Hygiene	Denture Care	96.82
5025	Oral Hygiene	Dental Floss Products	116.75
5030	Oral Hygiene	Interdental Implements	26.76
5035	Oral Hygiene	Oral Analgesics	115.45
5040	Oral Hygiene	Portable Oral Care	52.84
5201	Hair Care	Professional Daily Hair Care	310.75
5205	Hair Care	Performance Daily Hair Care	983.31
5210	Hair Care	Value Daily Hair Care	290.11
5215	Hair Care	Dandruff Hair Care	116.37
5220	Hair Care	Therapeutic Hair Care	20.54
5225	Hair Care	Hair Growth	12.85
5230	Hair Care	Kids Hair Care	46.75
5235	Hair Care	Hair Color	430.18
5250	Hair Care	African American Hair Care	59.91
5301	Skin Care	Bar Soap	395.65
5305	Skin Care	Liquid Hand Soap	138.95
5308	Skin Care	Liquid Waterless Sanitizer	41.00
5310	Skin Care	Body Wash	339.04
5312	Skin Care	Bath Care	29.82
5314	Skin Care	Image Bath Boutique	36.07
5315	Skin Care	Acne Prevention	140.02
5318	Skin Care	Acne Treatment	12.57
5320	Skin Care	Basic Facial Care	427.17
5322	Skin Care	Anti-aging/Treatments skin care	27.99
5325	Skin Care	Hand & Body Skin Care	312.46
5330	Skin Care	Lip Care	91.97
5335	Skin Care	Cotton	169.72

5340	Skin Care	Depilatories	33.61
5345	Skin Care	Adult Skin Care	172.57
5350	Skin Care	Child/Baby Sun Care	26.06
5401	Shave Needs/Men's Personal Care	Razors	161.13
5405	Shave Needs/Men's Personal Care	Cartridges	389.02
5410	Shave Needs/Men's Personal Care	Disposable Razors	195.95
5415	Shave Needs/Men's Personal Care	Shave Preps	210.23
5420	Shave Needs/Men's Personal Care	Mens Skin Care	14.98
5601	Vitamins and Dietary Supplements	Multiple Vitamins	264.95
5605	Vitamins and Dietary Supplements	Joint Relief	89.57
5610	Vitamins and Dietary Supplements	Calcium	72.59
5615	Vitamins and Dietary Supplements	Letters	120.32
5620	Vitamins and Dietary Supplements	Specialty Supplements	65.91
5625	Vitamins and Dietary Supplements	A/O Minerals	31.65
5630	Vitamins and Dietary Supplements	Herbal Supplements	74.18
5701	Pain Relief	Adult Aspirin	48.23
5703	Pain Relief	Enteric/Antacid/Buffered Aspirin	14.90
5704	Pain Relief	Low Strength Aspirin	62.19
5705	Pain Relief	Adult Acetaminophen	203.24
5710	Pain Relief	Ibuprofen Adult	252.89
5715	Pain Relief	Naproxen Sodium	54.63
5716	Pain Relief	Adult Compunds	86.75
5718	Pain Relief	Specialty Indication Pain	88.92
5725	Pain Relief	Childrens/Infants Analgesics	187.25
5730	Pain Relief	Sleeping Aids	64.99
5735	Pain Relief	Stimulants	14.82
5750	Pain Relief	Nightime Pain Relief	76.19
5760	Pain Relief	External Analgecis	144.08
5799	Pain Relief	GM/HBC Trial Size	66.88
5801	Respiratory	Pediatric Cold/Flu/Cough/Allergy/Sinus	229.73
5805	Respiratory	Adult Cough, Cold, Flu	925.93
5835	Respiratory	Adult Allergy/Sinus	500.74
5840	Respiratory	Nasal Products	269.19
5845	Respiratory	Broncial Asthma	41.45
5850	Respiratory	Cough Drops/Throat Relief	252.64
5855	Respiratory	Thermometers/Covers	37.72
5901	Digestive Health	Acid Neutralizers	243.37
5905	Digestive Health	Acid Combination	17.21
5910	Digestive Health	Acid Blockers	131.62
5915	Digestive Health	Proton Pump Inhibitors (PPI)	92.82
5920	Digestive Health	Multi Sympton Gastro Intestinal Relief	70.60
5925	Digestive Health	Gas Relief	49.46
5930	Digestive Health	Motion Sickness/Anti-Nausea	24.32
5935	Digestive Health	Anti-diarrheal	82.70
5940	Digestive Health	Laxatives	265.29
5945	Digestive Health	Lactose Intolerance	22.14
5950	Digestive Health	Rectal/Hemmorhoidal	58.79
5955	Digestive Health	Pediatric Laxatives	31.57
6001	Eye/Ear Care	Soft Contact Lens Care	155.16
6005	Eye/Ear Care	Rigid Gas Permeable Contact Lens Care	18.55
6010	Eye/Ear Care	General Eye Care	203.62
6040	Eye/Ear Care	Reading Glasses	71.66
6042	Eye/Ear Care	Sunglasses	43.87

6045	Eye/Ear Care	Misc. Eye Glass Accessories	15.28
6050	Eye/Ear Care	Ear Care/Ear Plugs	33.25
6101	Foot Care	Insoles/Inserts	75.90
6105	Foot Care	Corns/Callous/Padding/Bunion/Blister	28.88
6110	Foot Care	Odor/Wetness Control	19.64
6115	Foot Care	Anti-Fungal/Athlete's Foot	107.49
6120	Foot Care	Jock Itch	20.22
6130	Foot Care	Wart Removers	37.76
6190	Foot Care	Grooming and Misc. Foot Care	12.70

NOTE--Weekly revenue statistics based on sales in calendar year 2005.

Figure 1a
Placebo Estimates: Sets of Control Products

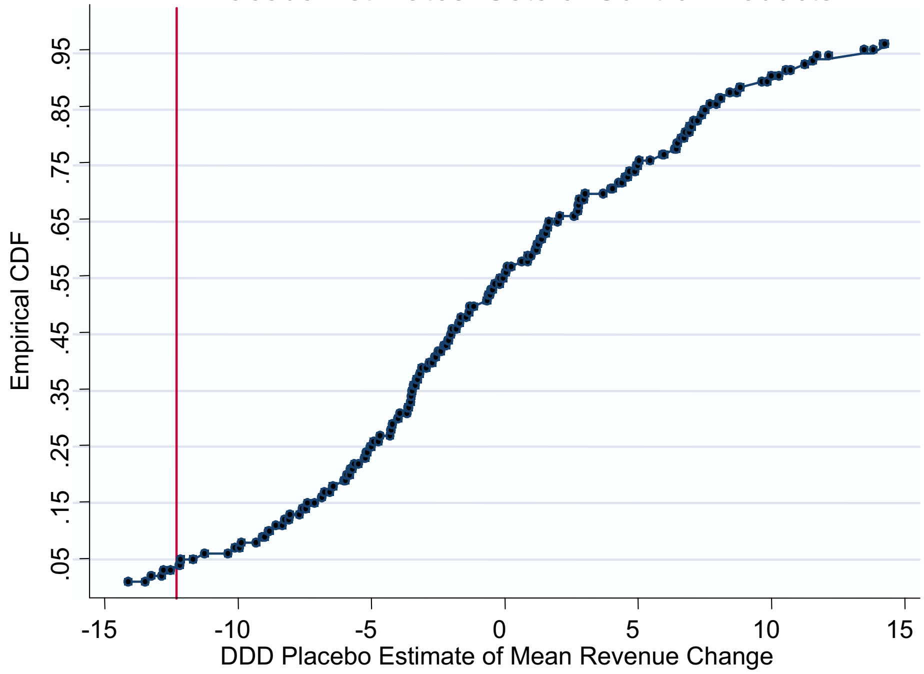


Figure 1b
Placebo Estimates: Other Time Periods

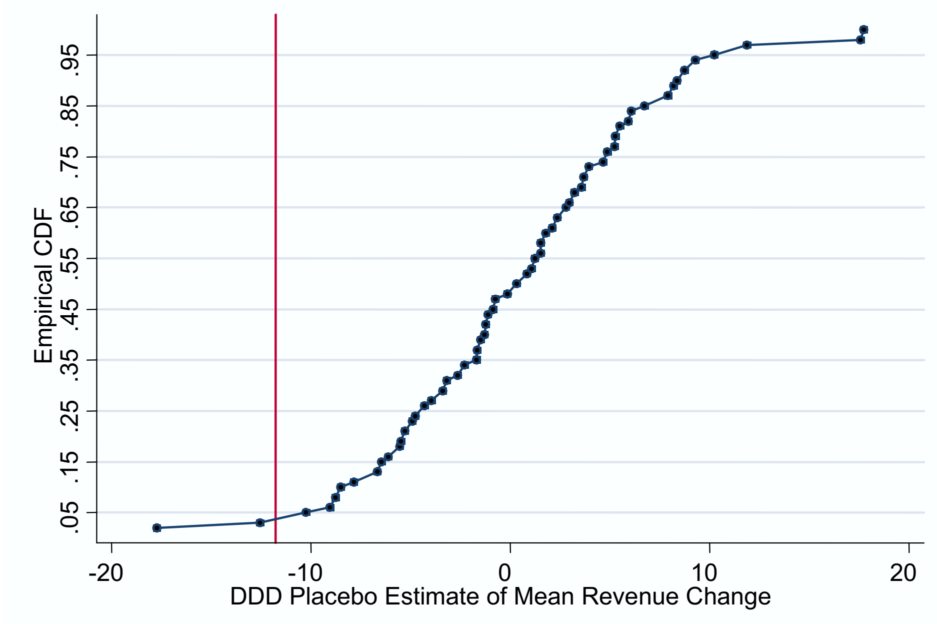


Figure 2a

Per Capita Beer Consumption and State Beer Excise Taxes

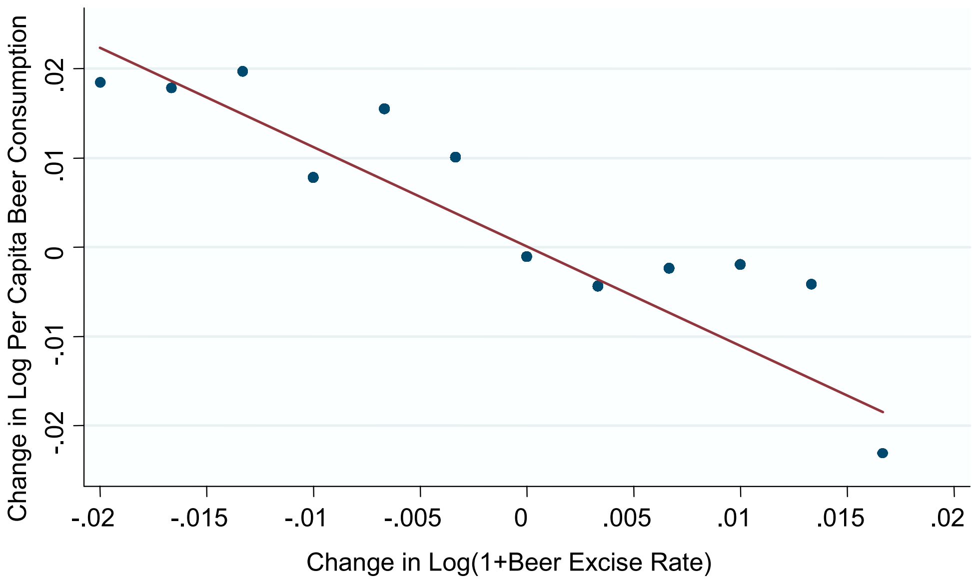
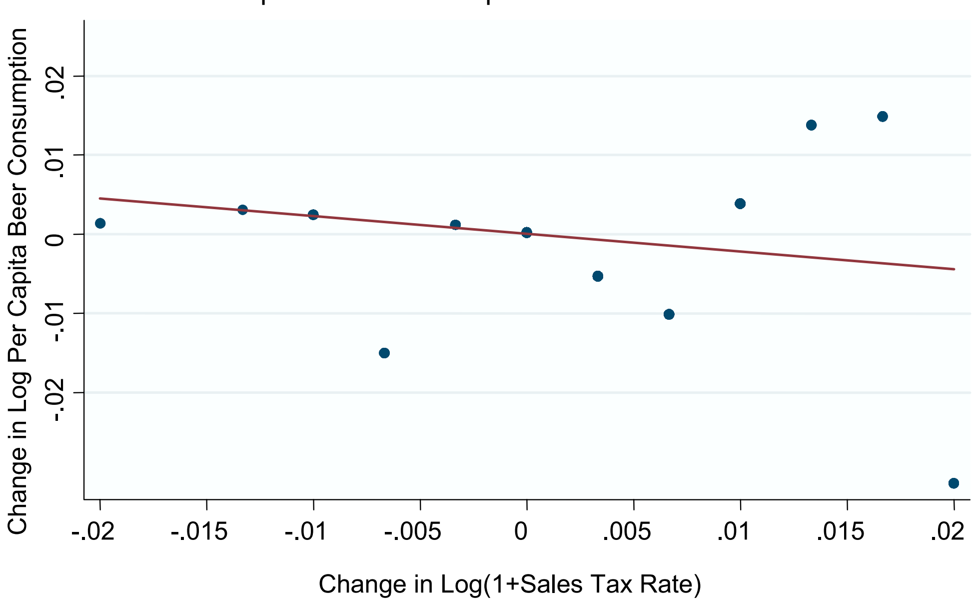


Figure 2b

Per Capita Beer Consumption and State Sales Taxes



NOTE–These figures plot the relationship between log changes in per capita beer consumption and log changes in gross-of-tax-prices ($1 + t^E$ and $1 + t^S$). See text for details on construction of these figures and data sources.

Figure 3a
Number of Correctly Reported Taxed Items on Survey

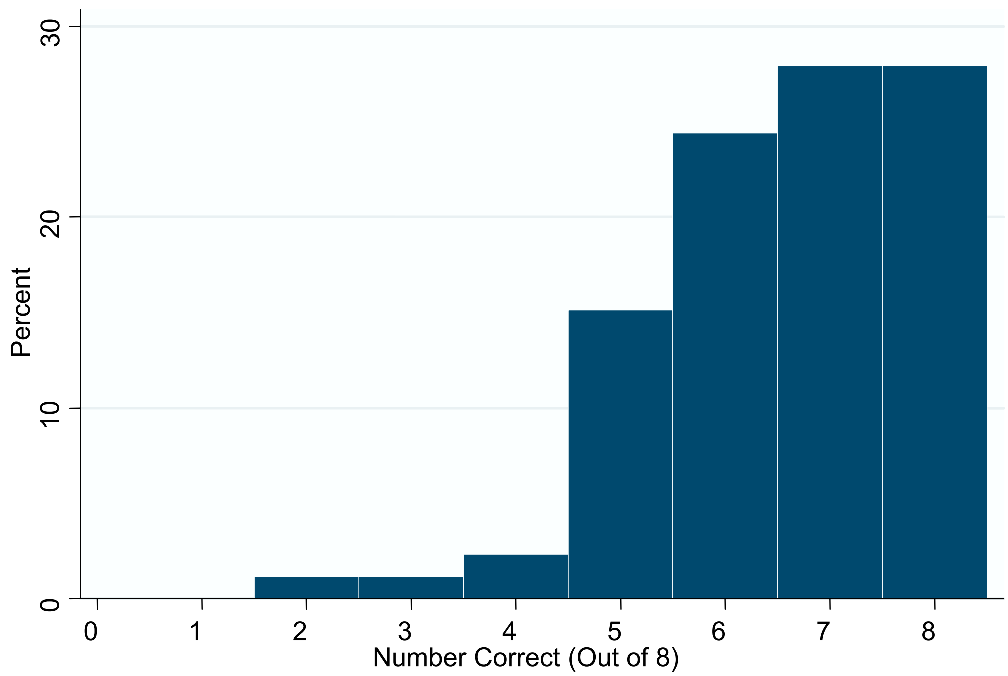


Figure 3b:
Fraction Incorrect by Item on Survey

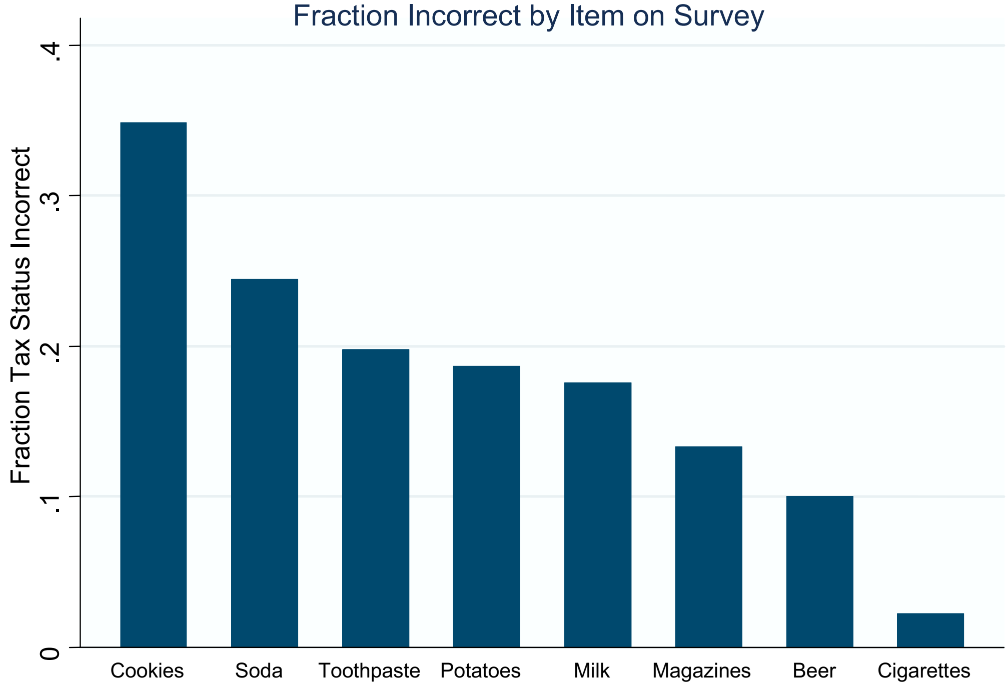
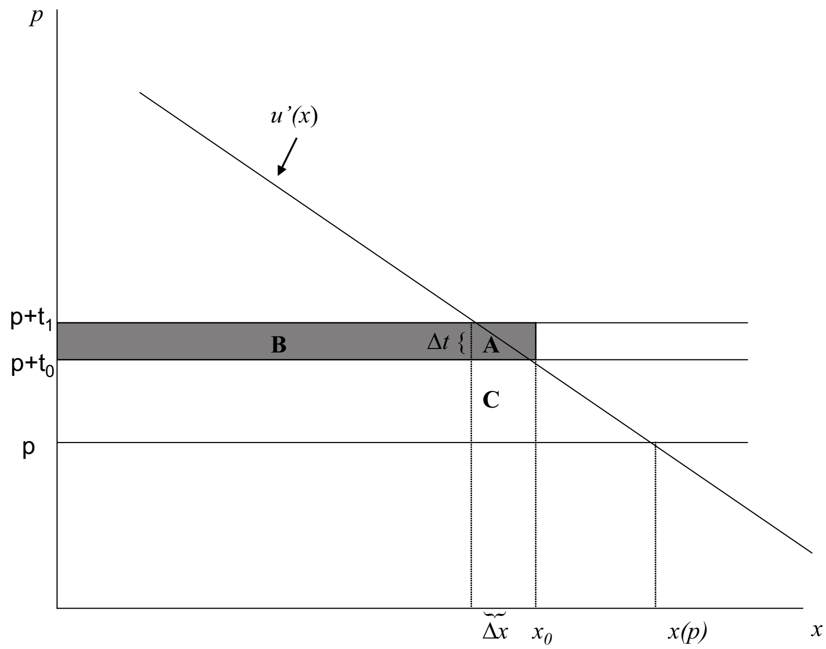


Figure 4

Efficiency Cost of Taxation with Quasilinear Utility



NOTE–This figure shows the effect of a tax change on consumer surplus, tax revenue, and excess burden for an agent with quasilinear utility. Region A (triangle) depicts individual utility loss (G) from failing to reoptimize in response to tax change. Region B (shaded rectangle) depicts change in government revenue ($\Delta R(\theta = 0)$) when the individual does not respond to tax change. Regions A+C (trapezoid) represent excess burden when the individual does respond to tax change ($EB^*_{\Delta t}$).

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!