



Salience and Taxation: Theory and Evidence

Citation

Chetty, Raj, Adam Looney, and Kory Kroft. 2009. Salience and taxation: theory and evidence. *American Economic Review* 99(4): 1145–1177.

Published Version

doi:10.1257/aer.99.4.1145

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:9748525>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

NBER WORKING PAPER SERIES

SALIENCE AND TAXATION:
THEORY AND EVIDENCE

Raj Chetty
Adam Looney
Kory Kroft

Working Paper 13330
<http://www.nber.org/papers/w13330>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2007

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, Shachar Kariv, Peter Katuscak, Botond Koszegi, Erzo Luttmer, James Poterba, Matthew Rabin, Ricardo Reis, Emmanuel Saez, Jesse Shapiro, Andrei Shleifer, Uri Simonsohn, and numerous seminar participants for helpful comments and discussions. Greg Bruich, Matt Grandy, Matt Levy, Ankur Patel, James Sly, and Philippe Wingender 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. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

© 2007 by Raj Chetty, Adam Looney, and Kory Kroft. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Salience and Taxation: Theory and Evidence
Raj Chetty, Adam Looney, and Kory Kroft
NBER Working Paper No. 13330
August 2007
JEL No. D11,H0,J0,K34

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. Using scanner data, we find that posting tax-inclusive prices reduced demand by roughly 8 percent 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 aggregate 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 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 government revenue. Using this theoretical framework, we develop elasticity-based formulas for the efficiency cost and incidence of commodity taxes when agents do not optimize fully.

A revised version of this paper with new theoretical results may be downloaded here:
http://elsa.berkeley.edu/~chetty/papers/taxsalience_aer.pdf

Raj Chetty
Department of Economics
UC- Berkeley
521 Evans Hall #3880
Berkeley, CA 94720
and NBER
chetty@econ.berkeley.edu

Kory Kroft
UC-Berkeley
kroft@econ.berkeley.edu

Adam Looney
Federal Reserve Board
Washington, DC 20551
adam.looney@frb.gov

1 Introduction

A central assumption in public finance is that agents optimize fully with respect to taxes. 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 with respect to such tax policies.

In the empirical component of 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 Specifically, 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). In Gabaix and Laibson’s (2006) terminology, this analysis can be interpreted as a test of whether commodity taxes are “shrouded attributes.” 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 supermarket over a three-week period in early 2006. In this store, prices posted on the shelf exclude sales tax of 7.375 percent. If the good is subject to sales tax, it is added to the bill only at the register, as in most other retail stores in the United States.² Our inter-

¹To 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.

²The 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.

vention 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 on demand using a 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 8 percent during the intervention relative to two “control groups” – other products in the same aisle of the treatment store that were not tagged and products in two other stores in the same chain in nearby cities. The null hypothesis that posting tax-inclusive prices has no effect on demand is rejected using both t-tests and non-parametric permutation tests. To interpret the magnitude of the treatment effect, we compare it with the price elasticity of demand for 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 percent price increase. This finding suggests that the vast majority of customers do not normally 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 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 total alcohol consumption by state, we find that increases in the excise tax reduce alcohol consumption by an order of magnitude more than similar increases in the sales tax. A simple calibration shows that the magnitude of the difference in the elasticity estimates cannot be explained purely by the fact that the sales tax applies to a broader base, especially since food and non-alcoholic beverages are exempt from sales tax in most states. The difference in elasticities persists over time, indicating that sales and excise taxes induce different behavioral responses even in the long run.

There are two potential explanations for the finding that consumers under-react to taxes that are not included in posted prices. 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 when shopping. 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 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 finding motivates the second portion of the paper, which focuses on developing a theoretical model that can explain the importance of tax salience while providing a tractable framework for analyzing 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. Small cognitive costs can lead agents to ignore a large range of taxes because the utility gain from computing the tax-inclusive price is often quite small. For example, when the tax rate is 10 percent, an individual with quasilinear utility gains only \$4 by computing and optimizing with respect to the tax-inclusive price for an item on which he spends \$1,000 and has a price elasticity of 1. Intuitively, since agents are near an optimum to begin with, the gain from reoptimizing relative to the true tax rate is second-order – an application of the envelope theorem, as in Akerlof and Yellen (1985) and Mankiw (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 this evidence, the model also helps explain several stylized facts about behavioral responses to taxation that pose problems for existing models.

Even though *individual* welfare may be minimally affected by ignoring certain taxes, the same taxes can have large impacts on *social* welfare and tax revenue. Consider a 10 percent tax increase that is implemented to raise revenue in a market without externalities. This tax increase raises a significant amount of revenue regardless of whether the agent reoptimizes his behavior, and 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 social welfare, we develop empirically implementable, Harberger-type formulas

for the incidence and efficiency costs of taxation. In our model, the deadweight loss of taxation is determined by two factors beyond the compensated price elasticity of demand: (1) the magnitude of “distortionary income effects” (budget 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 new factors, the efficiency cost and incidence of tax policies can differ substantially from predictions based on existing formulas. For example, a tax increase can have a substantial efficiency cost even when demand for the taxed good does not change by distorting consumption allocations for inattentive individuals. Another implication is that incidence depends on whether the tax is levied on consumers or firms, violating the classic tax neutrality result in competitive markets. More generally, the theoretical analysis indicates that prevailing views about the social welfare consequences of a wide range of tax policies could be affected if bounded-rationality and salience effects are taken into account.

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 cost 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); real vs. nominal prices (Shafir, Diamond, and Tversky 1997); internet price search engines (Ellison and Elli-

son 2004); prices vs. shipping fees (Hossain and Morgan 2005); financial markets (Barber, Odean and Zheng 2005; DellaVigna and Pollet 2005); and the pass-through of manufacturer rebates for car purchases (Busse, Silva-Risso, and Zettlemeyer 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). See DellaVigna (2007) for a detailed review of this literature.

Salience has received less attention in the public finance literature, and as a result its normative implications have not been fully explored. A small set of studies has examined whether individuals understand marginal income tax rates. 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. In contrast, Rosen (1976) finds that the cross-sectional correlation between marginal tax rates and work hours and wage rates and work hours is similar in a survey of married women, suggesting that there is limited “tax illusion.” 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 Feldman and Katuscak (2006) present 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 (2006) 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 using credibly exogenous variation to test whether the salience of commodity taxes affects behavioral responses in the field.

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 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. The seminal work of Akerlof and Yellen (1985) and Mankiw (1985) showed that failing to re-optimize

in response to shocks generates second-order losses for agents, but has first-order effects on the macroeconomy. More recently, Mullainathan (2002), Sims (2003), Reis (2006), and Mackowiak and Weiderholt (2006) develop models of boundedly rational and inattentive consumers, and show that they can explain puzzles in aggregate consumption and pricing dynamics. 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. While previous studies have focused on individual and firm behavior, our theoretical contribution is to explore the implications of bounded rationality and limited attention for social welfare.

In this respect, 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 of the efficiency cost of taxation can be viewed as a special case of Bernheim and Rangel’s approach, where we assume that choices when

taxes are salient reveal an agent's true ranking.

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 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 . The total price of x is $p_t = p(1 + t^S)$. 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 the sales tax. 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 $x^*(p, t) = \left(\frac{p(1+t^S)}{a}\right)^{-1/b}$. The second type is a consumer who is inattentive, and

³In 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.

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

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

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

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

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

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

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. Our goals are to test this hypothesis and 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.

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 p_t on the shelf. When tax-inclusive prices are posted, all individuals presumably optimize relative to the

⁴In practice, there could be other “types” in the population who use different heuristics, e.g. adding 10% to 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 test remains valid: an estimated $\theta < 1$ constitutes a rejection of a model where *all* individuals optimize fully.

tax-inclusive price (i.e., $\theta = 1$). Hence, the effect of posting p_t on demand is

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

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

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

where $\rho = -\frac{\log \hat{x}(p, t, 1) - \log \hat{x}(p, t, \theta)}{\log(1+t)}$ denotes the normalized “tax visibility” effect. The parameter ρ can be interpreted as the (absolute value of the) change in demand caused by making a 1 percent 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 identifies the fraction of individuals who ignore the sales tax. If all consumers normally optimize relative to the sales tax, posting p_t should have no effect on demand ($\rho = 0 \Leftrightarrow \theta = 1$), since it is redundant information. If all consumers ignore the sales tax, posting p_t should reduce demand by $\varepsilon_{x,p} \log(1 + t)$, implying $\rho = \varepsilon_{x,p}$ and $\theta = 0$.

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, as in Rosen (1976). In particular,

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

Under the null hypothesis that salience does not matter, prices and taxes should affect demand equally: $\varepsilon_{x,1+t^S} = \varepsilon_{x,p} \Leftrightarrow \theta = 1$.

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 percent of the products on the shelves are subject to the local sales tax rate of 7.375 percent. 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 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 asked to restrict the intervention to three categories that were not “sales leading” categories, and limit the duration of the intervention to three weeks. The three product groups were chosen in collaboration with the managers based on this requirement and two additional criteria: (1) having relatively high prices, so that the dollar amount of the sales tax is non-trivial; and (2) belonging to what the store terms “impulse purchase categories” – goods that exhibit high price elasticities – so that the demand response to the intervention would be detectable. This led us to run the experiment on three product groups – cosmetics, hair care accessories, and deodorants – over a three week period.⁵

We posted tax-inclusive prices for products in these product group beginning on February 22, 2006 and ending on March 15, 2006. Exhibit 1 shows how the 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

⁵In 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.

tag stated “Total Price: $\$p + \text{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 giving 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 used by the store for the original price.⁶ Additional details on experiment implementation are given in the appendix.

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 treatment categories were in two adjacent aisles, and take up space equivalent to roughly half an aisle in the store. 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 the months prior to the experiment.

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

⁶An important concern with this intervention 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.

4.2 Data and Summary Statistics

We use scanner data from the treatment store and the two control stores, spanning week 1 of 2005 to week 15 of 2006. Price and quantity data for individual products are observed by “promotional week” – weeks beginning on Wednesdays and ending on Tuesdays – in correspondence with the grocery chain’s pricing cycle. Details on the dataset and sample construction are given in the appendix.

Summary measures of store characteristics are displayed in Table 1. The three stores are large (roughly 37,000 sq. feet) and have been open for about 15 years. The cities in which these stores are located are higher income than the U.S. average: the median household income in 2000 was 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 overlap between the treatment and control groups is generally quite good. Within the treatment group, there are 13 product “categories” (e.g. lipsticks, eye cosmetics, roll-on deodorants, body spray deodorants). In these categories, the treatment store sold an average of 25.08 items and earned \$97.85 of revenue per week over the sample period (column 1 of Table 2). The treatment products thus account for approximately \$1,300 of revenue per week as a whole. Average weekly quantity sold per category is similar for the control products, but products in these categories are somewhat more expensive on average (column 2). The first control store is very similar to the treatment store, but the second control store has higher volume in all categories (columns 5-6).

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 aggregate to the category level rather than analyzing the data at the product level for two reasons: (1) the intervention was done at the category level and (2) we cannot distinguish products that were on the shelf but did not sell (true zeros) from products that were not on the shelf (missing data). Analyzing the data at the category level circumvents this problem because there are relatively few category-weeks with missing data (4.7 percent of all observations). Since all the categories always existed in all

stores throughout the sample period, we believe that these observations are true zeros, and code them as such.⁷

4.3 Results

Comparison of Means. We begin our analysis with a simple cross-tabulation 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 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

⁷If we impute quantity as zero for items that sold in both an earlier week and subsequent week, we find that 31 percent of products sell in a given week. We have analyzed data at the product level using this imputation procedure, and obtained results similar to those below.

evolved similarly in the treatment store had the intervention not taken place.

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 7.6 percent. Given the sales tax rate of 7.375 percent, the normalized tax visibility effect is $\rho = 1.03$. As we discuss below, the estimated price elasticity of demand at the category level is $\varepsilon_{x,p} = 1.59$. Since $1 - \theta = \rho / \varepsilon_{x,p}$, the point estimate of θ is 0.35. Note, however, that the hypothesis that $\theta = 0$ is not rejected given the standard error on the estimate of θ . 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 various covariate sets and sample specifications in Tables 4 and 5. Let the outcome of interest (e.g. quantity, log quantity, revenue) be denoted by y . Let the variables TS (treatment store), TC (treatment categories), and TT (treatment time) denote indicators for whether an 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:

$$\begin{aligned}
y = & \alpha + \beta_1 TT + \beta_2 TS + \beta_3 TC + \gamma_1 TT \times TC + \gamma_2 TT \times TS + \gamma_3 TS \times TC \\
& + \delta TT \times TC \times TS + \xi X + \varepsilon
\end{aligned} \tag{3}$$

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. Specification 2 replicates 1, controlling for the mean price of the products in each category using a quadratic specification and including category, week, and store fixed effects.⁸ The estimate on the treatment coefficient is essentially unchanged with the price control, which is unsurprising given that there were no unusual price changes during our intervention period. Specification 3 shows that the intervention led to a significant reduction in revenue (price \times quantity) 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 by the mean revenue over time by category by store, placing greater weight on the larger categories as in the levels regressions. An advantage of the log specification is that it may be a better model for comparisons across categories with different baseline quantities. A disadvantage of the logs specification is that it forces us to omit observations that have zero quantity sold. The logs specifications yield slightly larger estimates than the levels models: a decline in quantity sold of 10.1 percent

⁸The mean price is defined as the average price of the products in each category in the relevant week, weighted by quantity sold over the sample period. The fixed weights eliminate any mechanical relationship between fluctuations in quantity sold and the average price variable. See the appendix for details.

and revenue of 12.3 percent. In the log quantity model, the estimated category-level price elasticity – i.e., the effect of a 1 percent increase in the prices of *all* goods within a category – is $\varepsilon_{d,p} = 1.59$. This elasticity is identified by exploiting the variation in average category-level prices across weeks within the stores. The estimate is consistent with the estimates of Hoch et al. (1995), who estimate a full product-level demand system and obtain category-level price elasticities of 1 to 1.5 for similar products using scanner data from the same grocery chain.⁹

The data suggest that revenue per category fell slightly more than quantity sold per category. 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 \$0.10 (1.9 percent) during the treatment period. One interpretation of this result is that individuals in the market for a more expensive product were more likely to buy nothing at all because the tax levied on more expensive products is larger in dollar terms. Another interpretation is that individuals substituted toward cheaper products within the treatment categories. We cannot distinguish between these alternative hypotheses with the available data.

Placebo Tests and Robustness Checks. As noted by Bertrand et al. (2002), a 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 (*BT*) and the three week period after the intervention ended (*AT*). We also include second- and third-level interactions of *BT* and *AT* with the *TC* and *TS* variables, as for the *TT* variable in (3). Column 1 of Table 5 reports estimates of the third-level interactions (e.g. $BT \times TS \times TC$) 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 changed

⁹This similarity of estimates is reassuring because the simple 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.

by $\delta = -2.27$ units during the intervention. The corresponding “placebo” estimates for the periods before and after the treatment are close to zero.¹⁰ These results indicate that the fall in demand coincides precisely with the intervention period, supporting the identification strategy.

Building on the logic underlying this specification check, we implement a non-parametric permutation test for $\delta = 0$, as suggested by Bertrand et al. (2002). We first choose a “placebo” triplet consisting of a store, three week time period, and a randomly selected set of 13 product categories (excluding the triplet that coincides exactly with the true treatment, i.e. $TT \times TC \times TS = 1$). We then estimate (3), pretending that the placebo triplet is the treatment triplet. Let $\hat{\delta}_P$ denote the estimate of δ from this regression. We repeat this procedure for all permutations of stores and contiguous three week periods and 25 different randomly selected groups of 13 categories, obtaining $63 \times 3 \times 25 = 4,725$ estimates of $\hat{\delta}_P$. The $\{\hat{\delta}_P\}$ values yield an empirical distribution of “placebo effects” in the sample. Let G represent the cdf for this distribution. The statistic $G(\delta)$ gives a p-value for the hypothesis that $\delta = 0$. 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 placebo triplets.¹¹ Since this permutation test does not make parametric assumptions about the error structure, it does not suffer from the overrejection bias of the t-test in the presence of serial correlation.

Figures 1a-b illustrate the results of the permutation test by plotting the empirical distribution of placebo effects G for log quantity and log revenue. The placebo estimates used to construct Figures 1a and 1b are obtained by replicating the specifications in column 4 and 5 of Table 4, respectively. The vertical line in each figure denotes the corresponding treatment effect estimate reported in Table 4. For log quantity, $G(\delta) = 0.07$ and for log revenue, $G(\delta) = 0.04$. Although these p-values are larger than those obtained using the t-tests (presumably because the permutation test has less power), they confirm that the intervention led to an unusually low level of demand. It is also reassuring that the G dis-

¹⁰We discuss why demand returns to pre-experiment levels after the tags were removed in section 6.

¹¹This test can be viewed as an extension of Fisher’s (1922) “exact test” for an association between two binary variables. See Rosenbaum (1986) for more on permutation tests.

tributions are approximately Normal, supporting the parametric assumption underlying the t-tests above. Placebo tests for the other specifications in Table 4 yield similar results.

As an alternative method of probing the robustness of our identification strategy, we consider subsets of the large set of “controls” across time, categories, and stores. In column 2 of Table 5, 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 of these DD estimates are similar to the DDD estimates reported in Tables 3 and 4. Restricting the control time period to the three months immediately before the intervention or limiting the control categories to nearby products also does 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 often above the integer threshold. It is interesting to ask whether demand fell more for the products whose price crossed the integer threshold once taxes were included (e.g. $\$3.99 + \text{Sales Tax} = \4.28). 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 no evidence that demand fell more for the products that crossed the threshold, though the interaction effect is imprecisely estimated.

We also tested whether the intervention in the treatment categories had “spillover” effects onto the nearby control categories by estimating (3) with separate indicators for control categories that were physically adjacent to the treatment categories vs. not adjacent to them. We find no significant difference in demand during the treatment period across the adjacent and more distant control categories within the treatment store, indicating that the effect of the intervention was confined to the products for which tax-inclusive prices were posted.

5 Evidence from Observational Data on Alcohol Sales

5.1 Research Design

We turn now to our second empirical test: comparing the effect of increases in posted prices and taxes on demand. We implement this strategy by focusing on alcohol consumption. Alcohol is subject to two state-level taxes in most states: (1) an *excise* tax that is levied at the wholesale level and is included in the price posted on the shelf or restaurant menu and (2) a *sales* tax, which is added at the register (except in Hawaii, which we exclude). The total price of alcohol is therefore $p_t = p_0(1 + t^E)(1 + t^S)$ where p_0 is the pre-tax price, t^E is the excise tax, and t^S is the sales tax. Since the excise tax creates variation in the posted price, it is more salient than the sales tax.

We compare the effects of state-level changes in the two taxes on alcohol consumption, using an estimating equation based on the demand specification in (1):

$$\log \hat{x}(t^E, t^S, \theta) = \alpha + \beta \log(1 + t^E) + \theta\beta \log(1 + t^S) \quad (4)$$

where $\beta = \frac{\partial \log \hat{x}}{\partial \log 1 + t^E}$ is the gross-of-excise-tax elasticity and $\theta\beta = \frac{\partial \log \hat{x}}{\partial \log 1 + t^S}$ is the gross-of-sales-tax elasticity.¹² We estimate (4) in first-differences because both the tax rates and alcohol consumption are highly autocorrelated series. Letting t index time (years) and j index states, define the difference operator $\Delta z = z_{jt} - z_{j,t-1}$. Introducing a set of other demand-shifters X_{jt} and an error term ε_{jt} to capture idiosyncratic state-specific demand shocks, we obtain the following estimating equation by first-differencing (4):

$$\Delta \log \hat{x}_{jt} = \alpha_0 + \beta \Delta \log(1 + t_{jt}^E) + \theta\beta \Delta \log(1 + t_{jt}^S) + X_{jt}\rho + \varepsilon_{jt} \quad (5)$$

We estimate variants of (5) using OLS, and test the hypothesis that the gross-of-excise-tax and gross-of-sales-tax elasticities are equal, as would be predicted in the absence of salience effects ($\theta = 1$). The identification assumption is that the changes in sales and excise taxes

¹²If tax increases are passed through fully to the consumer ($\frac{\partial p}{\partial t^E} = 1$), β can be interpreted as the price elasticity of demand. Irrespective of the incidence of taxes, the traditional model predicts that the two taxes should affect demand equally ($\theta = 1$).

are uncorrelated with state-specific shocks to alcohol consumption. We assess the validity of this assumption below.

An important simplification made in the test proposed above is the assumption 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.¹³ Hence, a 1 percent increase in t^S changes the relative price of x and y less than a 1 percent increase in t^E . After presenting the baseline estimates, we present some additional evidence and calibrations which indicate 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

Tax rates on alcohol vary across beer, wine, and spirits. In the interest of space, we present results for beer, which accounts for the largest share of alcohol consumption in the U.S. A parallel analysis using data on wine taxes and consumption for a smaller subset of years yields similar results.¹⁴ We use data on aggregate annual beer consumption by state from the National Institute of Alcohol Abuse and Alcoholism (2006) from 1970-2003. These data are compiled from administrative state tax records, and are more precise than comparable data from surveys because they reflect total consumption in each state.

We obtain data on beer excise tax rates and sales tax rates and revenues by state from the Brewer’s Almanac (various years), World Tax Database (2006), and other sources.¹⁵ The state sales tax is an ad valorem tax (proportional to price), while the excise tax is typically

¹³In 2004, sales tax revenues were 2.1% of personal consumption expenditures (PCE). The average (state income-weighted) sales tax rate was 5.3 percent. Hence the tax base is approximately 40 percent of PCE.

¹⁴A potential concern with focusing on one type of alcohol is that tax changes on that type may induce substitution to other types of alcohol. In practice, however, changes in excise tax rates on the three types are highly correlated. For example, the correlation coefficient of changes in beer and wine tax rates is 0.94, and in 85.5 percent of the instances in which a state changes its beer excise tax, it also changes its wine excise tax rate. We show below that there is no evidence of substitution from beer to other types of alcohol when beer excise tax rates are changed.

¹⁵We exclude West Virginia from our analysis because of problems with the sales tax rate series described in the appendix. Including WV magnifies the difference between the excise and sales tax elasticities.

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 in year 2000 dollars by the average cost of a case of beer in the United States in the year 2000.¹⁶ Details on the data sources and construction of tax rates are given in the appendix.

Table 6 lists summary statistics for the pooled dataset. Between 1970 and 2003, mean per capita consumption of beer is 23 gallons per year, equivalent to roughly 240 cans. The average state excise tax rate is 6.4 percent of the average price, which the mean sales tax rate is 4.3 percent. The excise tax rate varies significantly more than the sales tax rate: the standard deviation of excise tax rates is 3 times that of sales tax rates. 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 time trends in alcohol consumption, we identify the elasticities from differential changes in tax rates across states by including year fixed effects in all regressions.

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 annual state-level 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 construct Figure 2a, we first round each state excise tax change to the nearest tenth of a percent (0.1%). We then compute the mean change in log beer consumption for observations with the same rounded excise tax change. Finally, we plot the mean consumption change against the rounded excise tax rates, superimposing a best-fit line on the points as a visual aid. Figure 2b is constructed analogously, rounding sales tax changes to the nearest 0.1%. 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 the tax changes to $\pm .02$ log points. Figure 2a shows that there is a sharp negative relationship between changes in the beer excise tax and beer consumption. Figure 2b shows

¹⁶We normalize by the average price in the nation because each state's price is endogenous to its tax rate.

that the relationship between changes in beer consumption and changes in the sales tax rate is considerably weaker.

To quantify the magnitude of the difference in the excise and sales tax elasticities, Table 7 presents estimates of the model for the state-level growth rate of alcohol consumption in (5). 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 0.87 percent ($\varepsilon_{x,1+t^E} = 0.87$).¹⁷ In contrast, a 1 percent increase in the gross-of-sales-tax price is estimated to reduce beer consumption by 0.20 percent ($\varepsilon_{x,1+t^S} = 0.20$). The null hypothesis that the excise and sales tax elasticities are equal is rejected with $p = 0.05$.

Columns 2-4 evaluate the robustness of these estimates to controls for factors that may be correlated with the tax changes. One concern is that sales tax changes are correlated with the business cycle. In column 2, we control for the state-level business cycle by including state per capita income and the state unemployment rate as covariates. In column 3, we include in addition the lags of these variables. Introducing these controls reduces the estimated sales tax coefficient, and as a result the null hypothesis of equal elasticities is rejected with $p = 0.01$. The sales tax effect is smaller because sales taxes are sometimes raised during budgetary shortfalls that occur in recessions. Since alcohol is a normal good (as indicated by the coefficients on per capita income and 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 rate appears to work *against* rejecting the null hypothesis that $\varepsilon_{x,1+t^E} = \varepsilon_{x,1+t^S}$.¹⁸

Another concern is that excise tax increases are sometimes associated with contemporaneous changes in alcohol regulations, particularly efforts to reduce drunk driving or underage

¹⁷This 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. $\log(t^E)$). For example, Cook, Osterman, and Sloan (2005) find that a \$0.01 increase in the beer tax per ounce of ethanol reduces beer consumption by 1.9 percent. This translates into a gross-of-tax price elasticity of 1.26 at the sample mean.

¹⁸Endogeneity of excise taxes is also likely to work against finding a difference in the elasticities: if policy-makers tend to raise alcohol excise taxes when alcohol consumption is rising, our estimate of $\varepsilon_{x,1+t^E}$ will be biased downward in magnitude.

drinking. We evaluate this concern using data on four measures of regulations: the legal drinking age, the blood alcohol content limit, implementation of stricter drunk driving regulations for youths, and introduction of administrative license revocation laws.¹⁹ To separate the effect of tax changes from changes in regulations, we control for the change in the legal drinking age (in years) and separate indicator variables for a shift toward stricter regulations in each of the other three measures in a state-year pair in column 4. The coefficient on the excise tax rate does not change significantly because regulation changes have modest effects on total beer consumption; on average, beer consumption falls by only 0.5 percent when one of the four regulations is tightened.

Robustness Checks. In Table 8, 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 of Table 8 replicates column 4 of Table 7 incorporating changes in local sales taxes, which are imputed from data on local tax revenues as described in the appendix. In this specification, the excise tax coefficient falls modestly, while the estimate of the sales tax coefficient is positive and statistically insignificant. The hypothesis that the two coefficients are equal is still rejected.

A concern in our identification strategy for the excise tax effect is that trends in tax rates may be correlated with changes in social norms, which directly influence alcohol consumption. For example, rising acceptance of alcohol consumption in historically conservative regions such as the South 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 such trends lead to significant bias, we include region fixed effects in column 2 of Table 8. The coefficient on the excise rate remains substantially larger than the coefficient on the sales tax, suggesting that our results are not spuriously generated by region-specific trends.

As an alternative approach to disentangling time trends from the effect of the tax, we focus on legislated policy changes in excise tax rates. There are two sources of variation identifying the excise tax coefficient. The first is policy changes in the nominal tax rate,

¹⁹We are grateful to Christopher Carpenter, Jeff Miron, and Lina Tetelbaum for sharing this data.

which produce sharp jumps in tax rates. The second is the gradual 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.²⁰ 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 beer at its sample average and compute the implied ad valorem excise tax as the nominal tax divided by this *time-invariant* price. The only variation in this simulated tax rate is due to policy changes. Using the simulated excise tax rate to instrument for the actual excise tax rate, we replicate the specification in column 4 of Table 7. 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 explanation of the difference between the sales and excise tax effects at the annual frequency is learning: people might immediately perceive excise taxes, but learn about changes in the sales tax over time. To test for such learning effects, 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 8 shows the effect of sales and excise tax changes on consumption over a three-year horizon. The estimates show that an increase in the excise tax rate continues to have a large negative effect on alcohol consumption after three years, whereas an equivalent increase in the sales tax does not.

As a supplementary analysis (not reported), we estimated the price elasticity of demand using data on the average price of beer by year by state from the ACCRA survey. We replicated specification 4 of Table 7, instrumenting the average price using the excise tax rate. The estimated price elasticity of demand is 0.9, and the hypothesis that the price elasticity of beer consumption equals the gross-of-sales-tax elasticity is rejected with $p < 0.05$.

²⁰To clarify why inflation generates identifying variation, consider the following example. Suppose the pre-tax price of beer is \$1 and that 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 $\frac{1.50-1.25}{1.50} = 17\%$ in state *A* but is unchanged in state *B*.

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 percent of consumption goods, and therefore leads to a smaller change in the relative price of alcohol than a change in the excise tax. We evaluate the magnitude of the resulting bias in two ways. First, we estimate the model using only the thirty states that fully exempted all food items from the sales tax in 2000.²¹ In these states, changes in the sales tax always affect the relative price of alcohol and food (and non-alcoholic beverages), which is the most plausible substitute for alcohol. Column 5 of Table 8 shows that the sales tax elasticity estimate remains quite small in this subsample, suggesting that the substantial difference between the sales and excise tax elasticities is not purely due to the difference in tax bases.

As an alternative approach, we calibrate the effect of a 1 percent increase in a (hypothetical) tax t^A that applies solely to alcohol (x) and is excluded from the posted price. Treating all goods other than alcohol as a composite commodity (y) of which 40 percent is subject to sales tax, observe that a 1 percent increase in the gross-of-sales-tax price $(1 + t^S)$ increases $\frac{p_x}{p_y}$ by $\frac{1.01}{1.004} - 1 \simeq 0.6$ percent. It follows that the effect of a 1 percent increase in the 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 calibration also indicates that the wider sales tax base is unlikely to explain the large difference between the excise and sales tax elasticities.

A related concern is that increases in the beer excise tax may induce substitution to wine and spirits, thereby biasing the beer tax elasticity up relative to the sales tax elasticity. To assess the extent of substitution, we estimate the effect of the beer excise tax on the share of beer in total alcohol (ethanol) consumption. The estimates in column 6 of Table 8 show that the beer share is insensitive to the beer tax rate. We also find that the effect of beer excise taxes on total ethanol consumption is much larger than the effect of sales taxes. These findings show that substitution between types of alcohol does not explain our results.

Summary. Averaging across the estimates in Tables 7 and 8, the mean estimate of the

²¹Unfortunately, we do not have historical data on which goods are subject to the sales tax. However, case studies of some states suggest that the set of items subject to sales tax is fairly stable over time.

gross-of-excise-tax elasticity is 0.84. The mean estimate of the gross-of-sales-tax elasticity is 0.03. Scaling up the sales tax coefficient by $\frac{5}{3}$, we obtain an implied elasticity of 0.05 for a tax that is applied solely to alcohol at the register. Combining these estimates yields a point estimate of $\theta = \varepsilon_{1,1+t^S} / \varepsilon_{1,1+t^E} = 0.06$. 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.

6 Information vs. Tax Salience: Survey Evidence

The evidence above 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 salient pre-tax price. In this section, we distinguish between these two mechanisms to understand how individual behavior should be modelled.

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 toiletries are 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 analysis, we find that individuals continue to respond less to the sales tax even after they have had considerable time (e.g. 2 or 3 years) to acquire new information.

To test between the information and salience hypotheses more directly, we surveyed 91 customers entering the treatment store in August 2006 about their knowledge of sales taxes.

Survey respondents were offered 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, toothpaste, beer) were subject to sales tax or not. Many 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 asked to mark their best guess to avoid nonresponse bias.²² 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 questions about knowledge of tax rates – the sales tax rate in the city where the store is located, the state income tax rate, 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, while inedible items and ‘sin’ goods are taxed.” Exceptions to this general heuristic led to 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. 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 was also high: 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 percent and 8.75 percent. The modal answer (15 percent of responses) was exactly 7.375 percent. 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 – was not correlated with any of the demographics.

Only 8 percent of individuals answered the estate tax question correctly, 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 closer to federal tax rates. Knowledge of sales taxes may be greater than knowledge

²²We solicited customers by asking them to fill out a survey for “a study being conducted at UC-Berkeley.” We were careful not to mention that the survey concerned taxes, as this could induce selection bias toward those knowledgeable about taxes. None of the individuals who initially agreed to fill out the survey declined to do so after learning that it concerned taxes.

of income or estate tax rates because consumers see the sales tax rate repeatedly (e.g., on receipts), but only see 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

We turn now to constructing a model that can explain the importance of salience in behavioral responses to taxation while providing a tractable framework for welfare analysis. This section sets up the model and characterizes individual behavior. In particular, we identify conditions under which agents ignore taxes, and show that the theoretical predictions match the evidence above as well as other stylized facts. We analyze the model's social welfare implications in the next section.

Setup. Consider an agent with wealth Z choosing consumption of two goods, x and y , over which he has utility $U(x, y) = u(x) + v(y)$. 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.

We depart from the neoclassical model of consumer choice by assuming that the agent pays 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 simply an opportunity cost of time.²³ The agent can alternatively make his consumption decision based on the pre-tax price, which is posted on the shelf and hence costless to compute (i.e., perfectly salient). This binary model of cognition – either computing the tax-inclusive price exactly or not computing it at all – can be viewed as a special case of a model where agents can calculate the tax-inclusive price more accurately by paying a greater computation cost.²⁴ The qualitative results below

²³This model assumes that the agent knows the tax rate t , as appears to be the case for most individuals in practice. In an earlier version, we studied a model where individuals are uncertain about the gross-of-tax price p_t and have a prior over p_t that can be refined by paying a cognitive cost. That model yielded results very similar to those below.

²⁴For example, adding 10% to the posted price is a less expensive approximation than adding 7.375%.

hold in the more general model, but the empirical estimates needed to obtain quantitative predictions differ.

An agent who computes the tax-inclusive price maximizes utility and chooses a consumption bundle $(x^*(t), y^*(t))$ that satisfies

$$\begin{aligned} u'(x^*(t)) &= (1+t)v'(y^*(t)) \\ (1+t)x^*(t) + y^*(t) &= Z \end{aligned} \tag{6}$$

An important conceptual issue in modelling the behavior of agents who do *not* compute tax-inclusive prices is that their consumption choices must still satisfy the true budget constraint $(1+t)x + y = Z$. Thus, one must specify a “budget adjustment rule” – a way in which an agent who uses the pre-tax price chooses x and y to maximize his utility while satisfying his real budget constraint – to close the model. With two goods, there are three intuitive budget adjustment rules, which correspond to variations in the order in which consumption of the goods is chosen. First, the agent can choose x to maximize utility while assuming that he pays no taxes, and spend his remaining wealth on y . This budget adjustment rule leads to a consumption bundle (x_1^p, y_1^p) that satisfies

$$\begin{aligned} u'(x_1^p) &= v'(Z - x_1^p) \\ y_1^p &= Z - (1+t)x_1^p \end{aligned}$$

Alternatively, the agent can choose y first to maximize utility assuming $t = 0$, and spend his remaining wealth on x . This leads to a bundle (x_2^p, y_2^p) that satisfies

$$\begin{aligned} u'(Z - y_2^p) &= v'(y_2^p) \\ x_2^p &= \frac{Z - y_2^p}{1+t} \end{aligned}$$

Figure 4 illustrates these two budget adjustment rules, showing how the initial (infeasible)

Such approximations are likely to be of greater importance in other contexts, such as income taxation. For instance, individuals choosing labor supply may use an approximation such as “one-fourth of income goes to taxes” rather than simply using their pre-tax wage.

choice (x', y') is adjusted by cutting either x or y to arrive at a feasible allocation.²⁵ The figure also suggests a third option, which is to cut consumption of *both* x and y (shifting the budget set inward) to accommodate the income lost to the tax. This leads to the third budget adjustment rule, where the agent chooses x and y simultaneously. Here, the agent ignores the tax on x when making his consumption decision, but recognizes that his net-of-tax income is lower because of the tax. This leads to a bundle (x_3^p, y_3^p) that satisfies

$$\begin{aligned} u'(x_3^p) &= v'(y_3^p) \\ x_3^p + y_3^p &= Z - tx_3^p \end{aligned} \tag{7}$$

The intuition underlying this budget adjustment rule is easiest to see in a dynamic setting where agents purchase x and y repeatedly. If such agents ignore the tax on x , they will later recognize that they do not have as much money as they expected in their bank account, and subsequently cut consumption of *both* x and y . As a result of this dynamic adjustment process, the inattentive agent fails to incorporate the substitution effect of the tax but fully accounts for the income effect, precisely as in (7).

Which of these budget adjustment rules best captures the consumption behavior of an agent who ignores the tax on x is essentially an empirical question.²⁶ The answer to this question can vary across agents and markets. For example, credit-constrained agents may be forced to cut back on consumption of y if they ignore the tax on x , while agents with sufficient liquidity may be able to cut back on both y and future purchases of x . In the interest of space, we focus below on the first rule, where the agent chooses x first and leaves y as residual. Our qualitative results are not sensitive to the assumed budget adjustment rule, with one exception: the efficiency cost of taxation depends on the adjustment rule.

Agent Behavior with Quasilinear Utility. We first characterize agent behavior when utility is quasilinear in y : $U(x, y) = u(x) + y$. Assume $u'(x) > 0$, $u''(x) < 0$, $\lim_{x \rightarrow 0} u'(x) =$

²⁵Thanks to Erzo Luttmer for suggesting this figure.

²⁶One can also attempt to answer the question theoretically by making the choice of the decision rule endogenous. The agent could calculate his utility under each rule, and follow the rule that yields highest utility. In the quasilinear case, Jensen's inequality implies that choosing the good with diminishing marginal utility first is optimal. See Reis (2006) for a related analysis of the choice between a consumption and savings rule in a lifecycle model.

∞ , and $u'(Z) < 1$ to guarantee an interior optimum at any $t \geq 0$.

The agent's problem can be solved in two stages: (1) choose an optimal bundle for any given perceived price p ; (2) decide whether to spend c on computing p_t .²⁷ Our main goal is to characterize the latter decision. We do so by answering 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 c , the agent will not compute p_t . The agent's utility gain from computing the tax-inclusive price is

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

where x^p is the agent's choice of x when he ignores the tax. Note that $G(t)$ is a money metric since utility is linear in y . Taking a second-order Taylor approximation of $u(x)$ around $x^*(t)$ and using the first order condition for $x^*(t)$ gives:

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

Under the first budget adjustment rule, $x^p(t) = x^*(0)$: demand for x coincides with what would be demanded by a fully optimizing agent when $t = 0$.²⁸ We can therefore use the linear approximation $x^*(t) - x^*(0) = t \frac{\partial x^*}{\partial t}$ and the condition $\frac{\partial x^*}{\partial t}(t) = 1/u''(x^*(t))$ to obtain:

$$G(t) \simeq -\frac{1}{2} \frac{\partial x^*}{\partial t} t^2 = \frac{1}{2} \frac{\varepsilon_{x,p}}{1+t} x^*(t) t^2 = \frac{1}{2} \varepsilon_{x,p}^0 x^*(0) t^2 \quad (8)$$

²⁷If 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 of fully optimizing consumption. This is an example of the "regression problem" in bounded rationality models (Conlisk 1996). In this particular case, the agent will presumably solve 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 apply that rule whenever he considers buying those products. In this repeated-decision setting, solving the cognitive problem once is likely to be less expensive than computing the tax-inclusive price each time.

²⁸With quasilinear utility, the first and third adjustment rules result in the same allocation, because the choice of x is independent of total income.

where $\varepsilon_{x,p} = -\frac{\partial x^*}{\partial t} \frac{1+t}{x^*(t)} = -\frac{\partial x^*}{\partial p_t} \frac{1+t}{x^*(t)}$ denotes the price elasticity evaluated at $p_t = 1 + t$ and $\varepsilon_{x,p}^0 = -\frac{\partial x^*}{\partial t} \frac{1}{x^*(0)}$ denotes the price elasticity at $p = 1$.²⁹ The key implication of equation (8) for the analysis below is that $G(t)$ is a second-order function of t .³⁰ This property follows directly from the envelope condition in utility maximization, which makes the first-order (u') terms drop out in $G(t)$. Figure 5a illustrates the result geometrically. The value of computing p_t is given by the lost consumer surplus from overconsumption of x , shown by triangle A. Letting $\Delta x = x^*(t) - x^*(0)$, the size of this triangle is given by $-\frac{1}{2}t\Delta x = -\frac{1}{2}t\frac{\partial x^*}{\partial t}t = G(t)$. As t approaches 0, the size of this triangle diminishes at rate t^2 because both its height and width diminish linearly with t .

It is optimal not to compute p_t if $G(t) < c$, i.e. if $t < T$ where $T = [\frac{-2c}{\partial x^*/\partial t}]^{1/2}$. The threshold T is a bound for the range of taxes which the agent will rationally ignore. Now consider how the threshold T varies relative to the cognitive cost as c approaches 0:

$$\lim_{c \rightarrow 0} \frac{T}{c} = \lim_{c \rightarrow 0} [\frac{2}{\partial x^*/\partial t} \frac{1}{c}]^{1/2} = \infty \quad (9)$$

This result shows that as cognitive costs become small, the range of taxes that are ignored grows small at a slower (square root) rate. As a result, small cognitive costs can lead to substantial inattention to taxes. To quantify what “small” and “substantial” mean in practice, we calibrate the value of computing p_t at various sales tax rates using (8). Consider a commodity on which the agent spends $x_0 = \$10,000$ prior to the implementation of the sales tax, and suppose the price elasticity of demand at that point is $\varepsilon_{x,p}^0 = 1$. The results of the calibration exercise are presented in the first column of Table 9. For example, a 10 percent tax is ignored if the cognitive cost is greater than \$50. Intuitively, the agent gains very little from adjustments in response to small changes in perceived prices because he has already equated the marginal utilities of x and y .

²⁹The point at which the price elasticity is evaluated matters because of the linear demand approximation. The same issue arises in the standard Harberger formula, which is often written in terms of $\varepsilon_{x,p}^0$, but may provide more reliable estimates if written in terms of $\varepsilon_{x,p}$ (Browning 1987). We provide the $\varepsilon_{x,p}$ representation because the formula for $G(t)$ cannot be written in terms of $x^*(0)$ and $\varepsilon_{x,p}^0$ when utility is not quasilinear.

³⁰This claim technically requires that $\frac{\partial x^*}{\partial t}(t)$ is bounded below by some value K as t approaches 0 (i.e., as p_t approaches 1). Under this condition, $f(t)$ behaves like a quadratic function locally around $t = 0$. We assume that this regularity condition holds in deriving results below.

General Utility. Next, we show that the result in (9) holds with arbitrary separable utility. The utility gain from computing p_t is:

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

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, an inattentive agent must be paid approximately $G(t) = \frac{\tilde{G}(t)}{v'(y^*(t))}$ to bring his utility up to the full-optimization level.³¹ Using a second-order Taylor approximation for the utility function as above, we derive the following expression for $G(t)$ in the appendix:

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

where $\gamma_y = -\frac{v''(y^*(t))}{v'(y^*(t))}y^*(t)$ measures the curvature of utility over y . When utility is quasilinear, $\gamma_y = 0$, and $G(t)$ reduces to (8). When $\gamma_y > 0$, the tax has an income effect that makes $|x^*(t) - x^*(0)|$ larger than in the quasilinear case. As a result, $G(t)$ 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). The second term emerges because the agent miscalculates his net-of-tax income when choosing x under the first budget adjustment rule. If the agent follows the third adjustment rule, he does not make this error, and the cost of ignoring the tax is the same as in the quasilinear case (with the compensated elasticity $\varepsilon_{x,p}^c$ replacing $\varepsilon_{x,p}$). Regardless of the budget adjustment rule and utility function, $G(t)$ is proportional to t^2 and hence the main analytical result $\lim_{c \rightarrow 0} \frac{T}{c} = \infty$ holds.

Aggregate Demand with Heterogeneous Agents. To link the model to the empirical evidence, consider an economy populated by a continuum of agents who have identical preferences but heterogeneous cognitive costs, distributed according to a smooth cdf $F(c)$. The

³¹More precisely, one can define the gain from computing p_t using the expenditure and indirect utility functions defined in section 8: $G(t) = [e(1, 0, V^*(1, t, Z))] - e(1, 0, V^p(1, t, Z))$. We show in the appendix that this expression equals (10) under a linear approximation for indirect utility.

fraction of individuals who compute p_t is

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

Intuitively, individuals with c below the threshold $G(t)$ compute p_t , while the rest focus on the salient pre-tax price p . Hence, aggregate demand is given by

$$\widehat{x}(1, t, \theta(t)) = \theta(t)x^*(t) + (1 - \theta(t))x^p(t). \quad (12)$$

Equation (12) shows that the bounded rationality model provides “micro foundations” for the two-type framework in section 3. The aggregate demand function generated by the model matches the evidence in sections 4-6. The calibrations indicate that many agents will not compute tax-inclusive prices despite having the information to do so, consistent with the low estimates of θ and the survey evidence. The model also explains why individuals remain inattentive to taxes in the long run. Since agents must pay a cost *every time* they calculate a tax-inclusive price, tax salience matters in steady-state, and not just on the transition path after a tax change.

Comparative Statics: Attention to Taxes. The fraction of agents who pay attention to a tax is endogenously determined by factors such as the tax rate and elasticity of demand, according to equation (11). The comparative statics of (11) yield the following predictions about when agents are most likely to pay attention and respond to taxes.

(1) $\frac{\partial \theta}{\partial x^*} > 0$. More individuals pay attention to taxes when making large, one time choices (e.g. buying a house) than small, repeated purchases (e.g. food, clothing). This prediction holds even if the small purchases aggregate to the same total expenditure over time as the large purchase, as is the case for food and housing in practice (each accounts for about 20% of the budget on average in the CEX). The reason is that the cognitive cost is a fixed amount paid at each transaction, whereas the benefit of computing p_t scales up with expenditure on the good. In the context of labor income taxation, the model predicts that high income households are more likely to respond to labor and capital income taxes because the stakes are larger for them, consistent with evidence in Feldstein (1995), Goolsbee (2000), and Saez (2004).

(2) $\frac{\partial \theta}{\partial t} > 0$. Individuals are more likely to pay attention to taxes when tax rates are high, because the cost of ignoring the tax grows with the square of the tax rate. By extension, in a dynamic setting, individuals will pay greater attention to large tax reforms than small changes in marginal rates. This could explain why cross-country comparisons (e.g. Prescott 2004, Davis and Henrekson 2006) find larger effects of taxes on behavior than studies that focus on a short window around relatively small tax reforms (e.g. Gruber and Saez 2002).

(3) $\frac{\partial \theta}{\partial \varepsilon_{x,p}} > 0$. Individuals are more likely to compute tax-inclusive prices for products where demand is elastic. If demand is inelastic, there is less benefit from calculating p_t , since one would not change demand much even if the price were higher. This 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.

(4) $\frac{\partial \theta}{\partial \gamma_y} > 0$. Individuals who have more curved utility – e.g. because of credit constraints or consumption commitments – are more likely to compute p_t (or at least net-of-tax income) because they stand to lose more by ignoring taxes.

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 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 policies that induce continuous changes in tax burdens. The cost of ignoring a discontinuous change in the tax burden is a linear (first-order) function of the amount of the tax. Consequently, behavioral responses to “cliffs” in the tax schedule will be large, whereas there will be limited “bunching” at kink points. A related prediction is that agents will respond more on the extensive margin than on the intensive margin to tax policies. Both of these predictions are consistent with empirical evidence (see e.g., Saez 2002).

8 Efficiency Cost and Incidence of Taxation

Auerbach (1985) and Kotlikoff and Summers (1987) present elasticity-based formulas for excess burden and incidence in the traditional model where agents fully perceive taxes. In this section, we develop analogous formulas for the excess burden and incidence of a tax that is not perfectly salient. Throughout, we restrict attention to tax policies implemented to raise revenue (e.g., to finance a public good). Pigouvian taxes intended to correct individual behavior are outside the scope of this paper.

8.1 Efficiency Cost

Definitions. Following Mohring (1971), we define the excess burden of a tax using the concept of equivalent variation. To incorporate salience effects, we first define generalized indirect utility and expenditure functions that permit prices and taxes to have different effects. Normalize the price of y to 1. Let $V(p, t, Z)$ denote the agent's indirect utility net of any cognitive cost as a function of the posted price of x (p), the tax levied at the register (t), and wealth (Z). Let $e(p, t, V)$ denote the agent's expenditure function, which represents the minimum wealth necessary to attain utility V given the posted price and tax, including any cognitive cost. Let $x(p, t, Z)$ denote uncompensated (Marshallian) demand for x and $x_c(p, t, V)$ denote the compensated (Hicksian) demand. Choosing units so that the pre-tax price of x is 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 p and a tax t^S that is not. For concreteness, one could think of t^E as the excise tax and t^S as the sales tax, in analogy with the empirical analysis.

The excess burden of introducing a sales tax t that is not included in the posted price in a previously untaxed market is:

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

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 a situation with pre-existing taxes. Suppose that there are two pre-existing taxes on good x : an excise tax t_0^E and a sales tax t_0^S . Let $t_0 = t_0^E + t_0^S$ denote the total tax. The excess burden of a sales tax increase Δt given initial tax rates t_0^S, t_0^E is

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

where $V_1 = V(1 + t_0^E, t_0^S + \Delta t, Z)$ denotes the agent's indirect utility after the tax increase. Following Auerbach (1985), this expression can be simplified to

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

The last term in (13) reflects the “fiscal externality” (lost tax revenue) that individuals impose on the government if they reduce consumption of x .

Our objective is to derive expressions for (13) in terms of empirically estimable elasticities when individuals behave according to the model in section 7. We begin with a graphical derivation for a case without income effects (quasilinear utility) in the next subsection. We then turn to the general case, where the formulas for excess burden are more complicated because income effects can be distortionary when agents are inattentive.

8.1.1 Quasilinear Utility

When utility is quasilinear in y , the Marshallian and Hicksian demands coincide: $x(p, t, Z) = x_c(p, t, V) \equiv x(p, t) \forall Z, V$. Hence, the excess burden of a tax increase can be characterized using the Marshallian demand curve and the notion of “consumer surplus.” We begin by illustrating the excess burden of a tax increase for two agents, one who does not pay attention to the sales tax and another who does, in Figure 5b. We then characterize aggregate deadweight loss in the economy populated by agents with heterogeneous cognitive costs. For simplicity, the figure assumes $t_0^E > 0$ and $t_0^S = 0$. However, the formulas below hold

with $t_0^S > 0$. An algebraic derivation of the formulas is given in the appendix.

First, consider a consumer whose cognitive cost is sufficiently high that he does not compute the tax-inclusive price even after the sales tax rate rises. We assume throughout this subsection that this inattentive agent meets his budget by cutting only y , deferring the calculation of excess burden under other budget adjustment rules to the general case below. The agent's consumption choice in the initial situation is depicted by x_0^p , the point where $u'(x_0^p) = 1 + t_0^E$. Since the agent does not reoptimize in response to the sales tax increase Δt , the tax increase reduces consumer surplus by $\Delta CS^p = x_0^p \Delta t$, shown in Figure 5b by the rectangle A+B+C. The increase in tax revenue is given by the same rectangle: $\Delta R^p = x_0^p \Delta t$. 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 for an inattentive agent is zero:

$$EB^p(\Delta t | t_0^S, t_0^E) = \Delta CS^p - \Delta R^p = 0.$$

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 . Note that $EB^p > 0$ when agents ignore tax changes if utility over y is concave, as we show below.

Now consider the effects of the same tax increase for an agent with zero cognitive cost ($c = 0$), who always computes p_t . Letting x_0^* denote this 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 is

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

where the second term reflects the loss in revenue from the agent's behavioral response, taking a linear approximation to the demand function as in Harberger (1964). The excess burden of the tax increase for the reoptimizing agent is given by the familiar "Harberger

trapezoid” with area B+D in Figure 5b:

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

where $\varepsilon_{x,p} = -\frac{\partial x^*}{\partial t}\frac{1+t_0}{x_0^*}$ is the price elasticity of demand at $p_t = 1 + t_0$.

Aggregate Deadweight Loss. We now characterize aggregate deadweight loss when the fraction of agents who compute p_t is endogenously determined as in section 7. Let θ_0 denote the fraction of agents who compute p_t when the sales tax rate is t_0^S and θ_1 denote the same after the tax increase. It is useful to divide the economy into 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 $\theta_0 EB^*(\Delta t|t_0^S, t_0^E)$ 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, leading to added deadweight loss of $(\theta_1 - \theta_0)EB^*(t_0^S + \Delta t|t_0^E)$. In addition, these agents now expend the cost of computing p_t , which further contributes to deadweight loss because no cognitive costs are incurred under lump sum taxation. To calculate the increase in deadweight loss, observe that the total cognitive cost incurred in the population from the computation of sales tax at rate t^S is

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

The effect of the tax increase on total cognitive costs can be calculated using a linear approximation to $C(t^S)$ as follows:

$$C(t_0^S + \Delta t) - C(t_0^S) \approx \frac{\partial C}{\partial t^S}(t_0^S)\Delta t = G(t_0^S)\frac{\partial \theta}{\partial t}\Delta t = (\theta_1 - \theta_0)G(t_0^S).$$

Hence the third group of agents contributes $(\theta_1 - \theta_0)\{EB^*(t_0^S + \Delta t|t_0^E) + G(t_0^S)\}$ to aggregate

deadweight loss. Summing the terms, we obtain

$$\begin{aligned}
DWL(\Delta t|t_0^S, t_0^E) &= \theta_0 EB^*(\Delta t|t_0^S, t_0^E) + (\theta_1 - \theta_0)(EB^*(t_0^S + \Delta t|t_0^E) + G(t_0^S)) \\
&= \theta_1 EB^*(\Delta t|t_0^S, t_0^E) + (\theta_1 - \theta_0)(EB^*(t_0^S|t_0^E) + G(t_0^S)) \\
&= \theta_1 [x_0^* \varepsilon_{x,p} t_0 \frac{\Delta t}{1 + t_0} + \frac{1}{2} x_0^* \varepsilon_{x,p} \frac{(\Delta t)^2}{1 + t_0}] + \theta_0 \varepsilon_{\theta,t^S} x_0^* \varepsilon_{x,p} \frac{t_0}{1 + t_0} \Delta t \quad (15)
\end{aligned}$$

where $\varepsilon_{\theta,t^S} = \frac{\partial \theta}{\partial t^S} \frac{t_0^S}{\theta_0}$. Equation (15) gives an empirically implementable formula for deadweight loss in the quasilinear case. It shows that an increase in t^S generates deadweight loss through two margins. 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), leading to a larger 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 a first-order revenue loss for the government, since they now consume $x^* < x^p$. The excess burden from the switchers is proportional to ε_{θ,t^S} , which measures the sensitivity of attention to the tax rate.

Discussion. An important implication of (15) is that taxes which generate small utility losses if ignored by individuals can nevertheless have large effects on social welfare and revenue. Mathematically, the source of this result is that an individual’s utility loss from ignoring a sales tax t ($G(t)$) is proportional to t^2 , whereas government revenue (ΔR) and deadweight loss (DWL) are proportional to t . Unlike individual utility, tax revenue has not been optimized, and therefore rises linearly with the tax rate. Similarly, when there is a pre-existing tax, excess burden rises linearly with t because attentive agents impose a first-order fiscal externality on the government when they reduce consumption of x .

To illustrate the difference between individual and social welfare, we calibrate G , ΔR , and DWL for a range of tax increases in columns 2-6 of Table 9. As above, consider a

good on which the agent initially spends $x_0^* = \$10,000$, and suppose the price elasticity of demand is $\varepsilon_{x,p}^0 = 1$. Suppose the initial excise tax (included in the posted price) is $t_0^E = 25\%$ and the initial sales tax is $t_0^S = 0$. The individual's losses from ignoring sales tax increases are small, as shown in column 2. For example, the utility loss from ignoring a 10% sales tax is \$40 when there is a pre-existing excise tax of 25%.³² To analyze the effects of a tax increase on social welfare, it is useful to contrast the extreme cases where θ is fixed at 0 or 1. When $\theta = 0$, the revenue from a 10% tax increase is $10\% \times 10,000 = \$1,000$. In contrast, when $\theta = 1$, the agent's 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 \$240 – six times the individual's private utility loss from ignoring the tax – because the sales tax amplifies the distortion created by the excise tax. These calibrations show that whether a tax increase is made salient or not can be quite important from a social perspective even when individuals find the issue unimportant.³³

The analysis above appears to imply that “hiding” taxes – to the extent possible given the endogeneity of θ – always reduces deadweight loss. In the next section, we show that this surprising result does not hold when utility is not quasilinear.³⁴

8.1.2 General Case

In the general case, the excess burden of raising the tax rate by Δt for an agent with $c = 0$ (who always computes p_t) is approximately

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

³²The individual's welfare gain from computing p_t is slightly smaller with the pre-existing excise tax because we have assumed that the elasticity equals 1 when the initial price is $p_t = 1.25$ in column 2. Since our approximations assume linear demand, this translates into a price elasticity below 1 at $p_t = 1$.

³³An analogous point applies to transfers. For example, individuals may not pay attention to the tax-benefit linkage in social security because of the small private gains from doing so. However, increasing the salience of the tax-benefit linkage could raise social welfare considerably by raising labor supply and income tax revenue.

³⁴Finkelstein (2007) shows that hidden taxes may also raise efficiency costs due to political economy issues, an important issue outside the scope of our analysis.

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

To calculate excess burden for an 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 an agent with $c = 0$ (who does optimize relative to p_t), some algebra yields

$$EB^p(t) = EB^*(t) + [e(1, 0, V^*(1, t, Z)) - e(1, 0, V^p(1, t, Z))] - t[x^p - x^*(t)] \quad (16)$$

where $EB^*(t)$ is the excess burden of introducing the tax t for a fully-optimizing agent. Equation (16) is an exact expression for the excess burden, which can be computed for any given expenditure function and budget adjustment rule. The properties of $EB^p(t)$ are sensitive to the budget adjustment rule. We therefore characterize EB_i^p under each of the budget adjustment rules $i = 1, 2, 3$ separately, and then provide a formula for aggregate deadweight loss where the fraction of agents using each budget adjustment rule can be estimated empirically.

We begin with the first rule, where the agent meets his budget by cutting consumption of only y . Using a linear approximation to the demand function locally around $x^*(t)$ as in Browning (1987), we prove in the appendix that

$$EB_1^p(t) \approx \frac{1}{2}t^2(1+t)x^*(t)\frac{x^*(t)}{Z}\frac{x^*(t)}{y^*(t)}\varepsilon_{x,Z}\gamma_y \quad (17)$$

where $\varepsilon_{x,Z} = \frac{\partial x^*}{\partial Z} \frac{Z}{x^*(t)}$ denotes the income elasticity of x and $\gamma_y = -\frac{v''(y^*(t))}{v'(y^*(t))}y^*(t)$.

Equation (17) shows that when utility is not quasilinear ($\gamma_y > 0$), a tax increase creates an efficiency loss even if demand for the taxed good *does not change*, contrary to traditional intuition based on the full-optimization model. 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 with the bundle he would choose under lump sum taxation, because the optimal choice of x does not depend on total income. When utility is not quasilinear, the agent optimally reduces consumption of *both* x and y when faced with a lump sum tax. Consequently, when the agent ignores the tax on x and cuts consumption of only y to meet his budget, he ends up overconsuming x relative to the optimum. The lost surplus from this budget allocation error is a function of the income elasticity $\varepsilon_{x,Z}$, which determines the extent of overconsumption of x , and the curvature of the utility function, which determines the utility cost of the resulting underconsumption of y .

To see the intuition more concretely, consider an individual choosing consumption of cars (x) and food (y). Suppose he chooses his car first, overspending because he does not perceive the tax, and therefore has to cut back on food, over which he has very curved utility (high γ_y). In this case, the ignored tax on cars leads to an inefficient allocation of net-of-tax income, which we term a “distortionary income effect.” The resulting efficiency loss could be substantial, potentially greater than the (Harberger) efficiency cost of a perfectly salient tax.

The size of the distortionary income effect depends critically on the budget adjustment rule that the agent uses when ignoring the tax. If the agent chose consumption of food first, he could end up with an allocation closer to what he would have chosen under lump-sum taxation. The excess burden of a tax t on good x if the inattentive agent follows the second budget adjustment rule (choosing y first and cutting consumption of x to meet the budget) is

$$EB_2^p(t) \approx \frac{1}{2} \frac{t^2}{1+t} x^*(t) \frac{x^*(t)}{Z} \frac{y^*(t)}{x^*(t)} \varepsilon_{y,Z} \gamma_x. \quad (18)$$

If curvature of utility over cars (γ_x) is low, $EB_2^p(t)$ could be much smaller than $EB_1^p(t)$.

Finally, if the agent follows the third rule – cutting back on both x and y based on his updated perception of net-of-tax income – the distortionary income effect disappears entirely. Intuitively, such an agent chooses exactly the allocation he would have chosen under lump sum taxation, because he accounts for the change in income but not the distortion in marginal price. Thus $EB_3^p(t) = 0$. The general lesson is that the efficiency cost of a tax that is

ignored by agents is determined by the goods upon which the impact of the tax is borne. If agents distribute the tax cost evenly across goods, excess burden is small. If the tax forces agents to cut back on items where curvature of utility is high, excess burden is large.

Aggregate Deadweight Loss. Now suppose the economy has three types of agents, who differ in the way they adjust their budgets if they ignore the tax on x . Let κ_i denote the fraction of the population that follow budget adjustment rule i . Let $f_i(c)$ denote the pdf of cognitive costs for agents who follow rule i , and $G_i(t)$ denote the gain to computing p_t . The fraction of agents who ignore the tax and follow rule i is given by $\sigma_i(t) = \kappa_i \int_{G_i(t)}^{\infty} f_i(c) dc$. As above, let the fraction of agents who ignore the tax in the aggregate economy be denoted by $1 - \theta(t) = \Sigma \sigma_i(t)$. With this notation, aggregate deadweight loss is

$$DWL(t) \approx \frac{1}{2} t^2 x^*(t) \left[\theta(t) \frac{\varepsilon_{x,p}^c}{1+t} + \frac{x^*(t)}{Z} \left\{ \sigma_1(t) (1+t) \frac{x^*(t)}{y^*(t)} \varepsilon_{x,z} \gamma_y + \sigma_2(t) \frac{1}{1+t} \frac{y^*(t)}{x^*(t)} \varepsilon_{y,z} \gamma_x \right\} \right] + C_g(t) \quad (19)$$

where $C_g(t) = \Sigma_i \kappa_i \int_0^{G_i(t)} c f_i(c) dc$ is the total cognitive cost expended by the three types of agents.

A tractable formula for the deadweight cost of raising the sales tax given a pre-existing sales tax t_0^S can be obtained by differentiating (19), subject to two conditions. First, assume that the optimal demand functions $x^*(t)$ and $y^*(t)$ are locally linear in income and price at t_0^S . Second, assume that the tax increase affects the σ_i s equi-proportionally, i.e. $\varepsilon_{\sigma_i,t} = \varepsilon_{\sigma_j,t} = \varepsilon_{1-\theta,t}$. We show in the appendix that under these two conditions, the marginal deadweight loss of raising the tax rate by Δt starting from an initial sales tax of t_0^S (and excise tax $t_0^E = 0$) is

$$\begin{aligned} DWL(\Delta t | t_0^S) = & \Delta t x_0^* t_0^S \left[\theta(t_0^S) \frac{\varepsilon_{x,p}^c}{1+t_0^S} + \frac{x_0^*}{Z} \left\{ \sigma_1(t_0^S) \frac{x_0^*}{y_0^*} (1+t_0^S) \varepsilon_{x,z} \gamma_y + \sigma_2(t_0^S) \frac{1 + \frac{1}{2} t_0^S}{(1+t_0^S)^2} \frac{y_0^*}{x_0^*} \varepsilon_{y,z} \gamma_x \right\} \right] \\ & - \Delta t \varepsilon_{x,p} (t_0^S)^2 x_0^* \frac{x_0^*}{Z} \left[\sigma_1(t_0^S) \frac{x_0^*}{y_0^*} \varepsilon_{x,z} \gamma_y + \sigma_2(t_0^S) \frac{y_0^*}{x_0^*} \frac{\varepsilon_{y,z}}{(1+t_0^S)^2} \gamma_x \right] \\ & + \Delta t \varepsilon_{\theta,t} \theta(t_0^S) [\bar{x}^p - x_0^*] \end{aligned} \quad (20)$$

where $\bar{x}^p = \Sigma \frac{\sigma_i}{1-\theta} x_i^p$ denotes the average consumption of x among agents who ignore the pre-

existing tax.³⁵ This equation shows that an increase in the sales tax generates deadweight loss through two channels. First, it amplifies distortions for both attentive agents and inattentive agents. The tax increase magnifies the wedge between individuals' choices and socially optimal consumption bundles in proportion to t_0^S , as shown in the first line of (20). For inattentive agents, part of the efficiency loss is offset by the fact that an optimizing agent would spend little on a highly taxed good ($x^*(t)$ becomes small), reducing the scale of the budget allocation error. This offset effect, which corresponds to the second line of (20), is proportional to $(t_0^S)^2$ and hence is relatively small unless the initial tax rate is high. The second channel through which a tax increase raises deadweight loss is by making more individuals compute the tax-inclusive price. This effect, which corresponds to the third line of (20), strictly raises deadweight loss despite reducing excess burden from the distortionary income effect because the added cognitive costs offset the benefit of computing p_t for the switchers. The net result is that switchers impose a first-order revenue loss on the government by reducing consumption of x by $[\bar{x}^p - x^*(t_0^S)]$ on average.

Although (20) has many terms, it is at least in principle empirically implementable. The elasticities can be estimated by examining behavioral responses to tax changes and other policies, as in conventional empirical studies. To see how the σ_i parameters can be estimated empirically, first note that the fraction θ can be estimated by either of the two methods proposed in section 3. The fractions $\{\frac{\sigma_i}{1-\theta}\}$ can then be estimated by comparing how a \$1 change in income (Z) affects the consumption allocation of individuals who ignore the sales tax relative to a \$1 change in net-of-tax income due to the sales tax. The parameter $\frac{\sigma_3}{1-\theta}$ is given by the fraction of agents whose consumption bundles coincide under the two scenarios. The remaining fractions can be calculated by examining which goods individuals cut (x or y in the two-good case) to meet their budget.

As a simple illustration, we calculate the marginal efficiency cost of increasing the sales tax, which applies to $\frac{x}{Z} = 0.4$ of consumption on average in the U.S. Suppose all agents ignore sales taxes when making consumption decisions, and absorb its impact entirely on untaxed goods ($\sigma_1 = 1$). If the income and price elasticity of food consumption are 1

³⁵When $\sigma_1 = 1$, $\gamma_y = 0$, and $t_0^E = 0$, equation (20) reduces to a first-order approximation (omitting the Δt^2 terms) of equation (15) derived in the quasi-linear case.

($\varepsilon_{x,p} = \varepsilon_{x,Z} = 1$) and the coefficient of relative risk aversion over the remaining consumption bundle is $\gamma_y = 1$, then $DWL(\Delta t|t_0^S) = x\Delta t \frac{x}{Z} \frac{x}{y} t_0^S$. Note that the revenue collected by raising the sales tax by Δt is approximately $R^p(\Delta t) = x\Delta t$. Hence, under these parameters, raising the sales tax starting from the current rate of approximately 10 percent would generate an efficiency loss of approximately $0.4 \times \frac{2}{3} \times 0.10 = 2.7\%$ percent of revenue raised.

Discussion. Although we motivated the efficiency analysis by a bounded rationality model, the expressions for excess burden for agents who do not compute p_t ($EB_i^p(t)$) do not rely on this particular model of individual 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. One only needs 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 the expressions for $EB_i^p(t)$ above without further assumptions about why salience matters. Hence, in the derivations of $EB_i^p(t)$, the bounded rationality theory can be viewed as a justification for why the choices made when taxes are salient reveal true preferences. The bounded rationality model is, however, needed to obtain the expressions for aggregate deadweight loss in (19) and (20), since calculating the fraction of agents who pay attention to the tax requires a positive theory of tax perceptions.

Our stylized model 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 seems to imply that the efficiency cost of a “hidden” tax will always be negligible in magnitude. While this logic is correct in our stylized model, we believe that deadweight loss from distortionary income effects could potentially be large in practice, for two reasons. First, many individuals are likely to be uncertain about the benefits of optimizing relative to various tax policies. When faced with uncertainty, boundedly rational agents may not pay attention to aspects of the tax code that have large financial consequences (e.g. tax credits, Roth vs. Traditional IRAs) because the cost of optimizing relative to all policies outweighs the expected benefit from doing so. As a result, a tax that is not salient could produce large

budget allocation errors and lead to a substantial efficiency loss. Second, individuals often make repeated small purchases that aggregate to a large fraction of total expenditure over time. A boundedly rational agent may ignore the tax because the value of computing p_t for each transaction is small; however, the total welfare loss over time from the resulting budget allocation errors could be large. These points suggest that incorporating uncertainty and dynamics into the analysis would be a useful direction for future work.

8.2 Incidence

How is the burden of a tax shared between consumers and producers in competitive equilibrium when some individuals are inattentive to taxes? To answer this question, we analyze the incidence of a tax t (which is not included in the posted price) in a partial equilibrium model. Let $D(p_t) = D(p + t)$ denote demand for individuals who compute p_t and $D(p)$ denote demand by inattentive individuals.³⁶ The supply curve $S(p)$ is a function solely of the pre-tax price. The market-clearing pre-tax price p equates supply and demand:

$$\theta(t)D(p + t) + (1 - \theta(t))D(p) = S(p) \quad (21)$$

Our objective is to calculate $\frac{dp}{dt}$ and $\frac{dp_t}{dt}$, the incidence of a tax on producers and consumers, respectively. Implicitly differentiating (21), we obtain:

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

Using a linear approximation to the demand function and collecting terms yields

$$\begin{aligned} \frac{dp}{dt} &= - \frac{\varepsilon_{D,p} \theta (1 + \varepsilon_{\theta,t})}{\varepsilon_{S,p} + \varepsilon_{D,p}} \\ \frac{dp_t}{dt} &= 1 + \frac{dp}{dt} = \frac{\varepsilon_{S,p} + \varepsilon_{D,p} [1 - \theta (1 + \varepsilon_{\theta,t})]}{\varepsilon_{S,p} + \varepsilon_{D,p}} \end{aligned} \quad (22)$$

³⁶The analysis below neglects any change in demand by inattentive agents that arises from an income effect of the tax on demand for x , as would occur under the second or third budget adjustment rules. Reduction in demand due to such income effects would shift a greater burden of the tax to producers.

where $\varepsilon_{D,p} = -\frac{\partial D}{\partial p} \frac{p}{D(p)}$ is the price elasticity of demand at the pre-tax price, $\varepsilon_{S,p} = \frac{\partial S}{\partial p} \frac{p}{S}$ 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. When $\theta = 1$, (22) reduces to the traditional formula for incidence.

Fixed θ . To characterize the effects of inattention on tax incidence, we begin with the case where the fraction of inattentive individuals is exogenously fixed at some positive level ($\theta < 1$). When $\varepsilon_{\theta,t} = 0$, the incidence of the tax on supply reduces to

$$\frac{dp}{dt} = -\frac{\varepsilon_{D,p}\theta}{\varepsilon_{S,p} + \varepsilon_{D,p}} = -\frac{\varepsilon_{D,1+t}}{\varepsilon_{S,p} + \varepsilon_{D,p}} \quad (23)$$

where $\varepsilon_{D,1+t} = \varepsilon_{D,p}\theta$ is the elasticity of aggregate demand with respect to the gross-of-tax price. Equation (23) shows that incidence on producers is attenuated by the factor θ relative to the traditional model without salience effects. Since demand is less sensitive to the tax, producers are under less pressure to reduce the pre-tax price, and consumers bear more of the burden in equilibrium when θ is low ($\partial[\frac{dp}{dt}]/\partial\theta < 0$). In the extreme case where $\theta = 0$, consumers bear all of the tax, because their demand is unchanged by the tax.

One interpretation of this result is that demand effectively becomes more inelastic when individuals are inattentive, leading to greater incidence on consumers. Though reductions in θ and $\varepsilon_{D,p}$ both increase incidence on demand, the magnitudes of the two effects are not equivalent. This is apparent in (23), since θ attenuates only the $\varepsilon_{D,p}$ term in the numerator in the expression for $\frac{dp}{dt}$. A 1 percent reduction in θ raises incidence on demand by *more* than a 1 percent reduction in $\varepsilon_{D,p}$, even though both changes affect the gross-of-tax elasticity $\varepsilon_{D,1+t}$ equivalently. 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-equilibrate the market when inattention is prevalent.

Distinguishing the two 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 $\varepsilon_{S,p}^A = \varepsilon_{S,p}^B = 0.1$. In market A , $\varepsilon_{D,p}^A = 0.3$ and $\theta^A = 1$; in market B , $\varepsilon_{D,p}^B = 1$ and $\theta^B = 0.3$. In both markets, $\varepsilon_{D,1+t} = 0.3$. 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 price elastic as supply, producers are able to shift most of the incidence of the tax because only 30 percent of the individuals compute p_t .

This example illustrates that the “shortcut” of estimating the tax-elasticity of demand $\varepsilon_{D,1+t} = \theta\varepsilon_{D,p}$ and applying the standard 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 (23).

Endogenous θ . Now consider the case where θ is endogenously determined according to (11). Letting $m = [1 + (\frac{x}{y})\gamma_y]$ denote the amplification factor for $G(t)$ when utility is not quasi-linear, note that $\theta = F(\frac{1}{2}\varepsilon_{D,p}D(p)mt^2)$ where $D(p)$ is demand at the pre-tax price and $\varepsilon_{D,p}$ is the price elasticity of demand at that point. Plugging in this definition of θ and $\varepsilon_{\theta,t} = \frac{f(\bar{c})}{F(\bar{c})}\varepsilon_{D,p}D(p)mt^2$ into (22), it follows that

$$\frac{dp}{dt} = -\frac{\varepsilon_{D,p}(F(\frac{1}{2}\varepsilon_{D,p}D(p)mt^2) + f(\bar{c})\varepsilon_{D,p}D(p)mt^2)}{\varepsilon_{S,p} + \varepsilon_{D,p}} \quad (24)$$

This equation shows that the endogeneity of θ has several additional implications for tax incidence. First, $\frac{dp}{dt}$ is more negative when θ varies with t than when θ is fixed exogenously. Producers are forced to reduce prices further when θ is endogenous, because more individuals notice taxes when tax rates are high. The endogeneity of attention therefore mitigates the extent to which large taxes are borne by consumers. Second, (24) shows 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 increased inattention. Third, consumers are more likely to bear taxes on inexpensive goods (D small) and goods subject to low taxes (t small), since they are less likely to compute tax-inclusive prices when buying such products.

Discussion. 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 inattentive.³⁷ If a producer wishes to pass a tax levied on the business (such as the excise tax on alcohol or a tariff on an intermediate input) to consumers, he generally must include it in the posted price of the good, limiting his ability to shift the tax to the consumer.³⁸ In contrast, many consumers may not pay attention to taxes levied on the demand side (such as the sales tax or vehicle excise fees) and thus could bear most of the burden of those taxes.

The model can also cast light on the distributional incidence of taxes if we allow for heterogeneity beyond cognitive costs. For example, consumers bear more of the incidence of taxes on inelastically demanded items that are purchased in small quantities, such as food or clothing. Insofar as such goods constitute a larger fraction of expenditure for lower income individuals, commodity taxes that are not salient (i.e., sales taxes as implemented in the U.S.) may have undesirable distributional effects. Another example arises from the fact that individuals with lower cognitive costs are more likely to substitute toward behaviors that reduce their tax burdens when tax policies are not salient. If higher income individuals have lower cognitive costs – e.g., if education lowers costs of cognition – opaque or complex tax and transfer policies could have undesirable effects on equity.

9 Conclusion

In this paper, we showed empirically that behavioral responses to commodity taxes differ significantly depending on whether taxes are included in posted prices. Individuals are well informed about commodity taxes when their attention is drawn to the topic, suggesting that *salience* is an important determinant of behavioral responses to taxation. We showed

³⁷Interestingly, 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 finding is that most consumers did not find out about the dealer rebates, but did know (by design) about the consumer rebates.

³⁸Producers may 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 2006).

that introducing small cognitive costs of computing tax-inclusive prices into the neoclassical model of consumer choice can explain our empirical findings as well as other stylized facts. These small cognitive costs can have substantial impact on the welfare consequences of tax policies. To characterize these welfare consequences, we developed Harberger-type formulas for the incidence and efficiency cost of taxation that can be easily adapted to other applications, such as income and capital taxation. Much as Harberger’s analysis identified the compensated elasticity as a key parameter to be estimated in subsequent work, the present analysis indicates that estimating tax perceptions $(\theta, \varepsilon_{\theta,t})$ and the magnitude of distortionary income effects due to inattention $(\varepsilon_{x,Z}, \gamma_y)$ would improve our understanding of the efficiency costs and incidence of taxation.

The finding that individuals optimize imperfectly even with respect to relatively simple, linear commodity taxes suggests that similar issues may arise in the analysis of a broad set of policies. In future work, 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. Such an analysis could shed light on which tax reforms are most likely to identify the fundamental price elasticities relevant for the analysis of incidence and efficiency. More generally, the willingness-to-pay approach proposed here can be applied to understand the heuristics agents use to simplify the tax code. For example, in ongoing work, we find that individuals are likely to value information about their average income tax rates more than information about marginal rates under plausible conditions. This could potentially explain why individuals are more aware of and responsive to average rates than marginal rates (Liebman and Zeckhauser 2004).

Our theoretical analysis should be viewed as a first step toward analyzing the welfare consequences of government policies when agents do not adhere to the neoclassical paradigm. The approach we have proposed is to model the heuristics that agents use to “filter” a tax policy before attempting to determine the effects of the policy on behavior and welfare. Building on this approach could ultimately shed light on a wide range of normative issues that have received attention in recent policy debates, such as consumption taxation (where taxes may be included in posted prices) and the value of tax simplification.

References

- [1] Akerlof, G. and J. Yellen (1985) "Can Small Deviations from Rationality Make Significant Differences to Economic Equilibria," *American Economic Review*, 74(4): 708-720.
- [2] Amador, M., I. Werning, and G. Angeletos (2006) "Commitment vs. Flexibility," *Econometrica*, 74(2): 365-396.
- [3] 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.
- [4] Atkinson, A.B. and J.E. Stiglitz (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*, 78: 2095-2102.
- [7] Bernheim, D. and A. Rangel (2007) "Beyond Revealed Preference: Toward Choice-Theoretic Foundations for Behavioral Welfare Economics," *Cal-Tech Working Paper*.
- [8] Bertrand, M., E. Duflo, and S. Mullainathan (2002) "How Much Should We Trust Differences in Differences Estimates," NBER Working Paper 8841.
- [9] Beer Institute (various years), *Brewer's Almanac*. Available online at <http://www.beerinstitute.org/statistics.asp?bid=200>
- [10] Brown, C.V. (1968) "Misconceptions About Income Tax and Incentives," *Scottish Journal of Political Economy*, 1-12.
- [11] Browning, E.K. (1987) "On The Marginal Welfare Cost of Taxation," *American Economic Review* 77: 11-23.
- [12] Busse, M., Silva-Risso, J. and F. Zettlemeyer (2006) "\$1000 Cash Back: The Pass-Through of Auto Manufacturer Promotions," *American Economic Review*, 96(4): 1253-1270.
- [13] Chetty, R. (2006) "A New Method of Estimating Risk Aversion," *American Economic Review* 96(5): 1821-1834.
- [14] Commerce Clearing House (various years), *State Tax Handbook*, Chicago: Commerce Clearing House Inc.
- [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] Cook, P.J., J. Osterman and F.A. Sloan (2005) "Are Alcohol Excise Taxes Good for us? Short and Long Term Effects on Mortality Rates," NBER Working Paper No. 11138.
- [18] Davis, S. and M. Henrekson (2006) "Economic Performance and Work Activity in Sweden after the Crisis of the Early 1990s," NBER Working Paper No. 12768.
- [19] de Bartolome, C. (1995) "Which Tax Rate Do People Use: Average or Marginal?," *Journal of Public Economics*, 56: 79-96.
- [20] DellaVigna, S. (2007) "Psychology and Economics: Evidence from the Field," UC-Berkeley Working Paper.
- [21] DellaVigna, S. and J. Pollet (2005) "Attention, Demographics, and the Stock Market," NBER Working Paper 11211.
- [22] Ellison, G. and S. Ellison (2004) "Search, Obfuscation, and Price Elasticities on the Internet," NBER Working Paper No. 10570.
- [23] Feldman, N. and P. Katuscak (2006) "Should the Average Tax Rate be Marginalized?," CERGE Working Paper No. 304.
- [24] Feldstein, M. (1985) "The Optimal Level of Social Security Benefits," *Quarterly Journal of Economics*, 100(2): 303-20.
- [25] 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.
- [26] Finklestein, A. (2007) "E-Z Tax: Tax Salience and Tax Rates," NBER Working Paper No. 12924.
- [27] Fisher, R.A. (1922). "On the interpretation of χ^2 from contingency tables, and the calculation of P". *Journal of the Royal Statistical Society* 85(1):87-94.
- [28] Fujii, E. and C. Hawley (1988) "On the Accuracy of Tax Perceptions," *The Review of Economics and Statistics*, 70(2): 344-347.
- [29] Gabaix, X. and D. Laibson (2006) "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets," *Quarterly Journal of Economics*, 121(2): 505-540.
- [30] Gabaix, X., D. Laibson, G. Moloche and S. Weinberg (2006) "Information Acquisition: Experimental Analysis of a Boundedly Rational Model," *American Economic Review*, 96(4): 1043-1068.
- [31] Goolsbee, A. (2000) "What Happens When You Tax the Rich? Evidence from Executive Compensation," *Journal of Political Economy*, 108(2): 352-378.

- [32] Gourville, J.T. (1998) "Pennies-a-Day: The Effect of Temporal Reframing on Transaction Evaluation," *Journal of Consumer Research*, 24(4): 395-408.
- [33] Gruber, J. (1994) The Incidence of Mandated Maternity Benefits," *American Economic Review* 84(3): 622-641.
- [34] Gruber, J. and B. Koszegi (2001) "Is Addiction "Rational"? Theory and Evidence," *Quarterly Journal of Economics*, 116 (4): 1261-1305.
- [35] Gruber, J. and E. Saez (2002) "The Elasticity of Taxable Income: Evidence and Implications," *Journal of Public Economics* 84: 1-32.
- [36] Harberger, A. (1964) "The Measurement of Waste," *American Economic Review*, 54(3): 58-76.
- [37] Hausman, J.A. and P. Joskow (1982) "Evaluating the Costs and Benefits of Appliance Efficiency Standards," *American Economic Review*, 72: 220-225.
- [38] Hossain, T. and J. Morgan (2005) "...Plus Shipping and Handling: Revenue (Non) Equivalence in Field Experiments on eBay," *Advances in Economic Analysis & Policy*.
- [39] 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.
- [40] Krishna, A. and J. Slemrod (2003) "Behavioral Public Finance: Tax Design as Price Presentation," *International Tax and Public Finance*, 10: 189-203.
- [41] Lakins, N.E., G.D. Williams, H. Yi and B.A. Smothers (2004) "Surveillance Report #66: Apparent Per Capita Alcohol Consumption: National, State, and Regional Trends, 1977-2002" Bethesda, MD: NIAAA, Alcohol Epidemiologic Data System.
- [42] Liebman, J. and R. Zeckhauser (2004), "Schmeduling," Harvard KSG Working Paper.
- [43] Mackowiak, B. and M. Wiederholt (2006) "Optimal Sticky Prices under Rational Inattention," Northwestern University Working Paper.
- [44] Mankiw, G. M. (1985) "Small Menu Costs and Large Business Cycles: A Macroeconomic Model," *Quarterly Journal of Economics*, 100 (2): 529-538.
- [45] McCaffery, E.J. and J. Baron. (2006). "Thinking about Tax." *Psychology, Public Policy & Law* 12(1):106-35.
- [46] Mirrlees, J.A. (1971) "An Exploration in the Theory of Optimum Income Taxation," *The Review of Economic Studies*, 38(2): 175-208.
- [47] Mohring, H. (1971) "Alternative welfare gain and loss measures," *Western Economic Journal*, 9: 349-368.

- [48] Morwitz, V.G., E.A. Greenleaf, and E.J. Johnson (1998) "Divide and Prosper: Consumers' Reactions to Partitioned Prices," *Journal of Marketing Research*, 35: 453-463.
- [49] Mullainathan, S. (2002) "A Memory-Based Model of Bounded Rationality," *Quarterly Journal of Economics*, 117(3): 735-774.
- [50] National Institute on Alcohol Abuse and Alcoholism (2006) "Per Capita Alcohol Consumption" Data available at <http://www.niaaa.nih.gov/Resources/DatabaseResources/QuickFacts/AlcoholSales>
- [51] Nephew, T.M., H. Yi, G.D Williams, F.S. Stinson, and M.C. Dufour (2004) "U.S. Alcohol Epidemiologic Data Reference Manual" NIH Publication No. 04-5563.
- [52] O'Donoghue, T. and M. Rabin (2006) "Optimal Sin Taxes," *Journal of Public Economics*, 90(10-11): 1825-1849.
- [53] Prescott, E. (2004) "Why do Americans Work so Much More than Europeans?" NBER Working Paper No. 10316.
- [54] Reis, R. (2006) "Inattentive Consumers," *Journal of Monetary Economics*, 53: 1761-1800.
- [55] Rosen, H. (1976) "Taxes in a Labor Supply Model with Joint Wage-Hours Determination," *Econometrica*, 44(3): 485-507.
- [56] Saez, E. (2002) "Do Taxpayers Bunch at Kink Points," NBER Working Paper No. 7366.
- [57] Saez, E. (2004) "Reported Incomes and Marginal Tax Rates, 1960-2000: Evidence and Policy Implications," NBER Working Paper No. 10273.
- [58] Shafir, E., P. Diamond, and A. Tversky (1997). "Money Illusion." *Quarterly Journal of Economics*, 112(2): 341-74.
- [59] Sheshinski, E. (2002) "Bounded Rationality and Socially Optimal Limits on Choice in A Self-Selection Model," Hebrew University Working Paper.
- [60] Simon, H. (1955) "A Behavioral Model of Rational Choice," *Quarterly Journal of Economics*, 69(1): 99-118.
- [61] Sims, C. (2003) "Rational Inattention," Princeton University Working Paper.
- [62] Tax Foundation (various years), *Special Report: State Tax Rates and Collections*, Washington D.C. <http://www.taxfoundation.org>
- [63] World Tax Database (2006), University of Michigan Business School. <http://www.bus.umich.edu/OTPR/otpr/introduction.htm>

Appendix A: Data Construction

Grocery Scanner Data. 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. The tags were printed using a template and card stock supplied by the store (often used for sales or other additional information on a product) in order to match the color scheme and layout familiar to customers.

The raw scanner data provided by the grocery chain contains information on weekly revenue and quantity sold for each product (UPC id) that was sold among the 108 categories listed in Appendix Table 1 in the three stores from 2005 week 1 to 2006 week 15. We construct price as revenue divided by quantity. The original dataset contains 331,508 product-week-store observations. The quantity and revenue variables are measured net of returns (i.e., returns count as negative sales). We exclude 1,756 observations where the weekly quantity or revenue was zero or negative, which are cases where as many or more items were returned than purchased in that week. Including these observations does not affect the results. Finally, we aggregate to the category-week-store level by summing quantity and revenue across products, setting the sum to zero if no products were sold in a given category-week-store.

The average price for each category of goods is defined as $P_{ct} = \sum_{i \in c} (p_{it} \bar{q}_i) / \sum_{i \in c} \bar{q}_i$ where c indexes the category, t time, and i products, p_{it} is the price of good i at time t , and \bar{q}_i is the average quantity sold of good i . This “category price” is effectively a price index for a fixed basket of products where each product’s weight in the basket is determined by its average weekly sales over the period before and during the experiment. Since the scanner data reports only items that have sold each week, we impute prices for unsold items when constructing P_{ct} . In particular, we use the price in the last observed transaction for unsold products; if no previous price is available, we use the next available price. Alternative imputation methods – such as using the closest observed price, or an average of previous and subsequent prices – give similar results. Varying the imputation technique has little impact on the estimates in Tables 4 and 5 because items requiring imputation have low sales volume, and therefore receive little weight in the category-level price variable.

Alcohol Consumption and Tax Data. Data on aggregate annual beer, wine, spirits, and ethanol consumption by state are available from the National Institute of Alcohol Abuse and Alcoholism (2006) from 1970-2003. These data 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.

State excise tax rates on beer are primarily obtained from the Brewer’s Almanac (various years), published annually by the Beer Institute. These rates have been verified and corrected using the Tax Foundation’s State Tax Collections and Rates (various years) and the State Tax Handbook. Our measure of the excise rate includes taxes that are statutorily ‘local’ excise taxes – which are sometime excluded from state statistics available in the Brewer’s Almanac – that are applied state-wide. Specifically, in Alabama, Georgia, and

Louisiana all counties or localities levy an excise tax in addition to the state excise tax.

Excise taxes on alcohol frequently differ by product, packaging, and whether sold for on- or off-premise consumption. In states where rates differ, our measure corresponds to the excise tax on packaged 12oz. beer, sold for off-premise consumption, with an alcohol content of 3.2 percent or more. Excise rates on other beer products are highly correlated with this measure across states, and the timing of tax changes for different categories of alcoholic beverages within a state are virtually identical. Per-gallon taxes are converted to per-case rates by multiplying by 2.25, the number of gallons in 24 12oz. cans or bottles. The excise tax rate is converted into an ad valorem rate by dividing the real CPI-adjusted beer excise tax per case in year 2000 dollars by the average cost of a case of beer in the United States in 2000, as measured by the Beer Institute. 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.

State sales taxes are obtained primarily from the World Tax Database (2006) at the University of Michigan. These data were verified and corrected using state Department of Revenue websites and the State Tax Handbook. Four states (KS, VT, DC, MN) apply a higher sales tax rate to alcohol than other products. In those states we include the alcohol rate rather than the general sales rate when they differ. We supplement the data on state-level sales taxes with data on average local sales tax rates, which are imputed from data on local revenues from the Census Bureau's Survey of State and Local Government Finances and a tax base defined as state revenues divided by the state rate.

Since our estimation strategy relies on the timing and magnitude of the tax changes, we evaluate the precision 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. In the pooled sample, the coefficient estimate on the sales tax rate is 0.76 (s.e. 0.03). A state-by-state analysis of changes in rates and changes in revenues also yields similarly high correlations, with the exception of West Virginia. In WV, the correlation between sales tax rates and revenues is near zero and statistically insignificant, perhaps because the tax base is often changed at the same time as the rate. Since this problem could artificially attenuate the sales tax elasticity, we exclude West Virginia from our analysis.

Appendix B: Proofs

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

Let (x^p, y^p) denote the agent's allocation when he ignores the tax. Start from the definition of the gain in utility from computing the tax-inclusive price:

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

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

$$\begin{aligned}\tilde{G}(t) &= -\frac{1}{2}u''(x^*)(x^* - x^p)^2 + \frac{1}{2}v''(y^*)(y^* - y^p)^2 \\ &= -\frac{1}{2}(x^* - x^p)^2[u''(x^*) + v''(y^*)(1+t)^2]\end{aligned}\tag{25}$$

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

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

It follows that

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

and hence

$$\tilde{G}(t) = -\frac{1}{2}(x^* - x^p)^2 \frac{[v'(y^*) - (1+t)x^*v''(y^*)]}{\partial x^*/\partial t}.\tag{26}$$

The derivation of (26) does not depend on the budget adjustment rule used by the inattentive agent. To simplify (26) further, we express $x^* - x^p$ in terms of elasticities by specifying how the agent sets x^p . Under the first budget adjustment rule, where the agent chooses x first, we can linearly approximate $x^* - x^p = \frac{\partial x^*}{\partial t}(t)t$, reflecting the fact that the difference between x^* and x^p is proportional to the effect of the tax on x for the optimizing agent. It follows that

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

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

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

Under the second budget adjustment rule, an analogous expression for $G(t)$ can be derived as follows. Differentiating the agent's f.o.c. and rearranging terms, it can be shown that $[u''(x^*) + v''(y^*)(1+t)^2] = -\frac{\partial y^*}{\partial t}[u'(x^*) - xu''(x^*)]$. Next, observe that $x^* - x^p = \frac{y^p - y^*}{1+t}$ and $y^* - y^p = \frac{\partial y^*}{\partial t}$. Plugging these expressions into (25) and normalizing by $v'(y^*) = u'(x^*)/(1+t)$, we obtain

$$G_2(t) = \frac{1}{2}t^2\varepsilon_{y,p}y^*(t)\frac{1}{(1+t)^2}[1 - \gamma_x] \quad (27)$$

where $\gamma_x = -\frac{u''(x^*)x^*}{u'(x^*)}$.

Derivation of Equation (15) for $DWL(\Delta t|t_0^S, t_0^E)$ in Quasilinear Case

To derive an exact measure for the excess burden in the quasilinear case algebraically, consider an agent with cognitive cost c . Let $\theta = 0$ if the agent does not compute p_t and $\theta = 1$ if he does. Then the agent's indirect utility is

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

and hence the expenditure function is

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

Assume temporarily that $\theta(t) = \theta(t + \Delta t)$, i.e. the agent does not change his cognitive decision following the tax change. Then substituting the preceding expressions for V and e into (13) with $p = 1 + t_0^E$ yields

$$EB(\Delta t|t_0^S, t_0^E) = [u(x(1 + t_0^E, t_0^S)) - x(1 + t_0^E, t_0^S)] - [u(x(1 + t_0^E, t_0^S + \Delta t)) - x(1 + t_0^E, t_0^S + \Delta t)] \quad (28)$$

For agents who ignore taxes that are not included in posted prices, $x(1 + t_0^E, t_0^S) = x(1 + t_0, t_0^S + \Delta t)$ and hence $EB^p(\Delta t) = 0$.

For agents who always re-optimize relative to tax-inclusive prices, (28) 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 $x_0^* = x^*(1 + t_0)$, we obtain

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

To characterize aggregate deadweight loss in the economy with heterogeneous agents, observe that deadweight loss at tax rates t^S, t^E is given by $EB^*(t^E) + \theta(t^S)EB^*(t^S|t^E) + C(t^S)$. Letting $\theta_1 = \theta(t_0^S + \Delta t)$ and $\theta_0 = \theta(t_0^S)$, we obtain

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

Using the definition of EB^* yields

$$DWL(\Delta t|t_0^S, t_0^E) = \theta_1[x_0^*\varepsilon_{x,p}t_0\frac{\Delta t}{1+t_0} + \frac{1}{2}x_0^*\varepsilon_{x,p}\frac{(\Delta t)^2}{1+t_0}] + \frac{\partial\theta}{\partial t}\Delta t(-\frac{1}{2}\frac{\partial x^*}{\partial t}(t_0^S)^2 - t_0^E\frac{\partial x^*}{\partial t}t_0^S - \frac{1}{2}\frac{\partial x^*}{\partial t}(t_0^S)^2)$$

and some algebra yields (15).

Derivation of Equation (17) for Excess Burden EB_1^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 $EB^*(t) = \frac{1}{2}\frac{t^2}{1+t}x^*(t)\varepsilon_{x,p}^c$ yields

$$EB^p(t) = \frac{1}{2}\frac{t^2}{1+t}x^*(t)\varepsilon_{x,p}^c + [e(1, 0, V^*(1, t, Z)) - e(1, 0, V^p(1, t, Z))] - t[x^p - x^*(t)] \quad (29)$$

We first establish that $[e(1, 0, V^*(1, t, Z)) - e(1, 0, V^p(1, t, Z))] = G(t)$ under a linear approximation of the Hicksian demand functions. Observe that

$$\begin{aligned} & [e(1, 0, V^*(1, t, Z)) - e(1, 0, V^p(1, t, Z))] \\ &= [x^c(1, 0, V^*) - x^c(1, 0, V^p)] + [y^c(1, 0, V^*) - y^c(1, 0, V^p)] \\ &\approx [\frac{\partial x^c}{\partial V} + \frac{\partial y^c}{\partial V}][V^*(1, t, Z) - V^p(1, t, Z)] \end{aligned}$$

Totally differentiating the utility constraint for the expenditure minimization problem with respect to V and using the agent's first-order condition when $t = 0$ yields

$$\frac{\partial x^c}{\partial V} + \frac{\partial y^c}{\partial V} = \frac{1}{v'(y^*)}$$

Hence

$$[e(1, 0, V^*(1, t, Z)) - e(1, 0, V^p(1, t, Z))] = \frac{1}{v'(y^*)}[V^*(1, t, Z) - V^p(1, t, Z)] = G(t).$$

Returning to (29), it follows that

$$EB^p(t) = \frac{1}{2}\frac{t^2}{1+t}x^*(t)\varepsilon_{x,p}^c + G(t) - t[x^p - x^*(t)] \quad (30)$$

This equation for EB^p holds irrespective of the budget adjustment rule. To go further, we plug in the expression for $G(t)$ when the agent uses the first budget adjustment rule in (10) to obtain

$$EB_1^p(t) = \frac{1}{2}t^2x^*(t)\frac{\varepsilon_{x,p}^c}{1+t} + \frac{1}{2}t^2x^*(t)\varepsilon_{x,p}(\frac{1}{1+t} + \frac{x^*(t)}{y^*(t)}\gamma_y) - t[x^p - x^*(t)]$$

Next, using the approximation $[x^p - x^*(t)] = [x^*(t=0) - x^*(t)] = -\frac{\partial x^*}{\partial t}t$ yields

$$EB_1^p(t) = \frac{1}{2}t^2x^*(t)\frac{\varepsilon_{x,p}^c}{1+t} + \frac{1}{2}t^2x^*(t)\varepsilon_{x,p}\left(\frac{1}{1+t} + \frac{x^*(t)}{y^*(t)}\gamma_y\right) - t^2x^*(t)\frac{\varepsilon_{x,p}}{1+t} \quad (31)$$

$$= \frac{1}{2}\frac{t^2}{1+t}x^*(t)[\varepsilon_{x,p}^c - \varepsilon_{x,p}] + \frac{1}{2}t^2x^*(t)\varepsilon_{x,p}\frac{x^*(t)}{y^*(t)}\gamma_y \quad (32)$$

The key step in simplifying this expression is to recognize that the coefficient of relative risk aversion is related to the ratio of the income effect to the substitution effect (Chetty 2006):

$$\gamma_y = \frac{y^*(t)}{Z} \frac{\varepsilon_{x,z}}{\varepsilon_{x,p}^c} \quad (33)$$

To derive (33), implicitly differentiate the fully-optimizing first-order-condition $u'(x^*(t)) = p_tv'(y^*(t))$ to calculate $\frac{\partial x^*}{\partial p_t}$ and $\frac{\partial x^*}{\partial Z}$:

$$\begin{aligned} \frac{\partial x^*}{\partial p_t} &= \frac{v'(y) - xp_tv''(y)}{u''(x) + p_t^2v''(y)} \\ \frac{\partial x^*}{\partial Z} &= \frac{p_tv''(y)}{u''(x) + p_t^2v''(y)} \end{aligned}$$

Using the Slutsky equation $\frac{\partial x^c}{\partial p_t} = \frac{\partial x^*}{\partial p_t} + x\frac{\partial x^*}{\partial Z}$, we obtain

$$\frac{\partial x^*/\partial Z}{\partial x^c/\partial p_t} = \frac{(1+t)v''(y)}{v'(y)}$$

Finally, defining $\varepsilon_{x,p}^c = -\frac{\partial x^c}{\partial p_t} \frac{1+t}{x_0}$ as the compensated elasticity at $p_t = 1+t$ and $\varepsilon_{x,z} = \frac{\partial x}{\partial Z} \frac{Z}{x_0}$, and rearranging yields (33). Plugging in the expression for γ_y into (31), using the Slutsky equation $\varepsilon_{x,p}^c - \varepsilon_{x,p} = -\varepsilon_{x,z} \frac{(1+t)x}{Z}$ and collecting terms, we obtain (17).

The formula for EB_p^2 in (18) is obtained using an analogous derivation. Starting from (30), we use the equation $\Delta x = \frac{\Delta y}{1+t}$ to write all the elasticities in terms of y . We then plug in the formula for G under the second budget adjustment rule in (27). Finally, we use the identity $\gamma_x = \frac{(1+t)x^*}{Z} \frac{\varepsilon_{y,z}}{\varepsilon_{y,p}^c}$ where $\varepsilon_{y,p}^c = \frac{\partial y^c}{\partial p} \frac{1+t}{y^c}$ to collect terms and simplify the expression.

Derivation of Equation (20) for $DWL(\Delta t|t_0^S)$ in General Case

Differentiating (19) yields

$$\frac{\partial DWL(t)}{\partial t} \approx \theta(t) \frac{\partial EB^*(t)}{\partial t} + \Sigma \sigma_i \frac{\partial EB_i^p(t)}{\partial t} + \Sigma \frac{\partial \sigma_i}{\partial t} [EB_i^p(t_0^S) - EB^*(t_0^S)] + \frac{\partial C_g(t)}{\partial t} \quad (34)$$

Note that $\frac{\partial C_g(t)}{\partial t} = \Sigma G_i(t) \kappa_i f_i(G_i) = \Sigma G_i(t) \frac{-\partial \sigma_i}{\partial t}$. Hence

$$\frac{\partial DWL(t)}{\partial t} \approx \theta(t) \frac{\partial EB^*(t)}{\partial t} + \Sigma \sigma_i \frac{\partial EB_i^p(t)}{\partial t} + \Sigma \frac{\partial \sigma_i}{\partial t} [EB_i^p(t_0^S) - EB^*(t_0^S) - G_i(t_0^S)]$$

Using equation (16), it follows that

$$EB_i^p(t) = EB^*(t) + G_i(t) - t[x_i^p - x^*(t)].$$

Substituting this expression for $EB_i^p(t)$ in (34), we obtain

$$\begin{aligned} \frac{\partial DWL(t)}{\partial t} &\approx \theta(t) \frac{\partial EB^*(t)}{\partial t} + \Sigma \sigma_i \frac{\partial EB_i^p(t)}{\partial t} - \Sigma \frac{\partial \sigma_i}{\partial t} t[x_i^p - x^*(t)] \\ &= \theta(t) \frac{\partial EB^*(t)}{\partial t} + \Sigma \sigma_i \frac{\partial EB_i^p(t)}{\partial t} + \theta \varepsilon_{\theta, t^S} [\bar{x}^p - x^*(t)] \end{aligned} \quad (35)$$

where the last equality follows from the equi-proportionality assumption $\varepsilon_{\sigma_i, t} = \varepsilon_{\sigma_j, t}$.

Under the approximation that compensated demand is linear in price, $\frac{\partial x^c}{\partial t}(x^*(t_0^S))$ is constant, and it follows that $\frac{\partial EB^*(t)}{\partial t} = \frac{t_0^S}{1+t_0^S} x_0^* \varepsilon_{x,p}^c$ (Browning 1987). To calculate $\frac{\partial EB_1^p(t)}{\partial t}$, note that we can write $EB_1^p(t) = -\frac{1}{2} t^2 x^*(t)^2 (\frac{\partial x}{\partial Z})^2 / \frac{\partial x^c}{\partial p}$. Under the additional approximation that demand is linear in income, $\frac{\partial x}{\partial Z}$ is constant, and it follows that

$$\frac{\partial EB_1^p(t_0^S)}{\partial t} = t_0^S (1 + t_0^S) x_0^* \frac{x_0^*}{Z} \frac{y_0^*}{y_0^*} \gamma_y \varepsilon_{x,z} - (t_0^S)^2 x_0^* \frac{x_0^*}{Z} \frac{y_0^*}{y_0^*} \gamma_y \varepsilon_{x,z} \varepsilon_{x,p}$$

where all the elasticities are evaluated at a price of $1 + t_0^S$. Similarly, note that $EB_2^p(t) = \frac{1}{2} \frac{t^2}{1+t} x^*(t)^2 (\frac{\partial y}{\partial Z})^2 / \frac{\partial y^c}{\partial p}$. Using linear approximations for $y^*(t, Z)$, we obtain

$$\frac{\partial EB_2^p(t_0^S)}{\partial t} = \frac{t_0^S}{1+t_0^S} \frac{1 + \frac{1}{2} t_0^S}{1+t_0^S} x_0^* \frac{x_0^*}{Z} \frac{y_0^*}{x_0^*} \gamma_x \varepsilon_{y,z} - \frac{(t_0^S)^2}{(1+t_0^S)^2} x_0^* \frac{x_0^*}{Z} \frac{y_0^*}{x_0^*} \gamma_x \varepsilon_{y,z} \varepsilon_{x,p}$$

Finally, plugging in the derivatives of $EB^*(t)$ and $EB_i^p(t)$ into (35), multiplying by Δt , and factoring yields the expression for $DWL(\Delta t | t_0^S)$ in (20).

TABLE 1
Descriptive Statistics: Grocery Stores

	Treatment Store	Control Store #1	Control Store #2
<i>A. Store Characteristics</i>			
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
<i>B. City Characteristics (in 1999)</i>			
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. Census Bureau, 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
<i><u>A. Category Level Statistics:</u></i>							
Weekly quantity sold per category	25.08 (24.1)	26.63 (38.1)	24.88 (24.5)	28.45 (41.5)	30.80 (29.7)	32.83 (51.9)	29.01 (42.5)
Weekly revenue per category	\$97.85 (81.9)	\$136.05 (169.9)	\$93.26 (82.7)	\$144.09 (187.9)	\$120.81 (99.1)	\$165.24 (225.3)	\$143.10 (187.1)
Number of categories	13	95	13	95	13	95	108
<i><u>B. Product Level Statistics</u></i>							
Pre-tax product price	4.46 (1.8)	6.26 (4.3)	4.37 (1.6)	6.30 (4.4)	4.64 (1.8)	6.32 (4.1)	6.05 (4.1)
Pre-tax product price (weighted by quantity sold)	4.27 (4.7)	5.61 (3.9)	4.16 (1.6)	5.58 (3.9)	4.38 (1.7)	5.60 (3.7)	5.45 (3.7)
Weekly quantity sold per product (conditional >0)	1.47 (0.9)	1.82 (1.6)	1.58 (1.0)	1.95 (1.8)	1.63 (1.1)	2.01 (2.0)	1.88 (1.7)

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. See Appendix A for data sources and sample definition.

TABLE 3
DDD Analysis of Means: Weekly Quantity by Category

<u>TREATMENT STORE</u>			
Period	<u>Control Categories</u>	<u>Treated Categories</u>	<u>Difference</u>
Baseline (2005:1- 2006:6)	26.48 (0.22) [5510]	25.17 (0.37) [754]	-1.31 (0.43) [6264]
Experiment (2006: 8- 2006:10)	27.32 (0.87) [285]	23.87 (1.02) [39]	-3.45 (0.64) [324]
Difference over time	0.84 (0.75) [5795]	-1.30 (0.92) [793]	DD_{TS} = -2.14 (0.64) [6588]
<u>CONTROL STORES</u>			
Period	<u>Control Categories</u>	<u>Treated Categories</u>	<u>Difference</u>
Baseline (2005:1- 2006:6)	30.57 (0.24) [11020]	27.94 (0.30) [1508]	-2.63 (0.32) [12528]
Experiment (2006: 8- 2006:10)	30.76 (0.72) [570]	28.19 (1.06) [78]	-2.57 (1.09) [648]
Difference over time	0.19 (0.64) [11590]	0.25 (0.92) [1586]	DD_{CS} = 0.06 (0.90) [13176]
DDD Estimate			-2.20 (0.58) [19764]

Notes: Each cell shows mean number of units sold per category per week, for various subsets of the sample. Standard errors (clustered by week) in parentheses, number of observations in square brackets. Experimental period spans promotional week 8 in 2006 to promotional week 10 in 2006. Baseline period spans promotional week 1 in 2005 to promotional week 6 in 2006. A promotional week is a standard calendar week beginning on a Wednesday and ending the following Tuesday. See Appendix Table 1 for list of treatment and control categories. Lower panel reflects averages across the two control stores.

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 (0.58)***	-2.20 (0.59)***	-13.12 (4.88)***	-0.101 (0.03)***	-0.123 (0.04)***	-0.102 (0.212)
Average Price		-3.15 (0.26)***	-3.24 (1.74)*			
Average Price Squared		0.05 (0.00)***	0.06 (0.03)**			
Log Average Price				-1.59 (0.11)***	-0.39 (0.11)***	
Category, Store, Week FEs		x	x	x	x	x
Sample size	19,764	19,764	19,764	18,827	18,827	18,827

Notes: Standard errors, clustered by week, reported in parentheses: * significant at 10%; ** significant at 5%; *** significant at 1%. All columns report estimates of the linear regression model specified in equation (3). Quantity and revenue reflect total sales of products within a given category in a given promotional week in a given store. Average price is a weighted average of the prices of the goods for sale in each category, using a fixed basket of goods (weighted by total quantity sold) over time. In columns 4 and 5, observations are weighted by total quantity by category-store. In column 6, dependent variable is the average price of products that sold in a given category, week, and store (weighted by total quantity sold as in the previous specifications).

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 (0.60)***	-1.55 (0.35)***	-2.61 (0.72)***
Before Treatment	-0.21 (1.07)		
After Treatment	0.20 (0.78)		
Average Price	-3.04 (0.25)***	-15.06 (3.55)***	-7.99 (0.81)***
Average Price Squared	0.05 (0.00)***	1.24 (0.34)***	0.12 (0.01)***
Category, Store, Week FEs	x	x	x
N	21,060	2,379	6,588

Notes: Standard errors, clustered by week, in parentheses: * significant at 10%; ** significant at 5%; *** significant at 1%. 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 and 3 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 the treatment store only. The "Treatment" variable is defined as the interaction between the treatment category dummy and the treatment time dummy. All specifications include average price and its square and category, store, and promotional week fixed effects.

TABLE 6
Descriptive Statistics for Alcohol Consumption, Taxes, and Regulation

State Beer Consumption (Gallons)	110,003 (120144)
State Per-Capita Beer Consumption (Cans/Pop.)	243.2 (46.1)
State Beer Excise Tax (\$/case)	0.51 (0.50)
Federal Beer Excise Tax (\$/case)	0.90 (0.03)
State Beer Excise Tax (Percent)	6.4 (8.2)
General Sales Tax (Percent)	4.3 (1.9)
State Drinking Age is 21	0.73 (0.44)
State has Drunk Driving Standard	0.66 (0.47)
Any Alcohol Regulation Change	0.19 (0.39)
N	1,690

Notes: Means; standard deviations in parentheses. "State Drinking Age is 21" is an indicator for state has raised the legal drinking age to 21; "State has Drunk Driving Standard" indicates state has set a threshold blood alcohol content level above which one is automatically guilty of drunk driving. "Any Alcohol Regulation Change" is a dummy variable equal to one in any year where a state has raised the drinking age, implemented a stricter drunk driving standard, implemented an administrative license revocation law, or a zero tolerance youth drunk driving law. See Appendix A for data sources and sample definition.

TABLE 7
Effect of Excise and Sales Taxes on Beer Consumption

Dependent Variable: Change in Log(per capita beer consumption)				
	Baseline	Bus Cycle	Bus Cycle Lags	Alc Regulations
	(1)	(2)	(3)	(4)
ΔLog(1+Excise Tax Rate)	-0.87	-0.91	-0.86	-0.89
	(0.17)***	(0.17)***	(0.17)***	(0.17)***
ΔLog(1+Sales Tax Rate)	-0.20	-0.00	0.03	-0.02
	(0.30)	(0.30)	(0.30)	(0.30)
ΔLog(Population)	0.03	-0.07	0.05	-0.07
	(0.06)	(0.07)	(0.19)	(0.07)
ΔLog(Income per Capita)		0.22	0.18	0.22
		(0.05)***	(0.05)***	(0.05)***
ΔLog(Unemployment Rate)		-0.01	-0.01	-0.01
		(0.01)**	(0.01)	(0.01)**
Lag Bus. Cycle Controls			x	
Alcohol Regulation Controls				x
Year Fixed Effects	x	x	x	x
F-Test for Equality of Tax Variables (Prob>F)	0.05	0.01	0.01	0.01
Sample Size	1607	1487	1440	1487

Notes: Standard errors, clustered by state, in parentheses: * significant at 10%; ** significant at 5%; *** significant at 1%. All specifications include year fixed effects and log state population. Column 2 controls for log state personal income per capita and log state unemployment rate (unavailable in some states in the early 1970s). Column 3 adds one year lags of personal income per capita and unemployment rate variables. Column 4 controls for changes in alcohol policy by including three separate indicators for whether the state implemented per se drunk driving standards, administrative license revocation laws, or zero tolerance youth drunk driving laws, and the change in the minimum drinking age (measured in years).

TABLE 8
Effect of Excise and Sales Taxes on Beer Consumption

	Dependent Variable: Change in Log(per capita beer consumption)					Dep Var: Share Ethanol from Beer
	Include Local Sales Taxes	Region Trends	IV for Excise w/ Policy	3-Year Differences	Food Exempt	
	(1)	(2)	(3)	(4)	(5)	(6)
ΔLog(1+Excise Tax Rate)	-0.68	-0.71	-0.62	-1.11	-0.91	0.16
	(0.24)***	(0.18)***	(0.28)**	(0.46)**	(0.22)***	(0.13)
ΔLog(1+Sales Tax Rate)	0.14	-0.05	-0.03	-0.00	-0.14	0.25
	(0.34)	(0.30)	(0.30)	(0.32)	(0.30)	(0.22)
ΔLog(Population)	-0.12	-0.09	-0.06	-1.30	0.03	0.09
	(0.08)	(0.08)	(0.07)	(0.37)***	(0.07)	(0.05)*
ΔLog(Income per Capita)	0.19	0.22	0.22	0.09	0.22	0.01
	(0.05)***	(0.05)***	(0.05)***	(0.07)	(0.05)***	(0.03)
ΔLog(Unemployment Rate)	-0.01	-0.01	-0.01	-0.00	-0.01	0.00
	(0.01)*	(0.01)**	(0.01)**	(0.01)	(0.01)	(0.01)
Alcohol Regulation Controls	x	x	x	x	x	x
Year Fixed Effects	x	x	x	x	x	x
F-Test for Equality of Tax Variables (Prob>F)	0.05	0.06	0.16	0.05	0.04	0.74
Sample Size	1104	1487	1487	1389	937	1487

Notes: Standard errors, clustered by state, in parentheses: * significant at 10%; ** significant at 5%; *** significant at 1%. Column 1 replicates column 4 of Table 7, defining the sales tax rate as the state sales tax plus average local sales tax, imputed as described in the appendix. Column 2 reverts to the state sales tax measure and replicates column 4 of Table 7 including fixed effects for each of nine census regions. In column 3, excise tax rate changes are instrumented by the nominal excise tax rate divided by the average price of a case of beer from 1970 to 2003, to eliminate variation in the tax rate due to inflation erosion. Column 4 shows estimates of the model with all variables defined using three-year differences instead of first-differences. Column 5 restricts the sample to states where food was exempt from taxation in 2000 and replicates the first-difference specification. In column 6, the dependent variable is the fraction of total ethanol consumption in each state-year accounted for by beer.

TABLE 9

Individual vs. Social Welfare with Quasilinear Utility

	(1) $t_0^E=0$	(2) $t_0^E=25\%$	(3) $\theta = 0$	(4)	(5) $\theta = 1$	(6)
t^S	G	G	ΔR	DWL	ΔR	DWL
1%	\$0.50	\$0.40	\$100.00	\$0	\$79.20	\$20.40
5%	12.50	10	500	0	380	110
10%	50	40	1000	0	720	240
20%	200	160	2000	0	1280	560
30%	450	360	3000	0	1680	960

Notes: Calibrations assume that utility is quasilinear in y , $x_0=10,000$, $\varepsilon_{x,p}^0=1$. Column 1 assumes no pre-existing taxes; columns 2-6 assume a pre-existing excise tax of $t_0^E=0.25$. Column 1 shows individual utility gain from computing tax-inclusive price when a sales tax at rate t^S is introduced. Column 2 shows the same when there is a pre-existing excise tax of 25%. Column 3 shows the revenue obtained from the sales tax when agent is inattentive and does not change demand in response to sales tax. Column 4 shows deadweight loss for this agent. Columns 5 and 6 show revenue obtained from the sales tax and deadweight loss when agent is attentive and re-optimizes in response to the tax.

APPENDIX TABLE 1

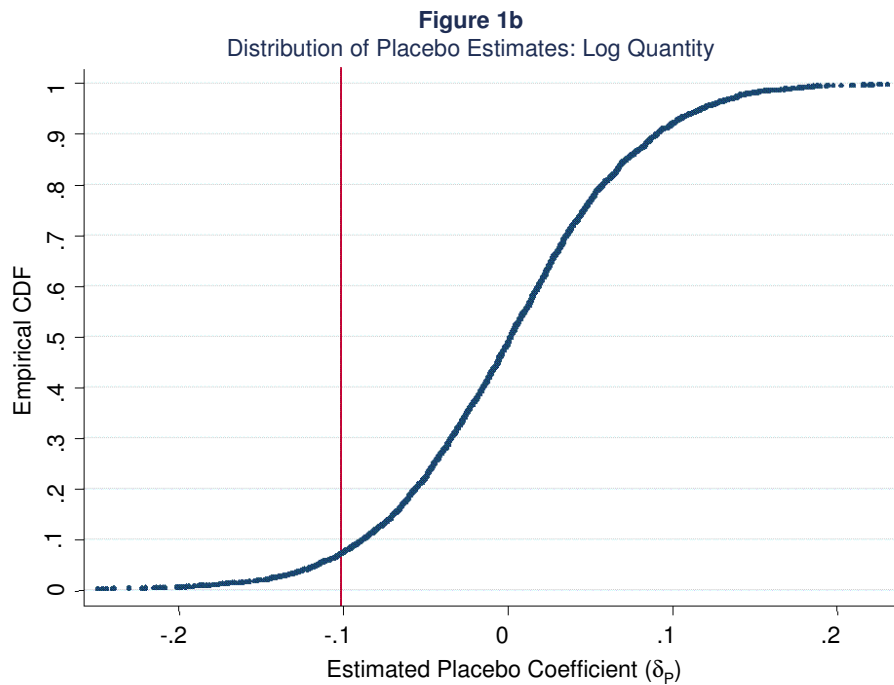
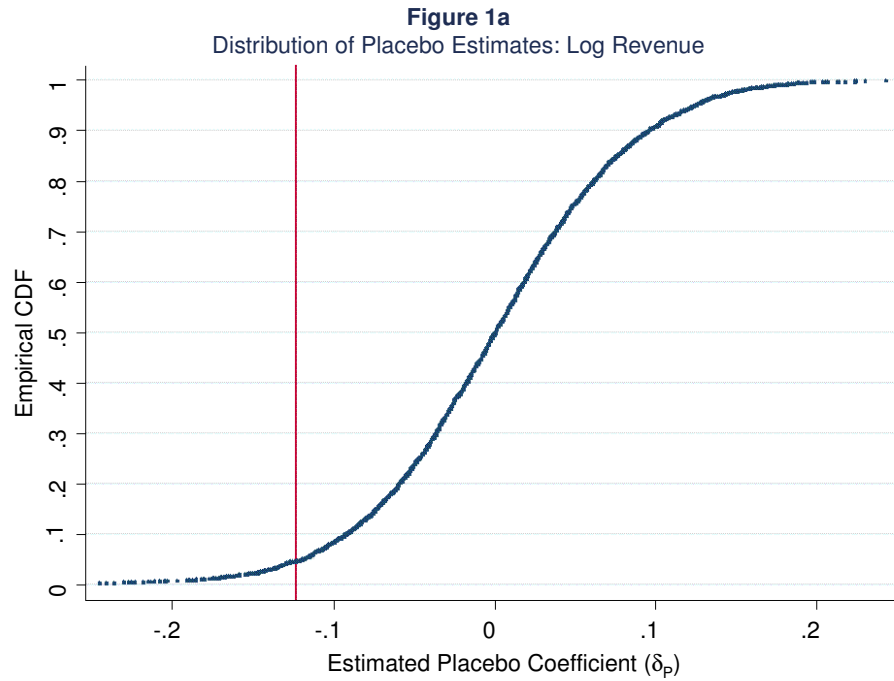
Category Classification in Grocery Store Data

Categories	Group Description	Category Description	Mean Weekly Revenue
<u>Treatment</u>			
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
<u>Control</u>			
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	Men's 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 Compounds	86.75
5718	Pain Relief	Specialty Indication Pain	88.92
5725	Pain Relief	Children's/Infants Analgesics	187.25
5730	Pain Relief	Sleeping Aids	64.99
5735	Pain Relief	Stimulants	14.82
5750	Pain Relief	Nighttime Pain Relief	76.19
5760	Pain Relief	External Analgesic	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	Bronchial 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 Symptom 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-diarrhea	82.70
5940	Digestive Health	Laxatives	265.29
5945	Digestive Health	Lactose Intolerance	22.14
5950	Digestive Health	Rectal/Hemorrhoidal	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.



NOTE— These figures plot the empirical distributions of placebo effects (G) for log revenue and quantity. The CDFs are constructed from 4,725 estimates of δ_P using the specifications in columns 4 and 5 of Table 4. No parametric smoothing is applied: the CDFs appear smooth because of the large number of points used to construct them. The vertical lines show the treatment effect estimates reported in Table 4.

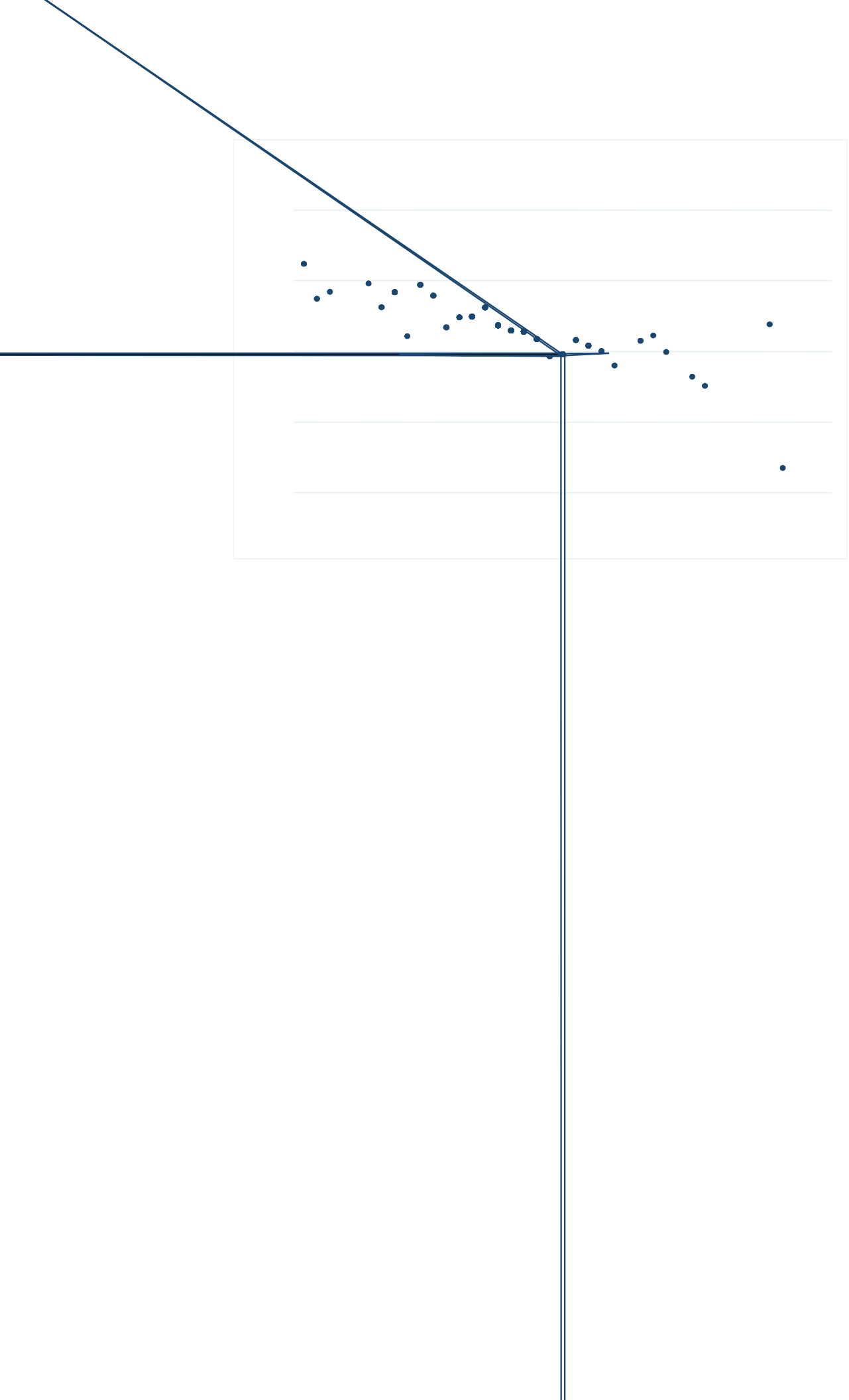


Figure 3a
Number of Correctly Reported Taxed Items on Survey

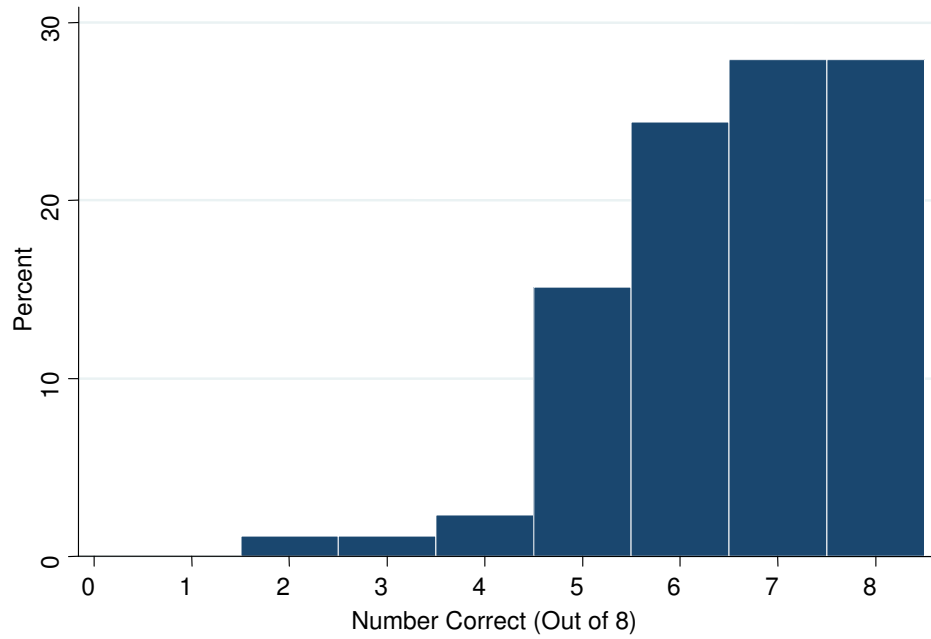
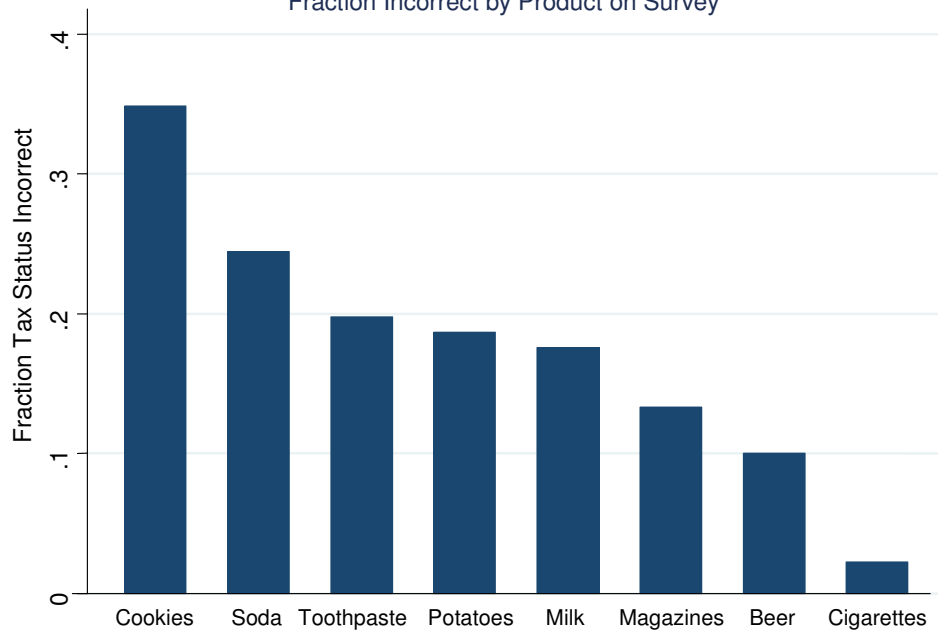
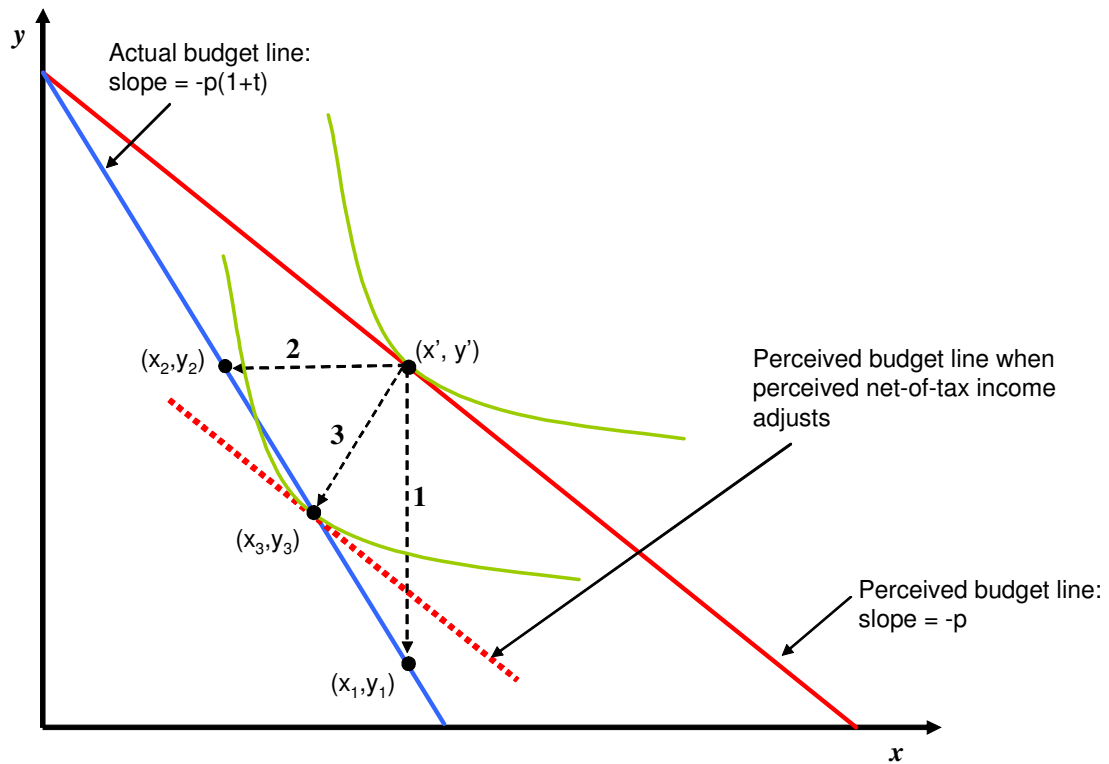


Figure 3b
Fraction Incorrect by Product on Survey



NOTE—Figure 3a is a histogram of the number of products whose tax status was correctly identified by survey respondents. Figure 3b plots the fraction of incorrect answers regarding tax status by product category.

Figure 4
Three Budget Adjustment Rules



NOTE—This figure illustrates three ways in which an agent who ignores the tax on good x can reach a feasible allocation given his true budget constraint. The agent's initial (infeasible) allocation under his perceived budget constraint is given by (x', y') . Under the first budget adjustment rule, the agent cuts consumption of y to meet his budget, leading to the choice (x_1, y_1) . Under the second budget adjustment rule, the agent cuts consumption of x , leading to (x_2, y_2) . Under the third rule, the agent cuts consumption of both goods, and chooses a point that is optimal when his budget set is shifted inward to match his true net-of-tax income at the chosen allocation (x_3, y_3) .

Figure 5a
Cost of Ignoring Tax with Quasilinear Utility

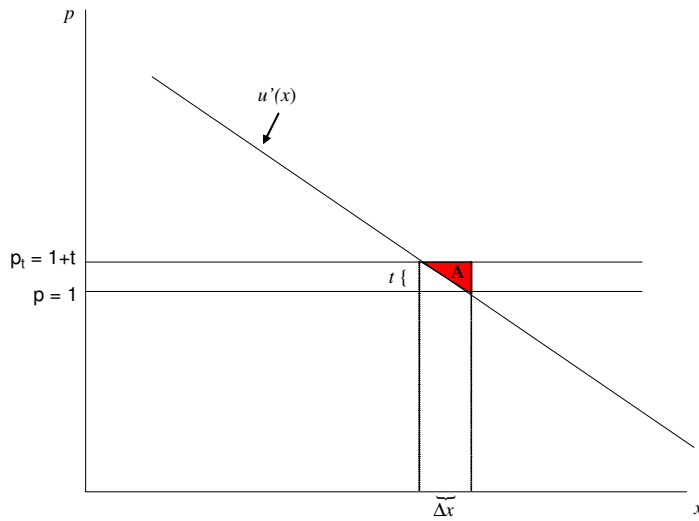
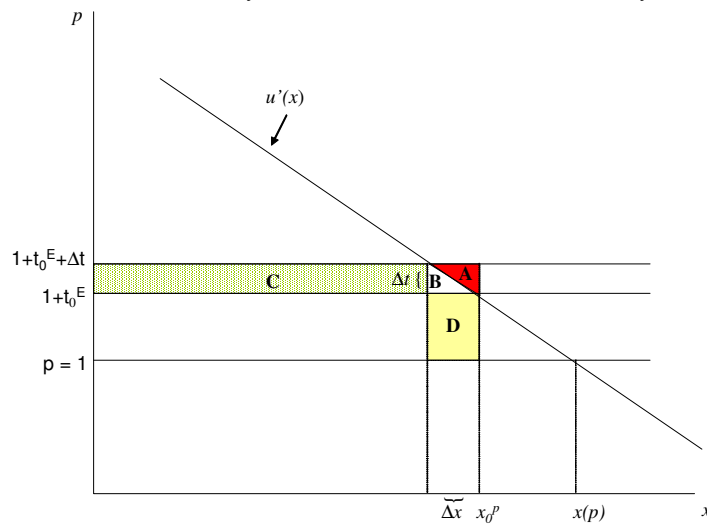


Figure 5b
Efficiency Cost of Taxation with Quasilinear Utility



NOTE—In Figure 5a, region A depicts individual utility loss (G) from ignoring the sales tax. In Figure 5b, region A depicts individual utility loss from ignoring a new sales tax Δt when there is already an excise tax of t_0^E included in the posted price. Region A+B+C (shaded rectangle) depicts change in government revenue ($\Delta R(\theta = 0)$) when the individual does not respond to the introduction of the sales tax. Regions B+D (trapezoid) represent excess burden when the individual does respond to the tax change ($EB^*(\Delta t)$).

Tax-Inclusive Price Tags



← Orig. Tag

Exp.
Tag



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: <input type="checkbox"/> Male <input type="checkbox"/> Female	Age:	Marital Status: <input type="checkbox"/> Married <input type="checkbox"/> Unmarried	Education: <input type="checkbox"/> High School <input type="checkbox"/> College Degree <input type="checkbox"/> Graduate Degree	Years You Have Lived in California:
--	-------------	--	---	--

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 [city]? _____%
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?
<div>< 2%</div> <div>2-10%</div> <div>10-25%</div> <div>25-50%</div> <div>> 50%</div>

Thank you for your time!

