Salience and Taxation with Imperfect Competition

Kory Kroft, Jean-William P. Laliberté, René Leal-Vizcaíno, and Matthew J. Notowidigdo[*]

April 2022

Abstract

This paper studies commodity taxation in a model featuring heterogeneous consumers, imperfect competition, and tax salience. We derive new formulas for the incidence and marginal excess burden of commodity taxation highlighting interactions between tax salience and market structure. We estimate the necessary inputs to the formulas by using Nielsen Retail Scanner and Consumer Panel data covering grocery stores and households in the U.S. and detailed sales tax data. We estimate a large amount of pass-through of taxes onto consumer prices and find that households respond more to changes in prices than taxes. We also estimate significant heterogeneity in tax salience across households. We calibrate our new formulas using these results and conclude that essentially all of the incidence of sales taxes falls on consumers, and the marginal excess burden of taxation is larger than estimates based on standard formulas that ignore imperfect competition and tax salience.

1 Introduction

Standard welfare analysis of commodity taxation typically makes two key assumptions: (1) the product market is perfectly competitive and (2) consumers respond to taxes in the same way they respond to price changes. Several papers in public economics have relaxed the first assumption (see Auerbach and Hines 2002 for a review of this literature), but these papers have maintained the second assumption that taxes are fully salient. More recently, researchers have relaxed the second assumption, developing new theoretical and empirical tools to analyze the welfare effects of taxes when taxes are less salient than prices, but have maintained the assumption of perfect competition (Chetty, Looney, and Kroft 2009; Taubinsky and Rees-Jones 2018; Farhi and Gabaix 2020; Morrison and Taubinsky 2020). If markets are characterized by imperfect competition and consumers misperceive taxes, however, neither of these approaches is likely to provide a fully accurate characterization of the welfare effects of commodity taxes.

In this paper, we derive new formulas for the incidence and marginal excess burden of commodity taxes (both unit taxes and ad valorem taxes) in a model featuring imperfect competition and tax salience with heterogeneous consumers. Using these formulas, we show how tax salience and market structure *interact* when considering tax incidence and the marginal excess burden.

For incidence, we show that greater attention to taxes can increase the incidence on consumers under imperfect competition in contrast to the standard model of perfect competition which predicts the opposite pattern. Thus, the standard intuition of how tax salience affects the incidence of taxation in perfectly competitive markets does not always carry over to imperfect competition. We also derive new results about how heterogeneity in consumer inattention to taxes affects incidence both under perfect competition and imperfect competition. We show that consumer heterogeneity affects pass-through and incidence under all market structures including perfect competition. With imperfect competition, there is an additional effect of heterogeneity on pass-through. Intuitively, when the consumer response to taxes is heterogeneous, this effectively changes the slope of the inverse demand curve facing the firm and firms take this into account when choosing prices. Of particular relevance

1

for firms with market power is how inattention correlates with the price elasticity of demand. We show that firms bear less of the burden of taxes when elastic consumers are more inattentive to taxes. Thus, the covariance between consumer inattention and price elasticity is important for incidence analysis.

Turning to welfare, we find that tax salience and market structure directly interact in the tax formula characterizing the marginal excess burden of taxation. In particular, while the expression for the marginal excess burden includes the additive effects of imperfect competition (via the markup) and tax salience (via the inattention parameter), it also includes the output response to the tax which in turn depends on the degree of inattention to the tax and the output response to prices. While salience has a second-order effect on excess burden under perfect competition when there are no pre-existing taxes (see Chetty, Looney, and Kroft 2009), we show that it has a first-order effect under imperfect competition and scales linearly with the markup. This is important since it highlights that, at least in some circumstances, "behavioral biases" may have a second-order effect on the welfare cost of taxation in competitive markets but a first-order effect on this welfare cost when firms have market power. Moreover, holding fixed market structure, we find that greater attention to taxes (lower "frictions") magnifies the distortionary effect of taxation under imperfect competition. We also find, similar to Taubinsky and Rees-Jones (2018) and Farhi and Gabaix (2020), that heterogeneous inattention to taxes induces misallocation, but that greater dispersion in inattention increases the welfare cost of taxes in similar ways under perfect and imperfect competition; these results extend the heterogeneity results in Taubinsky and Rees-Jones (2018) to the case with continuous demand and imperfect competition.

After presenting our theoretical results, we provide new estimates of all of the necessary inputs to our tax formulas using Nielsen Retail Scanner (RMS) data covering grocery stores selling consumer goods in the U.S., Nielsen HomeScan Consumer Panel (HMS) data covering household purchases, and county-level and state-level sales tax data. Using the RMS data, we estimate the effect of taxes on consumer prices and quantity demanded using a regression model that leverages variation in sales taxes within states and counties over time, and another regression model that focuses on differences

between "border pair" counties located on opposite sides of a state border (Holmes 1998; Dube, Lester and Reich 2010). We also estimate the price elasticity of demand based on an instrumental variable strategy that exploits the "uniform pricing" across stores within retail chains (DellaVigna and Gentzkow 2019). Our estimates indicate nearly-complete pass-through of taxes onto consumer prices and a tax elasticity of demand that is smaller in magnitude than the price elasticity of demand. We combine these estimates to provide a new estimate of tax salience, which is fairly similar to other estimates reported in the literature.

One concern with the RMS estimates is that our tax salience parameter estimate will be biased if there is meaningful cross-border shopping and other substitution across stores within a county. Intuitively, tax changes affect all stores within a county, while the "uniform pricing" instrument that we use only affects certain stores within a county that are part of the same retail chain, which may cause some households to substitute across stores in a way that is different from how the same households respond to tax changes. This would cause our price elasticity and tax elasticity estimates to diverge for reasons unrelated to tax salience. To investigate this potential bias, we replicate all of our main results in the HMS data, and, reassuringly, we find very similar tax elasticity and price elasticity estimates in the HMS data compared to the RMS data, once we adjust for differences in coverage between the two datasets. This suggests limited bias from cross-store substitution within and between counties.

We also use the HMS data to directly estimate household-level heterogeneity in tax salience using two complementary empirical approaches. First, we estimate models which allow the price elasticity and tax elasticity to vary with household-level demographic characteristics such as age, income, and education. We find strong evidence of statistically and economically significant heterogeneity in tax salience along these observable dimensions. Second, we estimate a mixed-effects model that allows for random coefficients for both the tax and price elasticity. We find an even larger estimated variance in the tax salience parameter using this richer model, which suggests an important role for both observable and unobservable differences across households in driving heterogeneity in consumer inattention. In our preferred specification, we find a mean tax salience parameter of about 30 percent

(implying that households on average respond to taxes only about 30 percent as much as they respond to prices), with a two-standard-deviation range around the mean of about 45 percentage points.

Lastly, we calibrate our new tax formulas using these empirical estimates. A novel feature of our approach is the use of the pass-through formula and the generalized Lerner index to calibrate the average markup, which enters in the marginal excess burden formula. We then consider two types of calibration exercises. First, we quantify the discrepancy if one "naively" computed the welfare cost of taxation using the standard Harberger formula (Harberger 1964) that ignores salience effects and assumes perfect competition. It is an empirical question whether the actual welfare cost is larger or smaller than the standard Harberger benchmark since inattention to taxes reduces it and imperfect competition increases it. Our results show that the standard Harberger formula *understates* the marginal excess burden of taxes. Second, we calibrate the excess burden of taxation in counterfactual scenarios that increase the salience of sales taxes and change the market structure taking into account the endogeneity of the output response to the tax and pass-through with respect to tax salience and market structure. Our analysis of these scenarios reveal various interactions between tax salience and imperfect competition in determining the pass-through of taxes onto consumer prices, the effect of taxes on quantity demanded, and ultimately the incidence and marginal excess burden. For example, we find that the change in the marginal excess burden from greater tax salience is larger under imperfect competition compared to perfect competition.

Our paper is related to several streams of research. First, our paper builds on and contributes to the literature on taxation and imperfect competition (see, e.g., Seade 1987; Stern 1987; Delipalla and Keen 1992; Anderson, de Palma, and Kreider 2001a; Anderson, de Palma, and Kreider 2001b; Auerbach and Hines 2001; Weyl and Fabinger 2013; Hackner and Herzing 2016; Miravete, Seim, and Thurk 2018; Adachi and Fabinger 2019).[1] Our paper innovates in several ways. First, we consider a general model of imperfect competition and do not impose a functional form for preferences or

---

[1]We focus on interaction between imperfect competition and tax salience but ignore firm heterogeneiety. Adachi and Fabinger (2019) consider the welfare and incidence effects of taxation with heterogeneous firms while assuming that consumers perfectly optimize.

technology, similar to Weyl and Fabinger (2013).[2] Second, we permit consumers to underreact to taxes and allow for heterogeneity in the degree of underreaction. Third, we derive our new formulas for both ad valorem and unit taxes (allowing for tax salience) and compare these formulas, which is important since existing theoretical work finds that these taxes are not equivalent under imperfect competition (Delipalla and Keen 1992). Lastly, we provide an empirical application that allows us to calibrate our new formulas, which contributes to the literature studying sales taxes empirically (see, e.g., Besley and Rosen 1999; Einav et al. 2014; Baker, Johnson, and Kueng 2018).

We also contribute to the literature in behavioral public economics (Liebman and Zeckhauser 2004; Chetty, Looney, and Kroft 2009; Goldin and Hominoff 2013; Allcott and Taubinsky 2015; Taubinsky and Rees-Jones 2018; Allcott, Lockwood, and Taubinsky 2018; Bradley and Feldman 2019; Farhi and Gabaix 2020; Morrison and Taubinsky 2020).[3] Most of the papers in this literature assume perfect competition. Bradley and Feldman (2019), which examines tax incidence in a monopoly setting with inattentive consumers, is an important exception. Relative to this paper, we allow for heterogeneity in tax salience across consumers, allow for more general forms of imperfect competition, and move beyond incidence to also study the efficiency cost of taxation. The joint consideration of incidence and efficiency analysis is important for our calibration approach, which combines both tax formulas to identify the markup which appears in the marginal excess burden formula.

The remainder of the paper is organized as follows: Section 2 begins with a model of imperfect competition and inattention and considers the welfare and incidence effects of a unit tax. Section 3 discusses the data and the empirical results. Section 4 presents the calibration results. Section 5 concludes.

---

[2]Weyl and Fabinger (2013) only consider tax incidence. They do not consider the efficiency costs of taxation.

[3]Our paper is also broadly related to other studies of salience and consumer choice such as Bordalo, Gennaioli, and Shleifer (2013). Our paper focuses specifically on the salience of taxes compared to posted prices, while Bordalo, Gennaioli, and Shleifer (2013) study salience of other attributes of consumer products.

2 Theory

We consider a differentiated product market (the "inside market") which is subject to a unit tax t on each product in the market.[4] Following Auerbach and Hines (2001) and Weyl and Fabinger (2013), we assume that markets for other goods are perfectly competitive and are not subject to taxation. There is a mass 1 of consumers each indexed by i with exogenous income Z^i. For each i, preferences are given by the quasilinear utility function $u^i(q_1, \ldots, q_J) + y$, where q_j is the quantity consumed of product $j = 1, \ldots, J$ and $y \in \mathbb{R}$ is the numeraire (representing consumption in all the outside markets). We assume that the subutility function, u^i, which represents preferences for the differentiated products, is strictly quasi-concave, twice differentiable, and symmetric in all of its arguments. The pre-tax (or producer) price for product j is given by p_j and the after-tax (or consumer) price is given by $p_j + t$ for all $j = 1, ..., J$. We define $u^i(Q^i) \equiv u^i(Q^i/J, \ldots, Q^i/J)$ to be the compact notation of utility for the symmetric case where individual i consumes $q^i = \frac{Q^i}{J}$ units of each product $j = 1, \ldots, J$, where Q^i is the aggregate quantity consumed by individual i. Throughout we assume that the number of products, J, is exogenous. In a companion paper, we allow for endogenous entry and exit of firms (see Kroft et al 2021).

Following Chetty, Looney, and Kroft (2009), consumer i's demand for product j is given by $q_j^i = q_j^i(p_1, \ldots, p_J, t)$ which is a function of both pre-tax prices and the tax. In order to connect our tax formulas to empirical objects, it is necessary to relate observed demand $q_j^i(p_1, \ldots, p_J, t)$ to consumer willingness to pay. We thus make the following assumptions which mirror assumptions $A1$ and $A2$ in Chetty, Looney, and Kroft (2009).

Assumption 1. *Taxes affect utility only through their effects on the chosen consumption bundle. Indirect utility is given by:*

$$V^i(p_1, \ldots, p_J, t, Z_i) = u^i(q_1^i(p_1, \ldots, p_J, t), \ldots, q_J^i(p_1, \ldots, p_J, t)) + Z^i - (p_1 + t)q_1^i - \cdots - (p_J + t)q_J^i$$

Assumption 1 requires that taxes or salience have no impact on utility beyond their effects on

---

[4]In the Online Appendix we derive results for the case of an ad valorem tax and compare unit and ad valorem taxes.

consumption. This assumption may be violated if, for example, attention is costly, as emphasized by theories of rational inattention (Gabaix 2019) and recent empirical evidence (Morrison and Taubinksy 2020). If there is a cost to processing non-salient taxes, then Assumption 1 is violated and one must additionally account for these information costs when examining the efficiency costs of taxation (Chetty, Looney, and Kroft 2007).

Assumption 2. *When tax-inclusive prices are fully salient, the agent chooses the same allocation as a fully-optimizing agent.*

$$(q_1^i, \ldots, q_J^i)(p_1 + t, \ldots, p_J + t, 0) = \arg\max_{(q_1, \ldots, q_J)} u^i(q_1, \ldots, q_J) + Z^i - (p_1 + t)q_1 - \cdots - (p_J + t)q_J$$

Assumption 2 implies that when tax-inclusive prices are fully salient, agents maximize utility. Our framework allows for salience effects by considering the possibility that $q_j^i(p_1, ..., p_J, 0) > q_j^i(p_1, ..., p_J, t) > q_j^i(p_1 + t, ..., p_J + t, 0)$.

In what follows, we assume that the demand function $q_j^i(\cdot)$ is symmetric in all other prices which we denote by $(p_k)_{-j}$ and twice differentiable and denote by $q^i(p, t)$ the demand function for individual i corresponding to symmetric prices and J firms: $q^i(p, t) \equiv q_j^i(p, ..., p, t)$. Without loss of generality on the functional form of $q^i(\cdot, 0) = \frac{(u_i')^{-1}(\cdot)}{J}$, and assuming $q^i(p, t) = q^i(p + \theta_i t, 0)$ for some $\theta_i > 0$, the salience parameter satisfies $\theta_i = \frac{\frac{\partial q_j^i}{\partial t}}{\frac{\partial q_j^i}{\partial p}}$ and is the same for all products j for individual i. This specification of inattention rules out the possibility that θ_i may depend on producer prices as suggested by the empirical evidence in Morrison and Taubinsky (2020). Relaxing this assumption would be a useful extension, but is beyond the scope of this paper. We adopted this assumption primarily for tractability to allow us to focus on interactions with imperfect competition. In our empirical analysis below, we briefly explore how $\mathbb{E}(\theta_i)$ varies between high-price and low-price products, and we do not find evidence that $\mathbb{E}(\theta_i)$ varies with average price levels in our data.

We define individual i's market demand as $Q^i(p, t) = Jq^i(p, t)$. Total market demand is given by $Q(p, t) = \int Q^i(p, t)di$, from where we define the market demand elasticity $\epsilon_D \equiv -\frac{\partial Q(p,t)}{\partial p} \frac{p+t}{Q}$ and $\epsilon_{Dt} \equiv \frac{dq(t)}{dt} \frac{p+t}{q(t)}$ as the elasticity of equilibrium output, $q(t) \equiv Q(p(t), t)$, with respect to the tax t. Note that ϵ_{Dt} need not equal $\frac{\partial Q(p,t)}{\partial t} \frac{p+t}{Q}$; the latter holds the pre-tax price, p, fixed, while the former includes

any indirect effect of taxes on the producer price. The pass-through rate is denoted by $\rho \equiv 1 + dp/dt$. Also, for an economy without taxes, we define the representative agent's willingness to pay $wtp(Q)$ as the inverse of $Q(\cdot, 0)$, and let $mwtp(Q) = wtp'(Q)$ be the marginal willingness to pay.[5] Then $ms(Q) = -mwtp(Q)Q$ is the marginal consumer surplus and the elasticity of inverse marginal surplus is given by $\epsilon_{ms} \equiv \frac{ms(Q)}{ms'(Q)Q}$. Furthermore, define $MS(Q,t) = -\frac{Q}{\frac{\partial Q}{\partial p}(p(t),t)} = \frac{ms(Q)}{mwtp(Q(t))*\frac{\partial Q}{\partial p}(p(t),t)}$, then $MS(Q,0) = ms(Q)$, and let $MS_t \equiv \partial MS$.

Let $q_j(p_1, ..., p_J, t) = \int q_j^i(p_1, ..., p_J, t)di$. On the supply side, we allow for different forms of competition by introducing the market conduct parameter $\nu_p = \frac{\partial p_k}{\partial p_j}$ ($k \neq j$) following Weyl and Fabinger (2013). Assume each firm produces a single product and has a homogeneous cost function $c_j(q_j) = c(q_j)$, where $c(\cdot)$ is increasing and twice differentiable with $c(0) = 0$ and $mc(q_j) \equiv c'(q_j)$. Firm j chooses p_j to maximize profits π_j:

$$\max_{p_j} \pi_j = p_j q_j(p_1 \ldots, p_J, t) - c(q_j(p_1 \ldots, p_J, t))$$

$$\text{s.t. } \frac{\partial p_k}{\partial p_j} = \nu_p \text{ for } k \neq j$$

The first-order condition for p_j is given by:

$$q_j + (p_j - mc(q_j)) \left(\frac{\partial q_j}{\partial p_j} + \nu_p \sum_{k \neq j} \frac{\partial q_j}{\partial p_k} \right) = 0.$$

In a symmetric equilibrium, $p_j = p$ solves:

$$q_j(p_j, p, \ldots, p, t) + (p_j - mc(q_j)) \left(\frac{\partial q_j(p_j, p, \ldots, p, t)}{\partial p_j} + (J-1)\nu_p \frac{\partial q_j(p_j, p, \ldots, p, t)}{\partial p_k} \right) = 0, k \neq j$$

We assume that $\frac{\partial \pi_j}{\partial p_j}(p_j, p)$ is strict single crossing (from above) in p_j and decreasing in p so that a unique symmetric equilibrium $p(t)$ exists.[6] By letting $\nu_q = \frac{1}{mwtp(Q)} \times \frac{1}{\frac{dq_j}{dp_j}} = \frac{1}{mwtp(Q)} \times \frac{1}{\frac{\partial q_j}{\partial p_j} + \nu_p \sum_{k \neq j} \frac{\partial q_j}{\partial p_k}}$

[5]Formally, there is no representative agent for the economy since even with quasilinear utility there is a problem of aggregation given the misoptimization with respect to taxes. However, when $t = 0$ the economy admits a representative agent (given that there are no income effects) and we use the inverse demand function of this representative agent to characterize average and marginal consumer surplus.

[6]The case of strategic complementarities, where $\frac{\partial \pi_j}{\partial p_j}(p_j, p)$ is increasing in p allows for the existence of multiple symmetric equilibria. However, in that case if we assume there is a continuous and symmetric equilibrium selection $p(t)$ the same results follow.

we can rewrite the first-order condition as a generalized Lerner index:

$$\frac{p - mc(q)}{p + t} = \frac{\nu_q}{J\epsilon_D} \tag{1}$$

Setting $\nu_q = J$ yields the monopoly (perfect collusion) outcome and setting $\nu_q = 0$ gives the perfect competition (marginal cost pricing) solution. Setting $\nu_q = 1$ corresponds to Cournot competition when goods are homogeneous and setting $\nu_p = 0$ yields the Bertrand-Nash equilibrium. The model thus captures a wide range of market conduct.

We assume that tax revenue $R = tQ$ and profits $J\pi$ are redistributed to the consumers as a lump-sum transfer. Consumers treat profits and tax revenue as fixed when choosing consumption, failing to consider the external effects on the lump-sum transfer. Given the assumption of quasilinear utility, consumers will choose to allocate the lump-sum transfer to the outside market y. Thus, total welfare, W, is given by the sum of consumer surplus (CS), producer surplus (PS) and government revenue (R).

$$W(p,t) = \underbrace{\int u^i(Q^i(p,t))di - (p+t)Q(p,t)}_{CS} + \underbrace{pQ - Jc(q)}_{PS} + \underbrace{tQ}_{R}$$

Following Weyl and Fabinger (2013), we define the incidence of a unit tax t as $I = \frac{dCS/dt}{dPS/dt}$ and the marginal excess burden as $\frac{dW}{dt}$. We begin by introducing a technical assumption which helps to simplify the analysis throughout.

Assumption 3. *The demand function $Q^i(p,t)$ can be represented by the linear approximation $\hat{Q}^i(p,t) = q_0^i + \frac{\partial Q(p_0,t_0)}{\partial p}(p - p_0 + \theta_i(t - t_0))$ around $(q_0^i, p_0, t_0, \theta_i)$ for $q_0^i = Q^i(p_0, t_0)$ for each individual i.*[7]

Assumption 3 is a formal statement of the approximation given in Bernheim and Taubinsky (2018, page 438) that allows us to focus on heterogeneity in salience effects across consumers, holding the price responses across consumers constant. Note that assuming price responses are the same across consumers at all quantity levels is not sufficient for Assumption 3 to hold exactly unless demand curves are linear. We next introduce a lemma which is quite useful for deriving the incidence formulas

---

[7]In case this assumption is violated, the model given by $\hat{Q}^i(p,t)$ is a linear approximation to the real model with a common slope for all i. The corollary that follows below applies to this linear approximation.

that we present in the paper.

Lemma 1. *The following relationship holds between the demand elasticities, pass-through and inattention to taxes:*

$$\epsilon_{Dt} = -(\mathbb{E}(\theta_i) + \rho - 1)\epsilon_D + \frac{p+t}{Q(t)}Cov\left(\theta_i, \frac{\partial Q^i(p,t)}{\partial p}\right)$$

Under Assumption 3, we obtain the following relationship:

$$\epsilon_{Dt} = -(\mathbb{E}(\theta_i) + \rho - 1)\epsilon_D$$

Proof. See Online Appendix. □

We can now state our main theoretical results. Consider a small increase in the tax t which applies to all goods in the inside market.

Proposition 1. *The incidence on consumers, producers, government, the pass-through rate and the marginal excess burden of a unit tax, t, may be expressed as:*

$$\frac{dCS}{dt} = -\rho Q - (1 - \mathbb{E}(\theta_i))t\frac{dQ(p(t),t)}{dt} + tCov\left(\theta_i, \frac{dQ^i(p(t),t)}{dt}\right)$$

$$\frac{dPS}{dt} = -\left(1 - \frac{\nu_q}{J}\right)[Q(1-\rho)] - \frac{\nu_q}{J}\left[Q\left(\mathbb{E}(\theta_i) + \frac{Cov\left(\theta_i, \frac{\partial Q^i}{\partial p}\right)}{\frac{\partial Q}{\partial p}}\right)\right]$$

$$\frac{dR}{dt} = Q + t\frac{dQ(p(t),t)}{dt}$$

$$\rho = 1 - (1-\omega)\left(\mathbb{E}(\theta_i) + \frac{Cov\left(\theta_i, \frac{\partial Q^i}{\partial p}\right)}{\frac{\partial Q}{\partial p}}\right) + \omega\frac{\nu_q}{J}MS_t \text{ where } \omega = \frac{1}{1 + \frac{\epsilon_D \frac{p}{p+t} - \frac{\nu_q}{J}}{\epsilon_S} + \frac{\frac{\nu_q}{J}}{\epsilon_{ms}}}$$

$$I = \frac{\rho + (1 - \mathbb{E}(\theta_i))\frac{t}{p+t}\epsilon_{Dt} - \frac{t}{Q}Cov\left(\theta_i, \frac{dQ^i(p(t),t)}{dt}\right)}{(1-\rho)\left(1 - \frac{\nu_q}{J}\right) + \frac{\nu_q}{J}\frac{\mathbb{E}\left(\theta_i\frac{\partial Q^i}{\partial p}\right)}{\mathbb{E}\left(\frac{\partial Q^i}{\partial p}\right)}}$$

$$\frac{dW}{dt} = (p - mc(q) + \mathbb{E}(\theta_i)t)\frac{dQ(p(t),t)}{dt} + tCov\left(\theta_i, \frac{dQ^i(p(t),t)}{dt}\right)$$

Proof. See Online Appendix. □

Moreover, we can make use of Assumption 3 to derive simplified versions of the formulas in

Proposition 1 which only make use of $Var\left(\theta_i\right)$ (instead of the covariance between θ_i and other demand parameters) and will be used in the calibrations below.

Corollary 1. *Under Assumption 3, the effect of the tax on consumer surplus, producer surplus, pass-through, incidence and welfare can be expressed as:*

$$\frac{dCS}{dt} = -\rho Q - (1 - \mathbb{E}(\theta_i))t\frac{dQ(p(t), t)}{dt} + tVar\left(\theta_i\right)\frac{\partial Q}{\partial p}$$

$$\frac{dPS}{dt} = -Q\left[\left(1 - \frac{\nu_q}{J}\right)(1 - \rho) + \frac{\nu_q}{J}\mathbb{E}(\theta_i)\right]$$

$$\rho = 1 - (1 - \omega)\,\mathbb{E}(\theta_i),\ where\ \omega = \frac{1}{1 + \frac{\epsilon_D\frac{p}{p+t} - \frac{\nu_q}{J}}{\epsilon_S} + \frac{\frac{\nu_q}{J}}{\epsilon_{ms}}}$$

$$I = \frac{\rho + (1 - \mathbb{E}(\theta_i))\frac{t}{p+t}\epsilon_{Dt} - \frac{t}{p+t}Var\left(\theta_i\right)\epsilon_D}{(1 - \rho)\left(1 - \frac{\nu_q}{J}\right) + \frac{\nu_q}{J}\mathbb{E}(\theta_i)}$$

$$\begin{aligned}\frac{dW}{dt} &= (p - mc(q) + \mathbb{E}(\theta_i)t)\frac{dQ(p(t), t)}{dt} + tVar\left(\theta_i\right)\frac{\partial Q}{\partial p}\\ &= (p - mc(q) + \mathbb{E}(\theta_i)t)\frac{\partial Q}{\partial p}(\rho - 1 + \mathbb{E}(\theta_i)) + tVar\left(\theta_i\right)\frac{\partial Q}{\partial p}\end{aligned}$$

Proof. See Online Appendix. ☐

Proposition 1 characterizes the consumer and producer burden, pass-through, incidence, and the marginal excess burden of a unit tax. There are several lessons that emerge. First, the expression for $\frac{dCS}{dt}$ does not depend directly on market structure, except insofar as market conduct determines pass-through and the behavioral response to the tax. When $t = 0$, the expression for the consumer burden is identical to the one in Weyl and Fabinger (2013), except that the pass-through term, ρ, is indirectly affected by salience effects. Intuitively, on the consumer side, when there are no taxes in the baseline equilibrium, consumers optimize and so the envelope theorem applies. Salience only affects consumers at the market level through changes in prices, as in Chetty, Looney, and Kroft (2009).

On the other hand, when $t > 0$, the effect of a change in the tax on consumer surplus depends on two additional terms, $-(1 - \mathbb{E}(\theta_i))t\frac{dQ}{dt}$ and $tCov\left(\theta_i, \frac{dQ^i}{dt}\right)$. In this case, one has to account for behavioral responses to the tax since the envelope theorem does not apply when consumers misopti-

mize in the baseline equilibrium. The first term, $-(1 - \mathbb{E}(\theta_i))t\frac{dQ}{dt}$, relates to the "internality wedge" in Bernheim and Taubinsky (2018) and Farhi and Gabaix (2020). It enters $\frac{dCS}{dt}$ positively whenever $\mathbb{E}(\theta_i) < 1$ and $\frac{dQ}{dt} < 0$ indicating that more inattention to taxes reduces the incidence on consumers, conditional on the pass-through rate and the behavioral response to the tax. Intuitively, if consumers are overspending on taxable goods at baseline (because $\mathbb{E}(\theta_i) < 1$), then a tax increase that causes them to reduce their demand brings them closer to their privately optimal choice. In this case, the degree of over-estimation of marginal utility is given by $(1 - \mathbb{E}(\theta_i))t$ and one can think of the welfare effects of tax salience as a Pigouvian correction. The second term, $tCov\left(\theta_i, \frac{dQ^i}{dt}\right)$, relates to the misallocation term in Taubinsky and Rees-Jones (2018). Our results extend the heterogeneity results in Taubinsky and Rees-Jones (2018) to the case with continuous demand and imperfect competition. Finally, Corollary 1 shows that under Assumption 3, $\frac{dCS}{dt} = -\rho Q - (1 - \mathbb{E}(\theta_i))t\frac{dQ(p(t),t)}{dt} + tVar\left(\theta_i\right)\frac{\partial Q}{\partial p}$ which will be useful in our calibrations below which incorporate our empirical estimate of $Var\left(\theta_i\right)$.

Second, the expression for $\frac{dPS}{dt}$ is a convex combination of the monopoly and perfect competition cases with weights $\frac{\nu_q}{J}$ and $1 - \frac{\nu_q}{J}$, respectively. Under perfect competition, $\frac{\nu_q}{J} = 0$ and $\frac{dPS}{dt} = -Q(1 - \rho)$ and the producer burden depends on the price effect, which in turn depends on the degree of salience, similar to Chetty, Looney, and Kroft (2009). Under monopoly, $\frac{\nu_q}{J} = 1$ and $\frac{dPS}{dt} = -Q\left(\mathbb{E}(\theta_i) + \frac{Cov\left(\theta_i, \frac{\partial Q^i}{\partial p}\right)}{\frac{\partial Q}{\partial p}}\right)$. To interpret this expression, it is instructive to consider the case where $\theta_i = 1$ for all i. Since the monopolist sets the price and equivalently the level of output, the effect of a small change in taxes is simply the mechanical effect of the tax change which is given by output, Q. Consumer inattention attenuates the reduction in consumer demand due to the tax increase on average by the amount $\mathbb{E}(\theta_i)$. The covariance term $Cov\left(\theta_i, \frac{\partial Q^i}{\partial p}\right)$ incorporates the correlation between θ_i and $\frac{\partial Q^i}{\partial p}$ which additionally determines the market-level demand response to the tax. When $Cov\left(\theta_i, \frac{\partial Q^i}{\partial p}\right) > 0$, the incidence on the monopolist is attenuated.[8] This can be easily seen in the binary case where there are two types of consumers: those who optimize perfectly ($\theta_i = 1$) and those who are fully inattentive to taxes ($\theta_i = 0$). If those who optimize are price inelastic and those who

---

[8]Since $\frac{\partial Q^i}{\partial p} < 0$, this requires that consumers that are attentive to taxes are price inelastic; in other words, the absolute value of $\frac{\partial Q^i}{\partial p}$ is negatively correlated with θ_i.

are inattentive are price elastic, then the monopolist earns higher profit compared to the case where inattention is uncorrelated with price elasticity. In fact, it may be optimal for the monopolist to fully disclose taxes (e.g., post tax-inclusive prices) if there are enough consumers who are both highly price elastic and overreact to taxes (so that $\theta_i > 1$).[9] This result on optimal disclosure of taxes relates to Veiga and Weyl (2016) on how firms can optimally use nonprice product features to sort profitable from unprofitable consumers.

Turning to pass-through, as is well-known from Chetty, Looney, and Kroft (2009), $\omega < 1$ with perfect competition and an increase in $\mathbb{E}(\theta_i)$ leads to a lower pass-through and incidence on consumers. In the presence of heterogeneous consumers, pass-through additionally depends on consumer heterogeneity as reflected in the term $Cov\left(\theta_i, \frac{\partial Q^i}{\partial p}\right)$.[10] Intuitively, what matters for incidence is the initial shift in demand in response to taxes and the price elasticities of demand and supply which determine how much prices need to adjust to re-equilibrate the market. Since individual-level responses to taxes can be written as $\frac{\partial Q^i}{\partial t} = \theta_i \frac{\partial Q^i}{\partial p}$, the market-level response to taxes depends on the covariance between θ_i and $\frac{\partial Q^i}{\partial p}$ across i. Under imperfect competition, there are interesting effects of salience on pass-through, ρ, which operate through the elasticity of inverse marginal surplus, which is positive (negative) if demand is log convex (log concave). In particular, under imperfect competition, it is theoretically possible that $\omega > 1$ which implies that an increase in $\mathbb{E}(\theta_i)$ raises incidence on consumers. To see this, consider the case of monopoly with constant marginal cost and suppose demand has constant pass-through form so that $\epsilon_{ms} = -\epsilon$ (Bulow and Pfleiderer 1983) and $\theta_i = \theta$. Under these assumptions, $\rho = 1 - \frac{\theta}{1-\epsilon}$ so that $\frac{d\rho}{d\theta} = \frac{1}{\epsilon-1}$, and so if demand is sufficiently convex, then $\frac{d\rho}{d\theta} > 0$ and increased attention to the tax makes consumers *worse off,* in contrast to the logic in Chetty, Looney, and Kroft (2009) under perfect competition.[11] Intuitively, the monopolist faces a demand curve of

---

[9]Morrison and Taubinsky (2020) find that some consumers overreact to taxes but do not investigate whether this is correlated with their price elasticity of demand. To see why disclosure is never optimal when $\theta_i < 1$ for all consumers, consider the case where the monopolist discloses taxes at some Q^*. If the monopolist then shrouds taxes, it could still sell Q^* but at a higher price since the inverse demand curve with hidden taxes lies *everywhere* above the inverse demand curve with salient taxes. Thus, there is a profitable deviation and so disclosure can never be optimal when all consumers are inattentive to taxes.

[10]This result was derived for the case of binary demand in Taubinsky and Rees-Jones (2018); see page 9 of the Appendix.

[11]Note that even when $\epsilon_S = \infty$, the full incidence is not on consumers, unlike the case of perfect competition, although

the form $P(q,t) = wtp(q) - \theta t$. This is a standard monopoly problem where θt can be interpreted as the "effective tax", and so a larger θ implies a larger effective tax. This implies that the change in the producer price is simply the change in the producer price in the standard case scaled by the amount of inattention, e.g., $\theta(\omega - 1)$. On the other hand, the consumer price is given by $P + t$ and pass-through is $1 + \frac{dP}{dt} = 1 + \theta(\omega - 1)$. Thus, if $\omega - 1 > 0$ (as would be the case with $\epsilon > 1$), greater attention to taxes amplifies the pass-through effect. By contrast, when $\omega - 1 < 0$ (as would be the case with linear demand where $\omega = 1/2$), greater attention to taxes leads to a greater cut in producer prices.[12] In both cases, greater attention to taxes increases the effective tax and this reinforces either an increase in the producer price in the case of overshifting or a decrease in the producer price in the case of undershifting.

We also see that the expression for ρ in the general case of imperfect competition depends on MS_t. Up to first order, this term can be approximated by $MS_t \approx \frac{-q}{\left(\frac{\partial Q}{\partial p}\right)^2} Cov\left(\frac{\partial^2 Q^i}{\partial p^2}, \theta_i\right)$ (see Online Appendix). This new term captures that when taxes change and consumers vary in their degree of inattention, this effectively changes the slope of the demand curve. Since the optimal price depends on the slope of the demand curve, firms exploit this change in market power when re-optimizing prices. If more attentive consumers become more price elastic when taxes change, then $Cov\left(\frac{\partial^2 Q^i}{\partial p^2}, \theta_i\right) < 0$ and $MS_t > 0$. Intuitively, when the tax increases there is a reallocation of demand, whereby the negative output response Q_i is bigger (in absolute value) for more attentive and price elastic consumers; in the case where $Cov\left(\frac{\partial^2 Q^i}{\partial p^2}, \theta_i\right) < 0$, the average (or market) demand becomes more inelastic as demand is reallocated to more inelastic and less attentive consumers. Therefore, pass-through increases ($MS_t > 0$). Depending on the magnitude of MS_t, it is possible to get overshifting of taxes onto consumer prices, even when the standard model predicts undershifting.

The relative incidence on consumers and firms is given by I. When $t = 0$ and markets are perfectly competitive ($\nu_q = 0$), we obtain the classical result that incidence is determined purely by

---

we note that this result holds independent of salience effects.

[12]It can also be shown that $\omega = \frac{P'(q)}{MR'(q)}$, so that $\omega > 1$ implies that the inverse demand curve is steeper than the marginal revenue curve. This case is shown in Online Appendix Figure OA.1, which illustrates graphically how greater attention to taxes can amplify the pass-through effect.

the price effect of the tax, $I = \frac{\rho}{1-\rho}$. Similarly, under monopoly ($\nu_q = J$), incidence is $I = \rho$ as shown by Weyl and Fabinger (2013). In the general case of inattentive consumers under Assumption 3, $I = \frac{\rho + (1 - \mathbb{E}(\theta_i)) \frac{t}{p+t} \epsilon_{Dt} - \frac{t}{p+t} Var(\theta_i) \epsilon_D}{(1-\rho)\left(1 - \frac{\nu_q}{J}\right) + \frac{\nu_q}{J}\mathbb{E}(\theta_i)}$. As discussed above, with pre-existing taxes, taxes play a corrective role and the numerator accounts for this, while the denominator is a weighted-average of perfect competition and monopoly. From this expression we see that pass-through is no longer sufficient for incidence analysis.

Finally, conditional on $\frac{dQ}{dt}$, the effects of salience on the marginal excess burden of the tax operate in similar ways under different market structures through the terms $\mathbb{E}(\theta_i)t$ and $tCov\left(\theta_i, \frac{dQ^i}{dt}\right)$. However, under imperfect competition the marginal excess burden depends additionally on the markup, $p - mc(q)$. This implies that changes in the degree of inattention to taxes have larger effects on excess burden in imperfectly competitive markets as compared to perfectly competitive markets. To see this, note that we can express $\frac{dQ}{dt} = \frac{\partial Q}{\partial p}(\rho - 1 + \mathbb{E}(\theta_i))$ under Assumption 1. Thus, salience enters linearly in the welfare formula $\frac{dW}{dt}$ and interacts directly with the markup. For example, with full pass-through and homogeneous consumers, $\frac{dW}{dt} = (p - mc(q))\theta\frac{\partial Q}{\partial p} + \theta^2 t \frac{\partial Q}{\partial p}$. Even when there are no pre-existing taxes, there is still a pre-existing distortion due to imperfect competition and thus introducing a small tax into the market has a first-order effect on welfare which scales with the degree of inattention to taxes. This shows that even when "behavioral biases" have a second-order effect on social welfare in competitive markets, they may have a first-order effect on welfare in markets where firms have market power. Holding fixed the markup, however, as frictions are reduced (i.e., increasing the value of θ), the excess burden of taxation is increased. In the calibrations below, we will impose Assumption 3 and calibrate $\frac{dW}{dt} = (p - mc(q) + \mathbb{E}(\theta_i)t)\frac{\partial Q}{\partial p}(\rho - 1 + \mathbb{E}(\theta_i)) + tVar(\theta_i)\frac{\partial Q}{\partial p}$.

To summarize, the analysis of the incidence and welfare consequences of a tax for the general case of imperfect competition suggests that the standard intuition for the case of perfect competition does not always apply when firms have market power. Instead, there are interesting interactions between tax salience and market structure. While the theoretical results in this section focus on unit taxes, we present a more general model with both unit taxes and ad valorem taxes in the Online Appendix. The

more general model allows us to derive analogous formulas for ad valorem taxes, which is useful since our empirical analysis below is based on sales taxes which are ad valorem sales taxes. Additionally, the more general model allows us to derive new results that compare the pass-through rate and the marginal cost of public funds (MCPF) between unit taxes and ad valorem taxes when consumers may under-react to both types of taxes. We show that if consumers under-react to ad valorem and unit taxes similarly (i.e., $\theta_t = \theta_\tau$, where θ_τ refers to the tax salience parameter for ad valorem tax, τ), then the pass-through rate will be lower for ad valorem taxes than unit taxes, which is consistent with existing results in the literature that ignore tax salience (Delipalla and Keen 1992; Adachi and Fabinger 2019). However, if consumers under-react more to ad valorem taxes than unit taxes, then the pass-through rate of ad valorem taxes can be higher, even under perfect competition. Turning to the MCPF, we show that if consumers under-react to ad valorem and unit taxes similarly, then we maintain the well-known result that ad valorem taxes are more efficient than unit taxes under imperfect competition. If consumers are less inattentive to ad valorem taxes than unit taxes, then this reinforces the result that ad valorem taxation is more efficient.

3 Data and Estimation

3.1 Data Description

Nielsen Retail Scanner (RMS) and Nielsen Consumer Panel (HMS) Data   We measure store-level prices and quantity using the Nielsen Retail Scanner (RMS) data from $2006 - 2014$, and complement these with household-level spending data from the HomeScan Consumer Panel (HMS) data.

RMS records sales and the number of units sold per week for roughly 2.5 million products which are designated as Universal Product Codes (UPC) for 35,000 stores in the United States (excluding Hawaii and Alaska) that are part of roughly 90 retail chains. We use these data, combined with sales tax data, primarily to estimate pass-through as well as price and tax elasticities of demand. Stores are assigned to one of five possible store types: grocery, drug, mass merchandise, convenience, and

liquor stores. Each store has a "parent company" that corresponds to the company that owns the store, and the data also indicates when multiple stores are part of the same retail chain. We limit our sample to grocery stores for two reasons. First, the distribution of store types varies considerably across counties. By focusing on one store type, we ensure that compositional differences across counties are not driving our results. Second, we use an instrumental variables strategy which relies on uniform pricing within retail chains following DellaVigna and Gentzkow (2019). There are too few retail chains for non-grocery stores, making this strategy infeasible for these store types. We follow DellaVigna and Gentzkow in further restricting our sample to (1) stores that belong to the same retail chain throughout $2006 - 2014$, (2) stores that are present in the data for at least two years, and (3) stores that belong to retail chains that were associated with the same parent company throughout the sample period.

HMS includes trip-level purchases for over 60,000 households each year, as well as household-level demographic information on an annual basis (e.g. income, education, presence of children). We use these data to estimate tax and price elasticities of demand, and to quantify the degree of heterogeneity in inattention to taxes.[13]

In both RMS and HMS, products (UPCs) are organized by Nielsen according to a hierarchical structure.[14] At the lowest rung are approximately 1,200 *product-module*s (e.g., fresh eggs, chocolate candy, olive oil, bleach, toilet tissue). Each module is assigned to one of roughly 120 *product-groups* (e.g. candy, shortening and oil, laundry supplies, paper products). These groups belong to one of 10 broader *product-departments* (e.g., dry grocery, fresh produce, non-food grocery). In all of our analyses, we keep all products in modules that are sold in all 48 continental states and we restrict the sample to top-selling modules that rank above the 80th percentile of total U.S. sales. These 198

---

[13]Because the HMS data tracks all household purchases at stores (whether or not the stores are covered by the RMS data), the estimated tax and price elasticities using the HMS data will reflect the full (or "total") quantity demanded response from consumers, inclusive of any endogenous cross-store substitution that occurs in response to changes in taxes or prices. Online Appendix Figure OA.1 shows the spatial distribution of RMS stores and HMS households in our analytical samples across US counties.

[14]Online Appendix Table OA.1 describes the hierarchy of the data using example UPCs. UPCs without a barcode such as random weight meat, fruits, and vegetables are excluded from our sample.

modules account for almost 80 percent of the total sales in grocery stores in the RMS data.[15]

The key variables for our empirical analysis are price and quantity. In RMS, we define these variables at the level of module (m), store (r), and time (n), where a unit of time is a year-quarter. This requires aggregating weekly revenue and quantities sold separately for each product to the quarterly level. A quarterly price is obtained by dividing quarterly revenue from the sales of product j by the number of units sold in that quarter. To address the concern that there may be compositional differences in price across stores due to different UPCs being offered, we follow Handbury and Weinstein (2015) and regress log quarterly price on UPC fixed effects and module-by-store-by-time fixed effects. The module-by-store-by-time fixed effects serve as the pre-tax price for the purpose of estimation. To measure quantity, we create a price-weighted quantity index based on the national price of products.[16] Specifically, for each product (j), store (r), and time (n), we multiply quantity purchased by the average national price (across all stores in our sample) of product j at time n, where the national price is an unweighted average. We then aggregate quantity across products within a module-by-store-by-time cell to arrive at a quantity measure that varies at the same level as the price index.

Price and quantity variables in HMS are defined at the level of module (m), household (i), and time (n). Their construction is analogous to their store-level counterpart, replacing store identifiers with household identifiers.

U.S. Sales Tax Exemptions and Rates   We collect data on local (county and state) sales tax rates and tax exemptions from a variety of sources, including state laws, state regulations, and online brochures.[17] In general, tax exemptions are set by U.S. states and are module-specific. The general rule of thumb is that states exempt food products from taxation and tax non-food products. However, there are several important exceptions to this rule which are reported in Table 1. First, several states

---

[15]We limit to the top 20 percent of modules for computational reasons, and we have explored some of our main specifications in the full sample of modules and found very similar results (results not reported).

[16]This normalization by the national price allows us to compare quantities across different goods and modules, and since all of our specifications include module-by-time fixed effects, the quantity measure is implicitly normalized relative to the module-by-time mean. In other words, explicitly normalizing the quantity measure by the national average price across products within a module-by-time cell leads to identical results.

[17]All data sources used to determine the exemption status of products are listed in Online Appendix Table OA.14.

tax food at the full rate or a reduced rate. Second, in a few states, food products are exempt from the state-level portion of the total sales tax rate, but remain subject to the county-level sales tax.[18] Third, in some cases where food is tax-exempt, there is a tax that applies at the product-module level. For example, prepared foods, soft drinks, and candy are subject to sales taxes in many states. Finally, some states exempt some non-food products from sales taxes. As a result, the effective sales tax rate varies by module (m), county (c), and time (n).[19]

There are two potential sources of measurement error in our sales tax rates. First, we do not incorporate county-level exemptions or county-specific sales surtaxes that apply to specific products or modules, although our understanding is that these cases are uncommon. Second, in some cases, there is some discretion in how we assign a taxability status to each module, based on interpreting the text of a state's sales tax law. While the bulk of the variation in taxes occurs at the module level or higher, there are some instances where taxability varies within module. For example, in New York, fruit drinks are tax exempt as long as they contain at least 70% real fruit juice, but are subject to the sales tax otherwise. Therefore, some products in Nielsen's module "Fruit Juice- Apple", may or may not be taxed in New York, but we code these products as tax exempt since we cannot readily identify the real fruit juice content. In cases where it is impossible to tell whether the majority of products in a given module are subject to the tax or not, we code the statutory tax rate as missing. This results in excluding less than 3 percent of the observations in our sample.

Overall, we are confident that we have measured sales tax rates with a high degree of accuracy. While sales tax exemptions are important for ensuring accurate measurement, the identifying variation

---

[18]Colorado, for example, allows each county to decide whether to subject food to the county-level portion of the sales tax rate.

[19]The Online Appendix shows the cross-sectional variation in sales tax rates and sales tax exemptions in our data. Online Appendix Figure OA.3 reports the total (state + county) sales tax rate in September 2008 and shows tax rates ranging from 0 in Montana, Oregon, New Hampshire, and Delaware to a maximum rate of 9.75 percent in Tennessee. Online Appendix Figure OA.4 reports the food tax exemptions across states and shows that many of the states that tax food are located in either the South or the Midwest. Online Appendix Figure OA.5 shows the changes in sales tax rates between Q1 2006 and Q4 2014, since our main results are based on local variation in sales tax rates over time (rather than cross-sectional variation across states and counties at a point in time). This figure shows that there is still meaningful variation in sales taxes within counties over time during this time period, including both increases and decreases in sales tax rates. Online Appendix Figure OA.6 shows the number of tax changes within counties during the same time period. Lastly, Online Appendix Table OA.3 decomposes the variance in tax rates in the data and reports that a two-standard-deviation change in tax rates within module-by-state-by-year-quarter "cells" is about 0.82 percentage points.

in our empirical analysis comes primarily from changes in sales tax rates within counties over time. Changes in exemptions are very rare during our sample period, and all of our main specifications include module-by-state-by-time fixed effects, so any changes in state sales tax rates or tax exemptions (regardless of the set of modules affected) are absorbed into these fixed effects and thus not used for identification of the effects of sales taxes.

Matched Samples   As a last step in constructing our analysis samples, we merge the tax data onto the RMS and HMS data. The stores in the RMS data are geolocated at the county level so we conduct the merge at the level of module (m), county (c), and time (n). To measure the sales tax rate by quarter of year, we use the tax rate effective at the mid-point of each quarter (February for quarter 1, May for quarter 2, etc). We have also tried using the quarterly average of monthly sales tax rates and found that our estimates were almost identical. Our final RMS sample includes 8,652 grocery stores, and includes price, quantity, and tax rates for 198 modules in 1,460 counties over 36 year-quarters.

Since households in the HMS data may purchase goods in multiple counties, we calculate the average tax rate $\bar{\tau}_{imn}$ paid by household i for their actual expenditures on module m in time n. This requires assigning tax rates to trip-level purchases using information on the location of stores where households shop. For about one-third of purchases, the location of the store is missing in the HMS data. In such cases, we assign the tax rate that prevails in the county of residence of household i. This may introduce measurement error in the average tax rates, which we address using an instrumental variable strategy described below. Our final HMS sample covers purchases for 198 modules over 36 year-quarters for 107,188 households who reside in 2,635 counties.

3.2   Store-level Estimation Strategy

The Effect of Taxes on Prices and Quantity   We use the following model to estimate the effect of sales taxes on consumer prices and quantity:

$$\log y_{mrn} = \beta^y \log(1+\tau_{mcn}) + \delta_{msn} + \delta_{mr} + \varepsilon_{mrn} \tag{2}$$

where the outcome y_{mrn} is either consumer prices $(p(1+\tau))$ or quantity (Q) for module m, store r, and time n. The term τ_{mcn} is the sales tax rate that applies to module m in county c at time n. The terms δ_{msn} and δ_{mr} are module-by-state-by-time and module-by-store fixed effects, respectively. The identifying assumption is that changes in sales taxes within counties over time are uncorrelated with changes in consumer demand (conditional on the fixed effects). This model allows for arbitrary trends across states and modules and therefore relies on within-county-over-time variation in tax rates. The parameter β^y is the elasticity of prices or quantity with respect to taxes ($\beta^{p(1+\tau)}$ and β^Q, respectively). Note that the quantity measure is a market-level measure so that the parameter represents a market-level elasticity of quantity demanded with respect to taxes.

For robustness, we also report results based on a model that uses a subsample of counties and a "county border pair" research design, following Holmes (1998) and Dube, Lester, and Reich (2010). This alternative model identifies the effects of sales taxes on prices and quantity by isolating variation in taxes within pairs of counties on opposite sides of a state border; it is designed to address the concern that sales tax rates are endogenous to local economic conditions. Under the assumption that local economic conditions are similar within a pair of border counties, the effects of sales tax rates on prices and quantity can be identified through "local" comparisons within each county border pair. As a result, for this analysis we restrict the sample to stores located in contiguous counties on opposite sides of a state border. Two contiguous counties located in different states form a county-pair d, and counties are paired with as many cross-state counties as they are contiguous with. The estimating equation is the following:

$$\log y_{mrn} = \beta^y \log(1+\tau_{mcn}) + \delta'_{mdn} + \delta'_{mr} + \varepsilon'_{mrn}. \tag{3}$$

where δ'_{mdn} and δ'_{mr} are module-by-border-pair-by-time and module-by-store fixed effects, respectively. This specification includes flexible trends for each module in each border pair. To estimate equation (3), the original dataset is rearranged by stacking all county pairs and weighting each store by the inverse of the number of times it is included in a border pair. In this regression model, the identifying assumption is that within a border pair, variation in tax rates for a given module over time

is not correlated with other unobserved determinants that differentially affect one of the counties in the pair.

The main results from estimating equations (2) and (3) are reported in Panel A of Table 2. Standard errors are clustered by state-module and reported in parentheses.[20] The first column uses the full sample, and the second column uses the "border pair" subsample. The first row reports results for log consumer prices. In column (1), the coefficient estimate $\widehat{\beta^{p(1+\tau)}} = 0.970$ (s.e. 0.046) indicates a large amount of pass-through of taxes onto consumer prices. The point estimate implies incomplete pass-through ($\beta^{p(1+\tau)} < 1$), but the estimate is not statistically significantly different from 1. However, classical measurement error in the effective tax rate biases our estimate of $\beta^{p(1+\tau)}$ towards 1, because classical measurement error biases the effect of taxes on pre-taxes prices towards 0. We assess the importance of measurement error using an instrumental variable (IV) approach that instruments the county sales tax rate with the state tax rate within the sample of "border pair" counties.[21] Online Appendix Table OA.4 reports these IV estimates of the effect of sales taxes on prices, which are slightly smaller than the corresponding border-pair-sample OLS estimates (0.952 vs 0.980), suggesting a small amount of attenuation bias from measurement error in our main results.[22]

The next row reports the estimate $\widehat{\beta^Q} = -0.775$ (s.e. 0.187), indicating a meaningful quantity response to tax changes. The results in column (2) show similar results using the county border pair sample. The similar results across the columns is consistent with limited endogeneity bias in the full sample.[23]

---

[20]Clustering by state-module is more conservative than clustering by state-by-module-time. We also note that the pass-through estimates are based on a module-level price index which is a generated regressor, but for computational reasons, we ignore the uncertainty in this generated regressor and treat it as measured without error.

[21]Instrumenting the total tax rate with the state tax rate also addresses the concern that the "border pair" counties may adjust their tax rates endogenously based on economic conditions within the border pair.

[22]Additionally, in Online Appendix Table OA.5 we find that the pass-through rate decreases with the degree of market concentration, as measured by a normalized Herfindahl index, where a market is a county-module pair and market shares are calculated separately for each UPC. This is consistent with the theoretical model in Section 2 under some additional parameter restrictions. For example, with constant marginal costs (so that $\epsilon_S = \infty$) and constant elasticity of demand (so that $\epsilon_{ms} = 1/\epsilon_D$), then the pass-through formula in Corollary 1 shows that as market power increases – i.e., as ν/J increases towards 1, which represents monopoly – the pass-through rate decreases.

[23]The minimum wage literature has also tended to find that results using all counties and "traditional" fixed effects models are similar to results based on the county border pair approach (see, e.g., Dube, Lester, and Reich 2010).

Tax salience parameter   To estimate the tax salience parameter θ_τ, the effect of sales taxes on

quantity needs to be scaled by the effect of price changes on quantity.[24] To estimate the price elasticity

of demand, we follow the recent literature on uniform pricing by retail chains and construct a store-

level instrument based on the pricing of products of other stores in a given retail chain (DellaVigna

and Gentzkow 2019). This instrumental variables strategy relates to earlier work by Hausman (1996)

and Nevo (2001), and has been used in several recent papers (e.g., Atkin, Faber and Gonzalez-Navarro

2018 and Allcott et al. 2019).

Specifically, we construct an instrument z_{mrn} that is equal to the average log pre-tax price for a

given model across all stores in the same retail chain excluding store r:

$$z_{mrn} = \frac{\sum_{x \in f} \log(p_{mxn}) - \log(p_{mrn})}{N_{fn} - 1}$$

where f denotes the retail chain to which store r belongs and N_{fn} is the number of stores in chain

f at time n. This is a valid instrument under the assumption that chain-level prices predict "own"

store prices, but are not correlated with unobserved store-level demand determinants. A threat to the

validity of this instrument is the presence of correlated demand shocks within retail chains. To address

this, we continue to include store-by-module fixed effects in all of our specifications. The inclusion of

module fixed effects accounts for the fact that more expensive modules may reflect chains responding

to strong demand for these modules. Intuitively, our identification is coming from differences in

relative prices across modules and chains. To the extent that this variation is driven by differences in

product-specific marginal costs across chains, differences in distribution costs across chains (such as

supply-sourcing costs), or differences in bargaining power across chains, we can consistently estimate

our elasticity of interest using this instrumental variables approach. Intuitively, this approach requires

that chains select store locations based on overall demand factors (that are common across modules),

but not module-specific demand factors. In Online Appendix Table OA.6, we report the reduced-form

---

[24]Formally, in the case of ad valorem tax (τ), the tax salience parameter is defined as follows: $\theta_\tau \equiv \frac{\frac{\partial D}{\partial \tau}}{\frac{\partial D}{\partial p}} \times \frac{1}{p}$. Intuitively, full salience of an ad valorem tax implies that an increase in the tax rate leads to the same demand response as an equivalent proportional change in prices. We begin by estimating a single tax salience parameter that corresponds to the average tax salience in the population (or $\mathbb{E}(\theta_i)$ in the model above), and then we will allow for heterogeneity in the tax salience parameter across households.

relationships between this instrument and price and quantity. To further verify that our results are not contaminated by local module-specific demand shocks, we present corresponding estimates based on an alternative instrument that is equal to the average log pre-tax price across stores in the same chain excluding all stores located in county c, and we show that our estimates are robust to using this alternative instrument. Across all specifications, we find a strong positive relationship between the instrument and price and a strong negative relationship between the instrument and quantity in both the full sample and the county border pair sample.

Using the chain-level instrument, we estimate the price elasticity of demand using the following Two-Stage Least Squares (2SLS) regression model:

$$\log(p(1+\tau)_{mrn}) = \lambda z_{mrn} + \kappa'_{msn} + \kappa'_{mr} + \upsilon'_{mrn} \tag{4}$$

$$\log Q_{mrn} = \alpha \log(p(1+\tau)_{mrn}) + \kappa_{msn} + \kappa_{mr} + \upsilon_{mrn} \tag{5}$$

where the log consumer price, $\log(p(1+\tau)_{mrn})$, is instrumented with z_{mrn}. The κ and κ' terms correspond to the same set of fixed effects as in the regression models in the prior section. Panel B of Table 2 reports the 2SLS estimates of α. The price elasticity of demand estimate in the full sample is $\widehat{\alpha} = -1.150$ (s.e. 0.027), and for the border pair subsample, the estimate is $\widehat{\alpha} = -1.170$ (s.e. 0.027). Both of these values are larger in magnitude than the estimated tax elasticity in Panel A. Given that we estimate pass-through to be very close to one, our finding of a larger price elasticity than tax elasticity suggests that consumers underreact to taxes relative to posted prices.

We next estimate the tax salience parameter θ_τ directly by plugging in each of the estimates in Panel A and Panel B of Table 2 using the following formula:

$$\theta_\tau = \frac{(1-\rho_\tau)\tilde{\epsilon}_D + \tilde{\epsilon}_{D\tau}}{(1+\tau\rho_\tau)\tilde{\epsilon}_D} \tag{6}$$

The derivation of equation (6) is provided in the Online Appendix and comes directly from a result that is analogous to Lemma 1 for the case of ad valorem taxes instead of unit taxes. In this formula, $\rho_\tau = \frac{d\log(p(1+\tau))}{d\log(1+\tau)}$ and corresponds to the estimate $\beta^{p(1+\tau)}$, $\tilde{\epsilon}_D \equiv \frac{d\log(Q)}{d\log(p(1+\tau))}$ and corresponds to the estimate α, and $\tilde{\epsilon}_{D\tau} \equiv \frac{d\log(Q)}{d\log(1+\tau)}$, which corresponds to the estimate β^Q. If there is complete pass-

through ($\rho_\tau = 1$), then the "plug-in" estimate of θ_τ reduces to the ratio of the tax elasticity ($\tilde{\epsilon}_{D\tau}$) to the price elasticity ($\tilde{\epsilon}_D$) when $\tau = 0$. The formula accounts for the fact that when pass-through is incomplete and taxes are not fully salient, manipulating the actual after-tax price is not the same as manipulating the perceived price.

Panel C of Table 3 reports our "plug-in" estimates of θ_τ using our reduced-form results and using $\tau = 0.036$, which is the sample average sales tax rate across all modules.[25] We estimate $\widehat{\theta_\tau} = 0.680$ (s.e. 0.151) using the full sample and $\widehat{\theta_\tau} = 0.556$ (s.e. 0.122) using the border-pair subsample. We also explore whether the tax salience parameter differs between high-price and low-price product modules, and in Online Appendix Table OA.7 we report similar tax elasticity, price elasticity, and implied tax salience parameter estimates across high-price and low-price modules. These results imply that assuming the tax salience parameter is independent of producer (pre-tax) prices is likely to be a reasonable approximation in our setting.[26]

We assess the robustness of our estimates of θ_τ in several ways. First, we consider an alternative method for calculating the tax elasticity and the price elasticity, as well as the associated value of θ_τ, in Online Appendix Table OA.8. In Panels A and B, we report the effect of taxes and of the price instrument on total expenditures.[27] We then back out the implied effect on quantity by subtracting the effect on prices (column 2) from the effect on expenditures (column 3). The implied values of the average tax salience parameter are $\widehat{\theta_\tau} = 0.625$ and $\widehat{\theta_\tau} = 0.501$ for the full sample and the border-pair subsample, respectively. Second, in Online Appendix Table OA.9 we present results based on alternative ways to account for spatial heterogeneity in consumption trends in our main

---

[25]In our baseline calibrations, we use the sample average tax rate covering all modules, including modules that are often tax-exempt, since we include these modules in our empirical analysis, and we have identifying sales tax variation in these modules, as well. For completeness, in the Online Appendix we report additional results using an alternative sample average tax rate that excludes modules which are often tax-exempt. For non-exempt products, state+county sales tax rates are typically in the range of 5 to 10 percent (see, e.g., https://taxfoundation.org/2020-sales-taxes/).

[26]Of course, our results do not rule out that tax salience may vary with prices in other settings, and some previous work (see, e.g., Morrison and Taubisnky 2020) finds some evidence that salience varies with pre-tax prices. We speculate that the similarity in the estimated tax salience parameter across the high-price and low-price sub-samples likely reflects the fact that the modules in our data are all fairly low-priced products.

[27]Total expenditures on module m is equal to $\sum_{j \in m} (q_{jrn} \times p_{jrn})$, where j denotes a UPC. The effect on expenditures, therefore, captures both the effects on prices and on quantity.

sample, and we find broadly similar (though slightly lower) values of θ_τ.[28] Lastly, we re-run our main specification excluding any module that is ever observed to have an excise tax (which excludes all alcohol and tobacco product modules), and we find very similar results excluding these modules (see Online Appendix Table OA.10).

Overall, we conclude that the tax salience parameter estimate is robust to a wide range of alternative samples, specifications, and estimation approaches. However, we identify three potential limitations with the store-level estimates of the average tax salience parameter. First, the average estimate may mask important consumer-level heterogeneity in tax salience. Tax salience could vary with income and education, for example, and heterogeneity in tax salience is one of the key inputs to the welfare formulas developed in Section 2. Second, while the store-level RMS data set tracks a very large number of stores and products, the geographic coverage is incomplete and is not nationally representative. This may lead to a biased estimate of the average tax salience parameter if there is consumer-level heterogeneity in tax salience that varies with the geographic coverage in the store-level data. Third, store-level price and tax elasticities may capture different substitution margins, and if this is the case, then the two elasticity estimates may differ for reasons unrelated to tax salience. For example, consumers may avoid high taxes rates by engaging in cross-border shopping, and they may avoid high store-level prices by substituting between stores within a county; in both cases, this behavior may lead to biased estimates of the average tax salience parameter when relying entirely on store-level data.

As a result, we address these potential limitations by estimating total consumer demand responses to prices and taxes using household-level data to complement the store-level results. Since the coverage of the store-level RMS and household-level HMS are different, we first re-estimate our store-level model weighting each observation by the total value of purchases made by HMS consumers for each

---

[28]Specifically, to account for county-level time-varying heterogeneity, we parameterize country-specific trends for each module as linear time trends (module-by-county-by-year-quarter fixed effects leave no residual variation in tax rates and therefore cannot be used). We also consider store-specific linear trends for each module. The tax and price elasticities under these alternative specifications are smaller than our preferred estimates and imply lower values of θ_τ, ranging between 0.372 and 0.509. The pass-through rate is also lower, around 0.94. We note that the inclusion of module-by-state-by-year-quarter fixed effects in our preferred specification effectively shuts down variation from state-level tax rates, whereas county-module linear trends do not.

module in each store. This effectively shifts the weight towards stores that have larger market shares in HMS in order to better align the two samples. Results are shown in column (3) of Table 2. Consistent with substantial underlying consumer-level heterogeneity, re-weighing stores does alter estimates of average effects. While our estimate of pass-through remains largely unchanged, the effect of taxes on quantity is about 50 percent smaller and less precisely estimated, at $\widehat{\beta^Q} = -0.311$ (s.e. 0.292), and the price elasticity of demand is slightly greater, at $\widehat{\alpha} = -1.258$ (s.e. 0.031). The associated estimate of θ_τ is 0.273.[29] For comparison, Chetty, Looney, and Kroft (2009) estimate $\widehat{\theta_\tau} = 0.35$ using a field experiment which posted tax-inclusive prices in a grocery store. Taubinsky and Rees-Jones (2018) and Morisson and Taubinsky (2020) conduct online shopping experiments in which participants face different tax rates on common household goods. Using experimental variation in tax rates along with a pricing mechanism used to elicit willingness to pay, they report ranges of experimental estimates of θ_τ between 0.23 and 0.54 and between 0.23 and 0.79, respectively. We now turn to household-level estimates of tax salience, which reassuringly show similar average levels of tax salience between the two data sets after adjusting for differences in coverage.

3.3 Household-level Estimation Strategy

The Average Effect of Taxes and Prices on Total Consumer Demand   The household-level estimating equations mimic the store-level design:

$$\log Q_{imn} = \bar{\beta} \log(1+\bar{\tau}_{imn}) + \delta_{msn} + \delta_{im} + \varepsilon_{imn} \tag{7}$$

$$\log Q_{imn} = \bar{\alpha} \log p_{imn} + \kappa_{msn} + \kappa_{im} + \varepsilon_{imn} \tag{8}$$

where $\log Q_{imn}$ is quantity purchased by household i for module m and time (year-quarter) n, and p_{imn} and $\bar{\tau}_{imn}$ are respectively the average price and taxes paid by household i for those purchases. The terms δ_{msn} and δ_{im} are module-by-state-by-time and module-by-household fixed effects, respectively.[30] We cluster standard errors at the household-module level. To address simultaneity bias

---

[29]The average sales tax rate is 0.024 in the weighted sample and 0.036 in the unweighted sample. The column (1) estimates, therefore, put slightly more weight on stores located in high-tax counties.

[30]The module-by-state-by-time fixed effects are based on household i's state of residence at time n.

between price, tax, and unobserved household-level demand shocks, we construct price and tax instruments following the methodology developed in Allcott et al. (2019). First, let $s_{ij(m)r,c}$ be the share of household i's expenditures on module m while residing in county c that are of UPC j in store r. These shares characterize household i's usual shopping behavior for module m while residing in county c. We then calculate predicted taxes Z^{τ}_{imn} and predicted prices Z^{p}_{imn} by fitting tax rates and retail-chain average prices for UPC j in store r at time n to these time-invariant shares:

$$Z^{\tau}_{imn} = \sum_{r \in RMS} \sum_{j \in m} s_{ij(m)r,c} \tau_{jrn}$$

$$Z^{p}_{imn} = \sum_{r \in RMS} \sum_{j \in m} s_{ij(m)r,c} z_{jrn}$$

where τ_{jrn} is the tax rate on product j in store r at time n, and z_{jrn} is the average log pre-tax price for UPC j at time n across all stores in the same retail chain excluding store r.[31] These chain-level averages are calculated in the RMS data and therefore are only observed at RMS stores. To maintain consistency across samples, we construct these instruments using the subset of grocery stores used in our main store-level analysis. Since the location of these stores is known, the tax instrument also helps address any attenuation bias associated with measurement error in average tax rates for all purchases.

2SLS estimates of the average effect of taxes and prices on total consumer demand are presented in Table 3, columns (1) and (3). The tax elasticity of demand is -0.401 (s.e. 0.277) and the price elasticity of demand is -1.365 (s.e. 0.013). These magnitudes are both fairly similar to the corresponding store-level weighted estimates (see Table 2, column (3)).[32] Using a pass-through rate of 0.968 and the sample average tax rate of 0.024 (see Table 2, column (3)), the plug-in estimate of θ_{τ} is 0.318, which represents the average tax salience parameter across consumers.

---

[31] In practice, to standardize the construction of price indices, the price instrument is obtained by regressing log quarterly retail-chain average price on UPC fixed effects and module-by-household-by-time fixed effects, weighting observations by $s_{ij(m)r,c}$. The estimated module-by-household-by-time fixed effects serve as the price instrument.

[32] In Online Appendix Table OA.11, we investigate whether consumers engage in cross-store substitution to avoid higher prices and tax rates. We find that when prices and taxes increase at the RMS stores used in the construction of the instrument, the quantity purchased at these stores decreases more than the total quantity purchased at any store does. This is consistent with cross-store (including cross-border) substitution being an important margin for both the tax and price elasticity of demand.

Heterogeneity in Consumer Inattention As a first step, we use household-level demographic information to investigate whether price and tax elasticities vary along observed dimensions. To this end, we supplement equations (7) and (8) with linear interaction terms, and also include the covariates on their own as regressors. We augment the set of instrumental variables to include linear interactions between the price and tax instruments and the covariates. The variables we include are income, age, and indicators for living in a single-family house, for being married, for the presence of children below the age of 18 in the household, for full-time employment, for college education, and for having internet at home. All covariates are centered around their sample mean. Results are shown in columns (2) and (4) of Table 3. The degree of responsiveness to both taxes and prices decreases with income, and increases with the presence of children. Several interactions are statistically significant at conventional levels, and a joint test of statistical significance of the interaction terms indicates we can reject the null of no heterogeneity.

Demographic variables vary across households and years. We therefore obtain household-year specific elasticities, by producing fitted values using the coefficients on the main effects of taxes and prices as well as the coefficients on the interaction terms. With a slight abuse of notation, we denote these household-year price and tax elasticities by $\tilde{\epsilon}_{D,i}$ and $\tilde{\epsilon}_{D\tau,i}$, although in practice they will vary across years within households. With these in hand, we calculate household-year-specific values of the inattention parameter

$$\theta_{\tau,i} = \frac{(1 - \rho_\tau)\tilde{\epsilon}_{D,i} + \tilde{\epsilon}_{D\tau,i}}{(1 + \tau\rho_\tau)\tilde{\epsilon}_{D,i}} \tag{9}$$

and calculate the variance of $\theta_{\tau,i}$ in the estimation sample.[33] We calibrate pass-through using our RMS estimate of $\rho_\tau = 0.968$ and set $\tau = 0.024$, the average tax rate on household purchases in HMS. This yields an estimate of $Var(\theta_{\tau,i}) = 0.009$. The variance of tax elasticities $Var(\tilde{\epsilon}_{D\tau,i}) = 0.027$ is twice as large as the variance of price elasticities $Var(\tilde{\epsilon}_{D,i}) = 0.013$.[34]

---

[33]Since the reduced-form effects of taxes includes the (endogenous) effect of taxes on consumer prices in both the household-level analysis and the store-level analysis, the inattention parameter is calculated the same way.

[34]Since they are functions of the same observables, the two elasticities are highly correlated, which attenuates the variance of $\theta_{\tau,i}$. To see why, consider the case of $\rho_\tau = 1$ and $\tau = 0$. Then, $\theta_{\tau,i}$ reduces to the ratio of the two elasticities, $\tilde{\epsilon}_{D\tau,i}/\tilde{\epsilon}_{D,i}$. If the two elasticities were perfectly correlated, the ratio would be constant across consumers.

While this approach is useful for identifying observable characteristics that are associated with consumer heterogeneity in the price elasticity and tax elasticity, it likely provides a lower-bound estimate of the true degree of dispersion in consumer inattention since it does not capture unobserved heterogeneity. To overcome this limitation, we turn to a mixed-effects model that allows for random coefficients, and we estimate the model using maximum likelihood. The model permits random household-year intercepts and random household-year coefficients on the tax and price coefficients. We identify the model using parametric functional form assumptions, assuming that the random effects are jointly normally distributed (with unknown variances and covariances).

The estimation proceeds in two steps. First, we re-estimate the effect of taxes and prices on quantity using a control function approach. For taxes, we estimate

$$\log Q_{imn} = \bar{\beta} \log(1 + \bar{\tau}_{imn}) + \underbrace{[\lambda \omega_{imn} + \delta_{msn} + \delta_{im}]}_{\Gamma X_{imn}} + \varepsilon_{imn}$$

where ω_{imn} are the residuals from the first-stage estimating equation (i.e., the residuals from first-stage regression of $\log(1 + \bar{\tau}_{imn})$ on Z^{τ}_{imn}, controlling for δ_{msn} and δ_{im}). We then partial out the covariates to obtain adjusted residuals, $\hat{\varepsilon}_{imn} \equiv \log Q_{imn} - \hat{\Gamma} X_{imn}$.

In the second step, we fit a mixed-effects model on the adjusted residuals allowing for a fixed (constant) effect of $\log(1 + \bar{\tau}_{imn})$, as well as random slopes β_i on $\log(1 + \bar{\tau}_{imn})$ and random intercepts ν_i that vary at the household-year level[35]:

$$\hat{\varepsilon}_{imn} = \bar{\beta} \log(1 + \bar{\tau}_{imn}) + \beta_i \log(1 + \bar{\tau}_{imn}) + \nu_i + \varepsilon'_{imn}$$

We use the mixed-effect model results to obtain empirical Bayes predictions $\hat{\epsilon}_{D\tau,i} = \hat{\bar{\beta}} + \hat{\beta}_i$. We then follow the same approach to estimate the effect of prices on quantity, and we obtain $\hat{\epsilon}_{D,i}$ estimated analogously. With these two estimates, we use equation (9) to calculate $\theta_{\tau,i}$, and from this we can directly estimate the variance of $\theta_{\tau,i}$ in the estimation sample.

Table 4 presents the mixed-effects model results. The variance of tax elasticity random coefficients $Var(\tilde{\epsilon}_{D\tau,i}) = 0.099$ is roughly 3 times greater than that based on observables characteristics. By

---

[35]This two-step approach mimics the methods developed in Chetty, Friedman, and Rockoff (2014) for the estimation of teacher value-added.

contrast, the variance of price elasticities $Var(\tilde{\epsilon}_{D,i}) = 0.002$ is close to zero.[36] Together, these results imply an average value of $\theta_{\tau,i}$ of 0.313, and a variance of $Var(\theta_{\tau,i}) = 0.051$, which is about 5 times larger than the variance estimate based on observable characteristics alone, and translates into a two-standard-deviation range around the mean salience parameter of 0.452.[37]

Taken together, both empirical approaches to measure heterogeneity in consumer inattention strongly reject the null hypothesis of no heterogeneity in the tax salience parameter across households. Additionally, both approaches find much more variation across households in their responsiveness to taxes than prices. This implies that there is meaningful variation in the tax salience parameter across households, and this comes primarily from household-level heterogeneity in the reduced-form tax elasticity.

4   Calibrations

In this section, we calibrate the incidence and marginal excess formulas for ad valorem taxes using the estimates in the previous section. We then compute the difference between our marginal excess burden formula and the standard Harberger formula and decompose it into the deviation coming from salience effects and the deviation coming from imperfect competition. Lastly, we consider counter-factual scenarios that increase tax salience and change the market structure, taking into account the endogeneity of the output response to the tax and the pass-through rate with respect to tax salience and market structure. These scenarios reveal the various interactions between tax salience and imperfect competition in determining the pass-through of taxes into consumer prices, the effect of taxes on quantity demanded, and ultimately the incidence and marginal excess burden.

To begin, we recover the markup and the conduct parameter in several intermediate steps shown in Panel A of Table 5. We assume constant marginal costs and a constant price elasticity of demand

---

[36]It must be noted that we implicitly assume constant effects for the first-stage relationships between average prices and price instruments, and between average taxes and tax instruments. To examine the validity of this assumption, we run similar mixed-effects models on the relevant first-stage residuals. Results are presented in Online Appendix Table OA.12 and confirm that the variances of the random slope coefficients in first-stage relationships are very close to zero.

[37]By comparison, Taubinsky and Rees-Jones (2018) report a lower-bound estimate of the variance of θ_τ around 0.1.

throughout this calibration exercise. Using our estimate of ρ_τ from Table 2, column (3), and our estimate of $\mathbb{E}(\theta_{\tau,i})$ from Table 4, along with the pass-through expression, we recover an estimate of $v_q/(J\epsilon_{ms}) = 0.070$ by exploiting the fact that the elasticity of inverse marginal surplus is equal to the inverse of the price elasticity of demand under constant elasticity of demand; i.e., $\epsilon_{ms} = 1/\epsilon_D$. Next, in order to estimate the markup $(p-mc)/p$, we translate $v_q/(J\epsilon_{ms})$ into $v_q/(J\epsilon_D)$, and since the latter determines the markup, we estimate $(p-mc)/p = 0.037$.[38] Our last intermediate step estimates $v_q/J = 0.051$.

With the estimated markup and conduct parameter in hand, we can calibrate the incidence and marginal excess burden formulas for ad valorem taxes using the estimates of price and tax elasticities in Table 4. The Online Appendix derives the following incidence formula for ad valorem taxes that allows for heterogeneity in θ_τ, which is analogous to the incidence formula in Proposition 1 above[39]:

$$I = \frac{\rho_\tau(1+\tau) + (1 - \mathbb{E}(\theta_{\tau,i}))\tau\tilde{\epsilon}_{D\tau} + \tau(1+\tau)\tilde{\epsilon}_D(1/p)Var(\theta_\tau)}{\left(1 - \frac{v_q}{J}\right)(1 - \rho_\tau) + \frac{v_q}{J}\mathbb{E}(\theta_{\tau,i})(1 + \tau\rho_\tau)} \tag{10}$$

In Panel B of Table 5, column (1) assumes no heterogeneity in θ_τ, while columns (2) and (3) illustrate sensitivity to heterogeneity in θ_τ. Column (2) uses our preferred estimate of $Var(\theta_\tau) = 0.051$ which comes from the mixed-effects model reported in Table 4. Column (3) considers a more conservative special case where $\theta_\tau \in \{0, 1\}$; i.e., some consumers fully optimize, while others are completely inattentive to taxes.[40] In column (1), we calculate $I = 21.088$, which suggests that essentially all of the incidence of sales taxes falls on consumers. In column (2), we allow for heterogeneity in tax salience, and we find this further increases the incidence on consumers to $I = 21.126$.

Turning to the marginal excess burden, we use the following formula that allows for heterogeneity in θ_τ, and we scale the formula so that it represents the change in welfare as a percentage of total

---

[38]Grocery stores operate on relatively low profit margins; industry analyst Jeff Cohen recently said that "It's a very competitive industry ... grocery stores can only slightly mark up the prices for their products." https://www.marketplace.org/2013/09/12/groceries-low-margin-business-still-highly-desirable/.

[39]Throughout this section, we use $Var(\theta_\tau)$ to denote the variance of the inattention parameter for ad valorem taxes.

[40]In this case, since θ_τ is binary, an average θ_τ of 0.313 (i.e., the share of consumers fully optimizing) implies a variance of $0.313 \times (1 - 0.313) = 0.215$. Also note that the incidence formula requires a value for $(1/p)Var(\theta_\tau)$, which normalizes the consumer heterogeneity in θ_τ by price. The values of $\log p$ in our data are centered around zero, implying an average price index p of 1.

revenue:

$$\frac{d\widetilde{W}}{d\tau} \equiv \frac{(1+\tau)}{pQ}\frac{dW}{d\tau} = \left(\frac{p-mc}{p} + \mathbb{E}(\theta_{\tau,i})\tau\right)\tilde{\epsilon}_{D\tau} + \tau(1+\tau)\tilde{\epsilon}_D(1/p)Var(\theta_\tau) \quad (11)$$

Using the sample average tax rate of 2.4 percent for τ and assuming that $Var(\theta_\tau) = 0$, we find $d\widetilde{W}/d\tau = -0.018$ (column 1). This implies that the marginal excess burden is about 1.8 percent of total revenue. Allowing for heterogeneity in θ_τ across consumers increases the welfare cost of taxation to $d\widetilde{W}/d\tau = -0.019$ if $Var(\theta_\tau) = 0.051$, and to $d\widetilde{W}/d\tau = -0.025$ if $Var(\theta_\tau) = 0.215$.

The final panel of Table 5 (Panel C) reports the excess burden calculation if one instead calibrated the standard Harberger formula (which assumes perfect competition and full optimization) using our empirical estimate for the tax elasticity. The standard Harberger formula is given by the following:

$$\frac{d\widetilde{W}^{Harberger}}{d\tau} \equiv \tau\tilde{\epsilon}_{D\tau} \quad (12)$$

The objective of this exercise is to quantify the discrepancy if one "naively" computed the welfare cost of taxation ignoring salience effects and imperfect competition. In this case, we find an estimate of $\frac{d\widetilde{W}^{Harberger}}{d\tau} = -0.009$. To understand the sources of the difference between excess burden according to formula (11) (-0.019) and excess burden according to formula (12) (-0.009), we decompose the difference into the change from accounting for tax salience (while assuming perfect competition) and the change from accounting for imperfect competition (while assuming full tax salience).

Accounting for salience effects but assuming perfect competition leads to the marginal excess burden formulas in Chetty, Looney, and Kroft (2009) and Taubinsky and Rees-Jones (2018), depending on whether or not we allow for heterogeneity in salience effects. These excess burden formulas are given respectively as:

$$\frac{d\widetilde{W}^{CLK}}{d\tau} \equiv \theta_\tau\tau\tilde{\epsilon}_{D\tau} \quad (13)$$

$$\frac{d\widetilde{W}^{TRJ}}{d\tau} \equiv \theta_\tau\tau\tilde{\epsilon}_{D\tau} + \tau(1+\tau)\tilde{\epsilon}_D(1/p)Var(\theta_\tau) \quad (14)$$

Assuming no heterogeneity gives an estimate of $\frac{d\widetilde{W}^{CLK}}{d\tau} = -0.003$ (column 1), which is smaller in magnitude than the excess burden according to the standard Harberger formula assuming full

optimization (-0.009). Allowing for heterogeneity in θ_τ gives an estimate of $\frac{d\widetilde{W}^{TRJ}}{d\tau} = -0.005$ (column 2), which is still smaller than but closer to the Harberger benchmark which ignores tax salience. In the special case where consumers either fully optimize or are fully inattentive (column 3), $\frac{d\widetilde{W}^{TRJ}}{d\tau} = -0.010$, which is very similar to the Harberger benchmark. Thus, allowing for salience effects leads to either smaller or fairly similar estimates of the marginal excess burden calculated using the standard Harberger formula (depending on the amount of heterogeneity in salience effects).

Next, we account for imperfect competition but assume full optimization so that taxes are fully salient. In this case, we calibrate the excess burden according to the formula in Auerbach and Hines (2001):

$$\frac{d\widetilde{W}^{AH}}{d\tau} \equiv \left(\frac{p-mc}{p} + \tau\right)\tilde{\epsilon}_{D\tau} \tag{15}$$

Accounting for imperfect competition but assuming full optimization gives an estimate of $\frac{d\widetilde{W}^{AH}}{d\tau} = -0.024$. This is larger than the marginal excess burden estimate in column (1) and fairly similar to the estimate in column (3).

Overall, we find that tax salience accounts for a smaller percentage of the deviation between excess burden calculated using formula (11) and excess burden calculated using formula (12) in both columns. Numerically, our findings in column (1) indicate that imperfect competition accounts for 180 percent of the difference between excess burden according to our new formula and excess burden according to the Harberger formula, while imperfect salience accounts for -80 percent of the difference. In column (2) the corresponding percentages are 149 percent and -49 percent, respectively. Thus, we conclude that naively applying the standard Harberger formula to calculate the welfare cost of taxation will lead one to understate the marginal excess burden, and our decomposition indicates that this is almost entirely due to accounting for imperfect competition.

We assess the robustness of these results in a number of dimensions. Online Appendix Tables OA.13 and OA.14 report all of the results in Table 5 using the unweighted store-level estimates instead of those based on households (for the full sample and the county border pair sample, respectively). The incidence continues to essentially all fall on consumers, and the excess burden is increased some-

what (e.g. from 1.9 percent in column (2), Table 5, to -3.6 and -2.4 percent in Tables OA.9 and OA.10, respectively).

In Online Appendix Table OA.15, we show sensitivity to alternative values of the elasticity of inverse marginal surplus. For the main results in Table 5, we assumed that this elasticity is equal to the inverse of the price elasticity of demand in order to recover estimates of the conduct parameter and the markup. Alternative functional form assumptions lead to different relationships between these parameters. Since we do not have sufficient data to estimate the elasticity of inverse marginal surplus directly, we instead show sensitivity to different values of this parameter. Varying this parameter by roughly 50 percent in either direction does not change our main qualitative results. Across all the columns, essentially all of the incidence falls on consumers, and the Harberger formula understates the welfare change relative to the general welfare formula that allows for both tax salience and imperfect competition.

In Online Appendix Table OA.16, we reproduce the decomposition using the sample average tax rates on taxed products of $\tau = 0.052$; i.e., excluding tax-exempt product modules from the calculation. This produces very slightly lower estimates of θ_τ and $Var(\theta_\tau)$. More importantly, a higher initial tax rate raises the marginal excess burden, but narrows the gap between our formula and the Harberger benchmark since the latter scales one-for-one with the initial tax rate.

Interestingly, even though we estimate a fairly small markup and a fairly large departure from full salience, we robustly find that imperfect competition "matters more" than salience in terms of accounting for the deviation from the standard Harberger analysis. Mathematically, this comes from the fact that the markup – while fairly small in absolute terms – is nevertheless larger than the product of the average tax rate and the salience "gap", or the gap between full salience and the tax salience parameter; i.e., $(1 - \theta_\tau)$.

4.1 Counterfactual Scenarios

The prior exercise sheds light on the discrepancy between our excess burden estimate and the estimate if one "naively" computed the welfare cost of taxation using the standard Harberger approach which ignores salience effects and imperfect competition. A different question one could ask is what are the welfare consequences of taxation under different market structures and under salient and non-salient taxes. This counterfactual analysis is more appropriate if one wants to compare the excess burden of taxation under different assumptions on the degree of market power and degree of inattention to taxes.

This section reports additional calibration exercises that estimate how the incidence and marginal excess burden would counterfactually be affected by changes to either tax salience or market structure (or both). The basic idea of the exercise is to use our knowledge of the incidence and efficiency formulas under each type of market structure and under different assumptions on tax salience. The key difference between these calibrations and the calibrations in the prior section is that we take into account the endogeneity of the pass-through rate and the output response to taxes with respect to changes in tax salience and market structure. This exercise allows us to capture the fact that when taxes become more salient, holding fixed the price elasticity of demand, consumers are expected to respond more, which leads to an increase in the tax elasticity of demand. Additionally, changes in either tax salience or market structure will change the pass-through of taxes onto consumer prices, and we can use the analytical pass-through expression to re-calculate the counterfactual pass-through of taxes into consumer prices in different scenarios (see the Online Appendix for the ad valorem tax pass-through expression that is analogous to the pass-through formula in Proposition 1 above).

Our main results are summarized in Table 6, which goes through five different counterfactual scenarios that explore the consequences of moving to full salience and either moving to perfect competition or increasing the amount of imperfect competition (relative to the baseline). Panel A of Table 6 shows that the underlying price elasticity of demand elasticity is always held constant across the scenarios. When varying the market structure, the markup and conduct parameter are either counter-factually set to zero (under perfect competition), or the conduct parameter is doubled (from the em-

pirical estimate in Table 5) to decrease the amount of competition. In each scenario, we re-calculate the pass-through and the tax elasticity (shown in the top of Panel B), and we then use these estimates to re-calibrate the incidence and marginal excess burden formulas (bottom of Panel B).

The first column reports the baseline calibration from Table 5 allowing for salience effects and imperfect competition. Column (2) holds constant the estimated average tax salience parameter (θ_τ) but considers the case of perfect competition where the conduct parameter is 0. This leads to full pass-through, which in turn leads to a slightly larger tax elasticity (as consumers now face slightly higher prices for the same change in taxes). The remaining rows in Panel B show full incidence on consumers and a lower marginal excess burden under perfect competition (as compared to column (1)). Column (3) continues to hold constant the tax salience parameter but instead doubles the conduct parameter from 0.051 to 0.102 so that there is less competition in the market. This leads to a lower pass-through, which, in turn, leads to a smaller tax elasticity of demand. In turn, there is a lower incidence of taxation on consumers and a larger marginal excess burden.

Next, column (4) holds constant the markup and conduct parameter and considers the scenario where taxes are fully salient. In this case, the tax elasticity is much larger since consumers respond much more to tax changes as compared to columns (1)-(3). However, the pass-through rate is actually lower than in column (1), reflecting the fact that under imperfect competition, firms will choose lower pass-through when salience is higher.[41] This reveals another interaction between tax salience and imperfect competition: we find that pass-through does not depend on the tax salience parameter under perfect competition given our calibration assumptions, but under imperfect competition, we find that increasing salience parameter leads to lower pass-through, a lower incidence on consumers, and a larger marginal excess burden.

Lastly, the scenarios in columns (5) and (6) continue to assume that taxes are fully salient but vary

---

[41]Our theoretical analysis shows that pass-through can increase with salience even under perfect competition if supply is less than perfectly elastic. Additionally, under imperfect competition pass-through can either increase or decrease with the salience parameter depending on the elasticity of inverse marginal surplus. Given our initial assumptions about supply elasticity and elasticity of inverse marginal surplus and our estimates of the other model parameters, we find that pass-through is generally increasing in the conduct parameter and decreasing with the salience parameter as long as the market is not perfectly competitive.

the degree of competition as in columns (2) and (3). In column (5), the tax elasticity is now the same as the price elasticity of demand since there is full pass-through under perfect competition. Interestingly, this leads to a larger marginal excess burden compared to our baseline calibration in column (1). In other words, "fixing" the misoptimization of consumers and reducing distortions from imperfect competition lead to a larger marginal excess burden than the status quo. While moving to perfect competition reduces the marginal excess burden in each "pair" scenarios (i.e., moving from (1) to (2) and from (4) to (5)), increasing the tax salience parameter increases marginal excess burden (going from (1) to (4), (2) to (5), and (3) to (6)). Additionally, the marginal excess burden from increasing the tax salience parameter increases even more under imperfect competition (as compared to perfect competition). Specifically, the marginal excess burden increases from 3.1 percent of revenue to 10.9 percent across columns (3) and (6) under imperfect competition (going to full salience), compared to an increase of 0.5 percent to 3.3 percent under perfect competition comparing columns (2) and (5). This highlights an additional interaction between tax salience and imperfect competition. Overall, the counterfactual scenarios highlight the interactions between tax salience and imperfect competition, ranging from the pass-through of taxes into prices to the incidence and marginal excess burden of taxation.

5   Conclusion

This paper develops new formulas for the welfare effects of commodity taxation in a model with heterogeneous consumers featuring imperfect competition and tax salience. We find important interactions between salience and the degree of competition for incidence and the efficiency cost of taxation. We also show that heterogeneity in inattention matters for incidence under all market structures, including perfect competition.

We estimate the inputs into the formulas using Nielsen Retail Scanner data, Nielsen Consumer Panel data, and detailed sales tax data. We find nearly-complete pass-through of sales taxes onto prices and meaningful effects of taxes on quantity. We also find that consumers substantially "underreact"

to taxes, with a tax elasticity of about 30 percent of the price elasticity. Lastly, we estimate a large amount of heterogeneity in tax salience; the mean tax salience parameter is about 30 percent, but a two-standard-deviation range around that mean is 45 percentage points.

We use our formulas to estimate other model parameters. For example, we estimate a markup around 3.7 percent, which is consistent with grocery stores operating with fairly low profit margins. We then use these model-based estimates to calibrate our new incidence and efficiency formulas. We find essentially all of the incidence falls on consumers, regardless of whether or not we account for imperfect competition and tax salience. Turning to welfare, we find that the standard marginal excess burden formula substantially understates the welfare costs of commodity taxation, even after accounting for consumers' underreaction due to salience effects. While we estimate substantial underreaction to taxes alongside a fairly small markup (and thus a fairly small departure from perfect competition), our calibration results suggest that both are important for welfare analysis. Ignoring imperfect competition but allowing for salience effects leads to a substantial underestimate of the marginal excess burden, while ignoring salience effects but allowing for imperfect competition leads to an overestimate of the marginal excess burden, although by a somewhat smaller magnitude.

It is perhaps a surprising result that imperfect competition "matters more" than salience effects in our welfare analysis (relative to the benchmark case of perfect competition with full optimization), since we find a fairly small departure from perfect competition alongside a large departure from fully-optimizing behavior. Our new welfare formula shows why this is the case: the marginal excess burden scales one-for-one with the estimated markup while the tax salience parameter scales with the tax rate. Nevertheless, we conclude that both imperfect competition and tax salience are important factors to consider together when analyzing the incidence and efficiency consequences of commodity taxation. Focusing on either one in isolation will generally lead to misleading estimates, and accounting for one (but not the other) will – in some circumstances – lead to conclusions that are even less accurate than a standard Harberger analysis that ignores both imperfect competition and salience effects.

Our general finding that tax salience and imperfect competition interact when considering the

welfare cost of taxation may have broader lessons for policy. Since imperfect competition is one type of a pre-existing distortion in the market, we conjecture that our formulas may be useful for analyzing how tax salience affects the welfare cost of taxation under other pre-existing distortions such as externalities. Our counterfactual scenarios showed that increases in tax salience lead to larger welfare effects of taxes under imperfect competition compared to perfect competition. This implies that policies such as carbon taxes and sugar taxes may be more effective "corrective taxes" if they are incorporated into posted prices to increase salience since tax salience presumably scales linearly with the marginal damage from the externality in the welfare cost formula.

References

1. Adachi, Takanori and Michal Fabinger (2019). "Multi-Dimensional Pass-Through and Welfare Measures under Imperfect Competition," Working paper.

2. Allcott, Hunt and Dmitry Taubinsky (2015). "Evaluating Behaviorally Motivated Policy: Experimental Evidence from the Lightbulb Market," *American Economic Review*, 105(8): 2501-2538.

3. Allcott, Hunt, Benjamin B. Lockwood and Dmitry Taubinsky (2019). "Regressive sin taxes, with an application to the optimal soda tax," *The Quarterly Journal of Economics*, 134(3), 1557-1626.

4. Allcott, Hunt, Benjamin B. Lockwood and Dmitry Taubinsky (2018). "Ramsey Strikes Back: Optimal Commodity Taxes and Redistribution in the Presence of Salience Effects," *American Economic Review, Papers and Proceedings*, 108(5): 88-92.

5. Allcott, Hunt, Rebecca Diamond, Jean-Pierre Dubé, Jessie Handbury, Ilya Rahkovsky, Molly Schnell (2019). "Food Deserts and the Causes of Nutritional Inequality," *The Quarterly Journal of Economics*, 134(4): 1793–1844, https://doi.org/10.1093/qje/qjz015.

6. Anderson, Simon P., Andre De Palma and Brent Kreider (2001a). "The Efficiency of Indirect Taxes under Imperfect Competition," *Journal of Public Economics*, 81(2): 231-251.

7. Anderson, Simon P., Andre De Palma and Brent Kreider (2001b). "Tax Incidence in Differentiated Product Oligopoly," *Journal of Public Economics*, 81(2): 173-192.

8. Atkin, David, Benjamin Faber and Marco Gonzalez-Navarro (2018). "Retail Globalization and Household Welfare: Evidence from Mexico," *Journal of Political Economy*, 126(1): 1-73.

9. Auerbach, Alan J. and James R. Hines Jr. (2001). "Perfect Taxation with Imperfect Competition," NBER Working Paper #8138.

10. Auerbach, Alan J. and James R. Hines Jr. (2002). "Taxation and Economic Efficiency," *Handbook of Public Economics*, Volume 3 (Chapter 21).

11. Baker, Scott R., Stephanie Johnson, and Lorenz Kueng (2018). "Shopping for Lower Sales Tax Rates," Working Paper.

12. Bernheim, Douglas B. and Dmitry Taubinsky (2018). "Behavioral Public Economics," In B. Douglas Bernheim, Stefano DellaVigna, and David Laibson (eds.), Handbook of Behavioral Economics, Volume 1, New York: Elsevier: 381-516.

13. Besley, Timothy J. and Harvey S. Rosen (1999). "Sales taxes and prices: an empirical analysis,"*National Tax Journal*, 52(2): 157-158.

14. Bradley, Sebastien and Naomi E. Feldman (2019). "Hidden Baggage: Behavioral Responses to Changes in Airline Ticket Tax Disclosure," forthcoming *American Economic Journal: Economic Policy*.

15. Bulow, Jeremy and Paul Pfleiderer (1983). "A Note on the Effect of Cost Changes on Prices," *Journal of Political Economy*, 91(1): 182-185.

16. Chetty, Raj, Adam Looney and Kory Kroft (2007). "Salience and Taxation: Theory and Evidence," *NBER Working Paper 13330.*

17. Chetty, Raj, Adam Looney and Kory Kroft (2009). "Salience and Taxation: Theory and Evidence," *American Economic Review*, 99 (4): 1145-77.

18. Chetty, Raj, John N. Friedman, and Jonah Rockoff (2014). "Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates," *American Economic Review*, 104(9), 2593-2632.

19. DellaVigna, Stefano and Matthew Gentzkow (2019). "Uniform Pricing in US Retail Chains," *The Quarterly Journal of Economics*, 134(4): 2011-2084.

20. Delipalla, Sophia and Michael Keen (1992). "The Comparison between Ad Valorem and Specific Taxation under Imperfect Competition," *Journal of Public Economics*, 49(3): 351-361.

21. Dube, Arindrajit, T. William Lester, and Michael Reich (2010). "Minimum Wage Effects Across State Borders: Estimates Using Contiguous Counties," *The Review of Economics and Statistics*, 92(4): 945–964.

22. Einav, Liran, Jonathan D. Levin, Neel Sundaresan (2014). "Sales Taxes and Internet Commerce," *American Economic Review*, 104(1): 1-26.

23. Farhi, Emmanuel and Xavier Gabaix (2020). "Optimal Taxation with Behavioral Agents," *American Economic Review*, 110(1): pp. 298-336.

24. Gabaix, Xavier (2019). "Behavioral Inattention," in in Douglas Bernheim, Stefano DellaVigna, and David Laibson, eds., *Handbook of Behavioral Economics*, 2: pp. 261-343.

25. Goldin, Jacob and Tatiana Homonoff (2013). "Smoke Gets in Your Eyes: Cigarette Tax Salience and Regressivity," *American Economic Journal: Economic Policy*, 5(1): 302-36.

26. Gruber, Jonathan, and Botond Kőszegi (2004). "Tax incidence when individuals are time-inconsistent: the case of cigarette excise taxes." *Journal of Public Economics*, 88(9-10): 1959-1987.

27. Hackner, Jonas and Mathias Herzing (2016). "Welfare Effects of Taxation in Oligopolistic Markets," *Journal of Economic Theory*, 163(C): 141-166.

28. Handbury, Jessie, and David E. Weinstein (2015), "Goods Prices and Availability in Cities," *Review of Economic Studies*, 82(1): 258-296.

29. Harberger, Arthur (1964). "The Measurement of Waste," *American Economic Review* 54(3): 58-76.

30. Hausman Jerry A. (1996). "Valuation of new goods under perfect and imperfect competition," *The Economics of New Goods*, eds. TF Bresnahan, RJ Gordon, chap. 5. Chicago: University of Chicago Press, 209–248.

31. Holmes, Thomas J. (1998). "The Effects of State Policies on the Location of Industry: Evidence from State Borders," *Journal of Political Economy*, 106(4): 667-705.

32. Kroft, Kory, Jean-William P. Laliberté, René Leal-Vizcaíno, and Matthew J. Notowidigdo (2021). "Efficiency and Incidence of Taxation with Free Entry and Love-of-Variety Preferences," *NBER Working Paper 28838.*

33. Liebman, Jeffrey B., and Richard J. Zeckhauser (2004). "Schmeduling," `https://scholar.harvard.edu/files/jeffreyliebman/files/Schmeduling_WorkingPaper.pdf`.

34. Miravete, Eugenio J., Katja Seim and Jeff Thurk (2018). "Market Power and the Laffer Curve," *Econometrica*, 86(5): 1651-1687.

35. Morrison, William and Dmitry Taubinsky (2020). "Rules of Thumb and Attention Elasticities," Working Paper.

36. Taubinsky, Dmitry and Alex Rees-Jones (2018). "Attention Variation and Welfare: Theory and Evidence from a Tax Salience Experiment," *Review of Economic Studies*, 85: 2462-2496.

37. Seade, Jesus (1987). "Profitable Cost Increases and the Shifting of Taxation: Equilibrium Responses of Markets in Oligopoly," Working Paper.

38. Stern, Nicholas (1987). "The Effects of Taxation, Price Control, and Government Contracts in Oligopoly and Monopolistic Competition," *Journal of Public Economics*, 32: 133-158.

39. Veiga, André and Glen Weyl (2016). "Product Design in Selection Markets," *Quarterly Journal of Economics*, 131(2): 1007-1056.

40. Weyl, Glen and Michal Fabinger (2013). "Pass-Through as an Economic Tool: Principle of Incidence under Imperfect Competition," *Journal of Political Economy*, 121(3): 528-583.

Table 1: Sales Tax Exemptions for Food and Non-Food Products Across States

Panel A: Food Modules

Module	Average Store-Level Expenditure	States taxing all food	States taxing module at reduced rate	States taxing module at full rate (but otherwise exempting food)
DAIRY - MILK	3.04%	AL, ID, KS, MS, OK, SD	AR, IL, MO, NC, TN,UT,VA,WV	
SOFT DRINKS - CARBONATED	2.88%	AL, ID, KS, MS, OK, SD	AR, IL, MO, TN,UT,VA	CA, CT, FL, IA, IN, KY, MD, ME, MN, NC, ND, NJ, NY, OH, PA, RI, TX, WA, WI, WV
BAKERY - BREAD - FRESH	2.19%	AL, ID, KS, MS, OK, SD	IL, MO, TN, UT, VA, WV	
CEREAL - READY TO EAT	1.93%	AL, ID, KS, MS, OK, SD	AR, IL, MO, NC, TN,UT,VA,WV	
SOFT DRINKS - LOW CALORIE	1.62%	AL, ID, KS, MS, OK, SD	AR, IL, MO, TN,UT,VA	CA, CT, FL, IA, IN, KY, MD, ME, MN, NC, ND, NJ, NY, OH, PA, RI, TX, WA, WI, WV
WATER-BOTTLED	1.42%	AL, ID, KS, MS, OK, SD	AR, IL, MO, NC, TN,UT,VA,WV	LA, MD, ME, MN, NY
ICE CREAM - BULK	1.22%	AL, ID, KS, MS, OK, SD	AR, IL, MO, NC, TN,UT,VA,WV	FL, MD
COOKIES	1.21%	AL, ID, KS, MS, OK, SD	AR, IL, MO, NC, TN,UT,VA,WV	
CANDY-CHOCOLATE	0.64%	AL, ID, KS, MS, OK, SD	AR, IL, MO, UT,VA,WV	CT, FL, IA, IN, KY, MD, ME, MN, NC, ND, NJ, NY, RI, TN,

Panel B: Non-Food Modules

Module	Average Store-Level	States with no sales tax	States exempting module	States taxing module at reduced rate
WINE - DOMESTIC	2.11%	DE, MT, NH, OR	PA, KS, KY, MA	
CIGARETTES	1.70%	DE, MT, NH, OR	CO, MN, OK	
TOILET TISSUE	1.07%	DE, MT, NH, OR	PA, NJ	
DETERGENTS - LIQUID	0.75%	DE, MT, NH, OR		
PAPER TOWELS	0.66%	DE, MT, NH, OR	NJ	
RUM	0.54%	DE, MT, NH, OR	PA, KS, KY, MA	
DISPOSABLE DIAPERS	0.50%	DE, MT, NH, OR	MA, MN, NJ, PA, VT	IL
MAGAZINES	0.41%	DE, MT, NH, OR	MA, ME, NY, OK	
CAT FOOD - DRY TYPE	0.35%	DE, MT, NH, OR		
COLD REMEDIES - ADULT	0.28%	DE, MT, NH, OR	CT, FL, MD, MN, NJ, NY, PA, TX, VA, VT	IL
DOG & CAT TREATS	0.25%	DE, MT, NH, OR		
ALE	0.25%	DE, MT, NH, OR	PA, KS, KY, MA	
DOG FOOD - WET TYPE	0.23%	DE, MT, NH, OR		
FACIAL TISSUE	0.22%	DE, MT, NH, OR	NJ	
TOOTH CLEANERS	0.22%	DE, MT, NH, OR	PA	IL

Notes: Tax exemption status as in September 2008 for selected list of modules. The list only includes modules in our analysis sample.

Table 2

Store-level Estimates of Pass-Through, Tax Elasticity, Price Elasticity of Demand, and Tax Salience

Sample:	Full Sample	County Border Pair Subsample	Full Sample
Weights:	Unweighted	Inverse of number of border-pairs	HMS-based weights
	(1)	(2)	(3)
Panel A: Reduced-form OLS Estimates of the Effects of Sales Taxes on Consumer Prices and Quantity			
$d \log(p(1+\tau))/d \log(1+\tau)$ [pass-through]	0.970	0.980	0.968
	(0.046)	(0.016)	(0.054)
$d \log(Q)/d \log(1+\tau)$ [tax elasticity]	-0.775	-0.649	-0.311
	(0.187)	(0.084)	(0.292)
Panel B: 2SLS Estimates of the Price Elasticity of Demand			
$d \log(Q)/d \log(p)$	-1.150	-1.170	-1.258
	(0.027)	(0.027)	(0.031)
Panel C: "Plug-in" Estimate of Tax Salience Parameter			
θ_τ	0.680	0.556	0.273
	(0.151)	(0.122)	(0.210)
Average tax rate, τ	0.036	0.034	0.024
Specification:			
Store $\times$ Module fixed effects	y	y	y
Module $\times$ State $\times$ Year-Quarter fixed effects	y		y
Module $\times$ Border Pair $\times$ Year-Quarter fixed effects		y	
N	53,987,430	33,749,257	48,051,884

Notes: This table reports estimates of the effects of sales taxes, of the price elasticity of demand, and of the tax salience parameter. In Panel A, the estimates come from OLS estimates of equations (2) and (3), and the independent variable is quarterly sales tax rate of module m in county c in state s. One observation is a module in a store in a given quarter. Consumer prices $p(1+\tau)$ are tax inclusive. The sample is restricted to modules above the 80th percentile of the national distribution of sales. In Panel B, the reported coefficients are 2SLS estimates of the effect of consumer prices on quantity sold, estimated using equations (4) and (5), where prices are instrumented with leave-self-out chain-level average prices. In Panel C, we report the estimate of the tax salience parameter using equation (6). For this parameter, standard errors are based on 100 bootstrap replications. All standard errors in this table are clustered at the state-module level and are reported in parentheses. In column (1), the sample includes our full sample of stores and the regression model includes module-by-store and module-by-quarter-by-state fixed effects. In column (2), the sample is restricted to stores in border counties and the regression model includes module-by-store and module-by-border-pair-by-year-quarter fixed effects, where border pairs denote pairs of contiguous counties on opposite sides of a state border. In column (2), observations are weighted by the inverse of the number of times a store appears in the data. In column (3), observations are weighted, where each module-store cell is weighted by the total expenditures by HMS consumers over the study period. Some cells have no observed purchase in HMS, and are therefore assigned a weight of zero.

Table 3

Household-level Estimates of Tax Elasticity, Price Elasticity of Demand, and Tax Salience

Independent variable (x):	Taxes, $(1+\tau)$		Prices, $p(1+\tau)$	
	(1)	(2)	(3)	(4)
Panel A: 2SLS Estimates of the Tax and Price Elasticities				
$\log(x)$	-0.401	-0.386	-1.365	-1.371
	(0.277)	(0.277)	(0.013)	(0.013)
$\log(x)$ × Income		0.094		0.065
		(0.025)		(0.007)
$\log(x)$ × Age		0.002		0.002
		(0.004)		(0.000)
$\log(x)$ × Single Family House		-0.025		-0.043
		(0.051)		(0.012)
$\log(x)$ × Married		-0.140		-0.083
		(0.057)		(0.011)
$\log(x)$ × Children		-0.217		-0.125
		(0.041)		(0.012)
$\log(x)$ × Full-time employment		0.033		0.056
		(0.037)		(0.010)
$\log(x)$ × College educated		0.032		0.039
		(0.052)		(0.010)
$\log(x)$ × Internet		0.073		-0.019
		(0.045)		(0.013)
Joint significance test of interactions, p-value		[0.000]		[0.000]

Panel B: "Plug-in" Estimate of Tax Salience Parameter	
θ_τ (no consumer heterogeneity, columns (1) and (4))	0.318
$E[\theta_\tau]$ (heterogeneous consumers, columns (2) and (4))	0.292
$\mathrm{Var}(\theta_\tau)$	0.009
Average tax rate, τ	0.024
Other variance terms:	
$\mathrm{Var}(\tilde{\epsilon}_{D\tau})$	0.027
$\mathrm{Var}(\tilde{\epsilon}_D)$	0.013
$\mathrm{Cov}(\theta_\tau, \tilde{\epsilon}_D)$	-0.010

Specification:				
Household × Module fixed effects	y	y	y	y
Module × State × Year-Quarter fixed effects	y	y	y	y
N	51,346,211		51,346,211	

Notes: This table reports estimates of the effects of sales taxes, of the price elasticity of demand, and of the tax salience parameter. In columns (1) and (2), the independent variable is the average quarterly sales tax rate faced by household i on module m. In columns (3) and (4), the independent variable is the average quarterly price paid by household i for purchases on module m. All reported coefficients in panel A are 2SLS estimates using the tax and price instruments described in the main text. In columns (2) and (4), the set of instruments include all linear interactions between tax (price) and the listed observable characteristics. In Panel B, we report estimates of the variance of the tax salience parameter. All standard errors in this table are clustered at the household-module level and are reported in parentheses. Observations are weighted using Nielsen's projection factors in order to obtain national representativeness.

Table 4

Mixed-Effects Estimates of Tax Elasticity, Price Elasticity of Demand, and Tax Salience

Panel A: Tax Elasticity Random Coefficients	
Average tax elasticity, $E[\tilde{\epsilon}_{D\tau}]$	-0.396
Variance of tax elasticity, $Var(\tilde{\epsilon}_{D\tau})$	0.099
Panel B: Price Elasticity Random Coefficients	
Average price elasticity, $E[\tilde{\epsilon}_D]$	-1.375
Variance of price elasticity, $Var(\tilde{\epsilon}_D)$	0.002
Panel C: Heterogeneity in Tax Salience	
Average tax salience, $E[\theta_\tau]$	0.313
$Var(\theta_\tau)$	0.051
$Cov(\theta_\tau, \tilde{\epsilon}_D)$	0.001
N	51,346,211

Notes: This table reports estimates of the variance of tax and price elasticities of demand based on mixed-effects models. Panel A and Panel B report the average elasticities as well as the sample variance of the associated empirical Bayes predictions of the random coefficients for the tax elasticity and price elasticity, respectively. Panel C reports estimates of the variance of the tax salience parameter. The mixed-effects models allow for random coefficients on the tax and price variables across household-year cells. Observations are weighted using Nielsen's projection factors in order to obtain a nationally representative sample.

Table 5
Calibration of Incidence and Marginal Excess Burden Formulas

	(1)	(2)	(3)
Panel A: Inputs and Intermediate Estimates Needed in Calibration			
<u>Inputs:</u>			
Average tax rate, τ		0.024	
Price elasticity, $\tilde{\epsilon}_D \equiv \partial\log(Q)/\partial\log(p(1+\tau))$		-1.375	
Tax pass-through, $\rho_\tau \equiv d\log(p(1+\tau))/d\log(1+\tau)$		0.968	
Tax elasticity, $\tilde{\epsilon}_{D\tau} \equiv d\log(Q)/d\log(1+\tau)$		-0.396	
<u>Intermediate estimates:</u>			
Implied estimate of $v_q/(J\epsilon_{ms})$		0.070	
Implied markup $(p-mc)/p$		0.037	
Implied estimate of v_q/J		0.051	
($v_q/J=0$ is perfect competition, $v_q/J=1$ is monopoly)			
<u>Tax salience:</u>			
Tax salience parameter, θ_τ		0.313	
Heterogeneity in θ_τ, $(1/p)\text{Var}(\theta_\tau)$	0.000	0.051	0.215
Panel B: Incidence and Marginal Excess Burden Formulas			
<u>Incidence (I)</u>			
$I \equiv (dCS/d\tau)/(dPS/d\tau)$	21.088	21.126	21.247
$= (\rho_\tau(1+\tau)+(1-\theta_\tau)\tau\tilde{\epsilon}_{D\tau}-\tau(1+\tau)\tilde{\epsilon}_D(1/p)\text{Var}(\theta_\tau))\ /$			
$\quad ((1-v/J)(1-\rho_\tau)+(v/J)\theta_\tau(1+\tau\rho_\tau))$			
<u>Marginal Excess Burden ($d\tilde{W}/d\tau$)</u>			
$d\tilde{W}/d\tau = ((p-mc)/p + \theta_\tau\tau)\tilde{\epsilon}_{D\tau} - \tau(1+\tau)\tilde{\epsilon}_D(1/p)\text{Var}(\theta_\tau)$	-0.018	-0.019	-0.025
Panel C: Decomposition of the Deviation Between General Formula and Harberger Formula			
Harberger formula (assuming perfect competition and full salience), $d\tilde{W}/d\tau = \tilde{\epsilon}_{D\tau}$	-0.009	-0.009	-0.009
Imperfect salience only, $d\tilde{W}/d\tau = \theta_\tau\tau\tilde{\epsilon}_{D\tau} - \tau(1+\tau)\tilde{\epsilon}_D(1/p)\text{Var}(\theta_\tau)$	-0.003	-0.005	-0.010
Decomposition as % of difference b/w Harberger and general formula	-80%	-49%	5%
Imperfect competition only, $d\tilde{W}/d\tau = ((p-mc)/p + \tau)\tilde{\epsilon}_{D\tau}$	-0.024	-0.024	-0.024
Decomposition as % of difference b/w Harberger and general formula	180%	149%	95%

<u>Notes:</u> This table reports calibrations of the tax incidence and marginal excess burden formulas. The results of these calibrations are shown in Panel B. Panel A presents the value of the input parameters taken from Tables 2 through 4, as well as estimates of intermediate parameters (see main text for details). Panel C presents a decomposition of the deviation between the general formula calibrated in Panel B. In column (1), we assume no heterogeneity in salience across consumers; in column (2) we allow for heterogeneity in the salience parameter by calibrating the variance of θ_τ using the estimate reported in Table 4. In column (3), we consider the special case of consumers being either fully attentive or fully inattentive to taxes.

Table 6
Counterfactual Scenarios Adjusting Tax Salience and Market Structure

	Imperfect salience, $\theta_\tau = 0.313$	Counterfactual scenarios				
Tax salience parameter, θ_τ:		Imperfect salience, $\theta_\tau = 0.313$		Full salience, $\theta_\tau = 1.00$		
	$\text{Var}(\theta_\tau) = 0.051$	$\text{Var}(\theta_\tau) = 0.051$		$\text{Var}(\theta_\tau) = 0$		
Market structure (conduct parameter, v_q/J):	Imperfect competition, $v_q/J = 0.051$	Perfect competition, $v_q/J = 0$	Imperfect competition, $v_q/J = 0.102$	Imperfect competition, $v_q/J = 0.051$	Perfect competition, $v_q/J = 0$	Imperfect competition, $v_q/J = 0.102$
	(1)	(2)	(3)	(4)	(5)	(6)

Panel A: Inputs and Intermediate Estimates Needed to Calibrate Formulas

Inputs: [constant parameters]						
Average tax rate, τ	0.024	0.024		0.024		
Price Elasticity, $\tilde\epsilon_D \equiv \partial\log(Q)/\partial\log(p)$	-1.375	-1.375		-1.375		
Intermediate estimates:						
Implied estimate of $v_q/(J\epsilon_{ms})$	0.070	0.000	0.140	0.070	0.000	0.140
Implied markup $(p-mc)/p$	0.037	0.000	0.074	0.037	0.000	0.074

Panel B: Counterfactual Tax Responses and Implied Incidence and Marginal Excess Burden

Counterfactual reponses:	Baseline estimates	Counterfactual responses				
Tax Pass-Through, $\rho_\tau \equiv d\log(p(1+\tau))/d\log(1+\tau)$	0.968	1.000	0.940	0.898	1.000	0.809
Tax Elasticity, $\tilde\epsilon_{D\tau} \equiv d\log(Q)/d\log(1+\tau)$	-0.396	-0.437	-0.352	-1.232	-1.375	-1.106
$= \tilde\epsilon_D(\rho_\tau - 1 + \theta_\tau/(1+\tau\theta_\tau))(1+\tau)$						
Incidence (I)						
$I \equiv (dCS/d\tau)/(dPS/d\tau)$	21.126	∞	11.076	6.184	∞	3.010
$= (\rho_\tau(1+\tau)+(1-\theta_\tau)\tau\tilde\epsilon_{D\tau}-\tau(1+\tau)\tilde\epsilon_D(1/p)\text{Var}(\theta_\tau))$ /						
$((1-v/J)(1-\rho_\tau)+(v/J)\theta_\tau(1+\tau\rho_\tau))$						
Marginal Excess Burden ($d\tilde W/d\tau$)						
$d\tilde W/d\tau = ((p-mc)/p + \theta_\tau\tau)\tilde\epsilon_{D\tau} - \tau(1+\tau)\tilde\epsilon_D(1/p)\text{Var}(\theta_\tau)$	-0.019	-0.005	-0.031	-0.075	-0.033	-0.109

Notes: This table reports calibrations of the tax incidence and marginal excess burden formulas for different counterfactual assumptions about tax salience and market structure. Columns (1)-(3) allow for heterogeneity in the salience parameter following column (2) of Table 5. Panel A shows the parameters that are held constant across scenarios as well as the implied markup (given assumed market structure that determines the degree of competition). Panel B shows the counterfactual responses to tax changes for each scenario, which can be compared to the empirical estimates in column 1 (based on Table 4). Lastly, Panel B reports the incidence and marginal excess burden. Column (1) reports results using the baseline estimates of salience and conduct parameter from Table 5, and the remaining columns report counterfactual results assuming different values of the tax salience parameter and the conduct parameter. Columns (1)-(3) use the baseline estimate of the tax salience parameter, and columns (4)-(6) assume taxes are fully salient. Columns (2) and (5) assume perfect competition, while columns (3) and (6) assume a conduct parameter that is double the baseline estimate from Table 5.