Salience and Taxation with Imperfect Competition

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

December 2020

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, and we find that tax salience and market structure interact when considering tax incidence and the marginal excess burden. We estimate the necessary inputs to the formulas by combining Nielsen Retail Scanner data from grocery stores in the US with detailed sales tax data. We calibrate our new formulas 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.

*Kroft: University of Toronto and NBER, kory.kroft@utoronto.ca; Laliberté: University of Calgary, jean-william.lalibert@ucalgary.ca; Leal-Vizcaíno: Bank of Mexico, rlealv@banxico.org.mx; Notowidigdo: University of Chicago Booth School of Business and NBER; noto@chicagobooth.edu. We thank Simon Anderson, Raj Chetty, Julie Cullen, Amy Finkelstein, Xavier Gabaix, Nathan Hendren, Louis Kaplow, Henrik Kleven, Nicholas Li, Jesse Shapiro, Rob Porter, Aviv Nevo, Stephen Coate, and numerous seminar participants for helpful comments. We thank Eileen Driscoll, Robert French, Adam Miettinen, Boriana Miloucheva, Pinchuan Ong, Shahar Rotberg, Marc-Antoine Schmidt, Stephen Tino, Jessica Wagner, Ting Wang, Haiyue Yu, and Ruizhi Zhu for extremely valuable research assistance. We gratefully acknowledge funding from the Social Sciences and Humanities Research Council (SSHRC). Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors(s) and do not necessarily reflect the views of the SSHRC. This research is based in part on data from The Nielsen Company (US), LLC and marketing databases provided through the Nielsen Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the Nielsen data are those of the researchers and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

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.[1] 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

[1] While Taubinsky and Rees-Jones (2018) study how consumer heterogeneity affects the efficiency cost of taxation under perfect competition, they do not consider incidence.

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 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, and we generalize these results by showing that this misallocation does not directly interact with market structure. Specifically, we find that greater dispersion in inattention increases the welfare cost of taxes in similar ways under perfect and imperfect competition.

After presenting our theoretical results, we provide new estimates of the necessary inputs to our tax formulas using Nielsen Retail Scanner data covering grocery stores selling consumer goods in the U.S. combined with county-level and state-level sales tax 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 which 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.

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 that ignores salience effects and assumes perfect competition. Our results show that the standard Harberger formula (Harberger 1964) *understates* the marginal excess burden of taxes, which contrasts with Chetty, Looney and Kroft (2009) who show that when consumers underreact to taxes, the standard Harberger formula *overstates* the true marginal excess burden of taxes. Intuitively, this is because there is a pre-existing distortion coming from firms' market power under imperfect competition. While the markup scales one-for-one in the excess burden formula, the mean and variance of the tax salience parameter scale with the tax rate, just as in the case of perfect competition. 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. This counterfactual analysis is more appropriate if one wants to compare the welfare consequences of taxation under different assumptions on the degree of market power and degree of inattention to taxes. Our analysis of these scenarios reveal 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.

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; Adachi and Fabinger 2019; Miravete, Seim, and Thurk 2018). 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 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; Bordalo, Gennaioli, and Shleifer 2013; Goldin and Hominoff 2013; Kőszegi and Szeidl 2013; Allcott and Taubinsky 2015; Caplin and Dean 2015; Taubinsky and Rees-Jones 2018; Allcott, Lockwood, and Taubinsky 2018; Bradley and Feldman 2019; Farhi and Gabaix 202; Morrison and Taubinsky 2020). 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 perfect

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

competition and considers the welfare effects of a unit tax. Section 3 extends the results to monopoly and the general model of imperfect competition. Section 4 derives analogous formulas for the case of an ad valorem tax and compares the incidence and efficiency costs of ad valorem and unit taxes. Section 5 discusses the data and the empirical results. Section 6 presents the calibration results. Section 7 concludes.

2 Perfect Competition

We are interested in characterizing the incidence and marginal excess burden effects of commodity taxation allowing for salience effects. 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 of the tax as $\frac{dW}{dt} = \frac{dCS}{dt} + \frac{dPS}{dt} + \frac{dR}{dt}$ where CS denotes consumer surplus, PS denotes producer surplus, R denotes government revenue, and $W = CS + PS + R$ denotes social welfare.[3] Finally, we assume that tax revenue and profits are redistributed to the consumers as a lump-sum transfer. We make the standard assumption that the consumer treats tax revenue and profits as fixed when choosing consumption, failing to consider the external effects on the lump-sum transfer.

Let p denote the producer price and $p + t$ denote the price paid by consumers. We assume that there is a mass 1 of consumers and we index each consumer by i. Consumer i has exogenous income Z_i and quasilinear utility given by $U_i(q, y) = u_i(q) + y$, where q is consumption of the taxed good and y is the numeraire good. Given the assumption of quasilinear utility, the consumer will choose to allocate the lump-sum transfer to the outside market y. The assumption of quasilinear utility, which is standard in the literature, is a convenient assumption as it allows us to use consumer surplus to measure welfare. We follow Chetty, Looney and Kroft (2009) by assuming that 1) taxes affect utility only through their effects on the chosen consumption bundle and that 2) in the absence of taxation, consumers perfectly optimize so that $p = u_i'(q)$ when $t = 0$. We define willingness to pay for

[3] It is well known that unit taxes and ad valorem taxes are equivalent under perfect competition. Section 4 considers the case of an ad valorem tax under imperfect competition.

consumer i as $wtp_i(q) \equiv u_i'(q)$ and marginal willingness to pay for consumer i as $mwtp_i(q) \equiv u_i''(q)$. When incomplete salience causes optimization errors, the marginal loss of foregone consumption of the numeraire commodity is constant.

Let $D_i(p,t)$ be the observed demand of individual i. Our specification for observed demand permits prices and taxes to have different effects, following Chetty, Looney and Kroft (2009). Assume that for $t > 0$, $D_i(p,0) > D_i(p,t) > D_i(p+t,0)$. By strict monotonicity and continuity, for all p and t there exists a $\theta_i(p,t) \in (0,t)$ such that $D_i(p+\theta_i(p,t),0) = D_i(p,t)$. For fixed t and all i, we assume that if $D_i(p+\theta_i,0) = D_i(p,t)$ for some price p, then $D_i(p'+\theta_i,0) = D_i(p',t)$ for any other price p'. This implies that $\theta_i(p,t) = \theta_i(t)$ does not depend on the producer price p. We follow the literature and assume that $\theta_i(t)$ is linear and write it as $\theta_i(t) = \theta_i t$ which is without loss of generality on the shape of the original inverse demand curve $u_i'(q) = wtp_i(q)$. This definition satisfies $\theta_i = \frac{\frac{\partial D_i}{\partial t}}{\frac{\partial D_i}{\partial p}}$ which is how this parameter is defined in Chetty, Looney and Kroft (2009), but we allow for consumer heterogeneity following Taubinsky and Rees-Jones (2018) and Farhi and Gabaix (2020). Total quantity demanded is given by $D(p,t) = \int D_i(p,t)di$. We assume that $D(p,t)$ is strictly decreasing in both arguments and continuous. Let $\epsilon_D \equiv -\frac{\partial D(p,t)}{\partial p}\frac{p+t}{D}$ denote the price elasticity of demand evaluated at the consumer price.

We define production similar to Chetty, Looney and Kroft (2009) and Taubinsky and Rees-Jones (2018). In particular, firms are price takers and use $c(S)$ units of the numeraire good to produce S units of output. The marginal cost of production is $c'(S)$ and we assume that firms perfectly optimize so that firm supply is given by $p = c'(S(p))$ where $S(p)$ is strictly increasing and continuous in p. Define $\epsilon_S \equiv \frac{\partial S}{\partial p}\frac{p}{S}$ as the price elasticity of supply.

The equilibrium price, p, in the market for the taxed good is determined by the condition $D(p,t) = S(p)$. Let $\epsilon_{Dt} \equiv \frac{dq(t)}{dt}\frac{p+t}{q(t)}$ be the elasticity of equilibrium output, $q(t) \equiv D(p(t),t)$, with respect to the tax t. Note that ϵ_{Dt} need not equal $\frac{\partial D}{\partial t}\frac{p+t}{q(t)}$; the latter holds the pre-tax price, p, fixed, while the former includes any indirect effect of taxes on the producer price. We denote the pass-through rate by $\rho \equiv 1 + dp/dt$.

We begin by introducing a technical assumption which helps to simplify the analysis throughout and connect our formulas to ones that exist in the literature.

Assumption 1. *The demand function $D_i(p,t)$ can be represented by the linear approximation $\hat{D}_i(p,t) = q_{i0} + \frac{\partial D(p_0,t_0)}{\partial p}(p - p_0 + \theta_i(t - t_0))$ around $(q_{i0}, p_0, t_0, \theta_i)$ for $q_{i0} = D_i(p_0, t_0)$ for each individual i.*[4]

Assumption 1 is a formal statement of the approximation given in Bernheim and Taubinsky (2018) that allows us to focus on heterogeneity in salience effects across consumers, holding the price responses across consumers constant. The expression in Assumption 1 is a first-order approximation rather than an exact expression because even assuming price responses are the same across consumers at all quantity levels is not sufficient for Assumption 1 to hold exactly unless demand curves are linear. We next introduce a lemma which turns out to be quite useful in deriving all of the incidence formulas that we present in the paper.

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

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

Under Assumption 1, we obtain the following relationship:

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

Proof. See Appendix. □

Given the definitions and Lemma 1, we can derive the following proposition and corollary for the incidence and efficiency costs of taxation. The corollary uses Assumption 1 to provide expressions that ignore heterogeneity in price responses across consumers. We note that while the results on efficiency already exist for perfect competition when consumers are heterogeneous (see Taubinsky and Rees-Jones 2018), the results in this section on incidence are novel. For example, Chetty, Looney

[4]In case this assumption is violated, the model given by $\hat{D}_i(p,t)$ is a linear approximation to the real model with a common slope for all i. The corollaries that follow below apply to this linear approximation.

and Kroft (2009) consider incidence under the assumptions of identical consumers and no pre-existing taxes.

Proposition 1. *Define* $q_i(t) \equiv D_i(p(t),t)$ *and* $q(t) \equiv D(p(t),t)$. *The incidence on consumers, producers, government, the pass-through rate and the marginal excess burden of a unit tax,* t, *under* ***perfect competition*** *may be expressed as:*

$$
\begin{aligned}
\frac{dCS}{dt} =& -\rho q - (1-\mathbb{E}(\theta_i))t\frac{dq}{dt} + tCov\left(\theta_i, \frac{dq_i}{dt}\right) \\
\frac{dPS}{dt} =& -(1-\rho)q \\
\frac{dR}{dt} =& q + t\frac{dq}{dt} \\
\rho =& 1-(1-\omega)\left(\mathbb{E}(\theta_i) + \frac{Cov\left(\theta_i, \frac{\partial D_i}{\partial p}\right)}{\frac{\partial D}{\partial p}}\right), \textit{ where } \omega \equiv \frac{1}{1+\frac{\epsilon_D}{\epsilon_S}\frac{p}{p+t}} \\
I =& \frac{\rho}{1-\rho} + \frac{1-\mathbb{E}(\theta_i)}{1-\rho}\frac{t}{p+t}\epsilon_{Dt} - \frac{t}{q(1-\rho)}Cov\left(\theta_i, \frac{dq_i}{dt}\right) \\
\frac{dW}{dt} =& t\mathbb{E}(\theta_i)\frac{dq}{dt} + tCov\left(\theta_i, \frac{dq_i}{dt}\right)
\end{aligned}
$$

Proof. See Appendix. □

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

$$
\begin{aligned}
\frac{dCS}{dt} =& -\rho q - (1-\mathbb{E}(\theta_i))t\frac{dq}{dt} + tVar\left(\theta_i\right)\frac{\partial D}{\partial p} \\
\frac{dPS}{dt} =& -(1-\rho)q \\
\rho =& 1-(1-\omega)\,\mathbb{E}(\theta_i), \textit{ where } \omega \equiv \frac{1}{1+\frac{\epsilon_D}{\epsilon_S}\frac{p}{p+t}}
\end{aligned}
$$

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

$$
\begin{aligned}
\frac{dW}{dt} &= t\mathbb{E}(\theta_i)\frac{dq}{dt} + tVar\left(\theta_i\right)\frac{\partial D}{\partial p} \\
&= t\mathbb{E}(\theta_i)\frac{\partial D}{\partial p}(\rho - 1 + \mathbb{E}(\theta_i)) + tVar\left(\theta_i\right)\frac{\partial D}{\partial p}
\end{aligned}
$$

We highlight several features of Proposition 1 and Corollary 1. First, when $t = 0$, the formulas for the effects of a tax on consumer surplus and producer surplus, and hence incidence, are identical to Weyl and Fabinger (2013), except that the pass-through term, ρ, is indirectly affected by salience effects.[5] 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 and producers at the market level through changes in prices, as in Chetty, Looney, and Kroft (2009). In particular, since $\omega < 1$ with perfect competition, an increase in $\mathbb{E}(\theta_i)$ leads to a lower pass-through and incidence on consumers. We also see that, in the presence of heterogeneous consumers, pass-through depends on the new term $Cov\left(\theta_i, \frac{\partial D_i}{\partial p}\right)$. 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 $\frac{\partial D_i}{\partial t} = \theta_i \frac{\partial D_i}{\partial p}$, the market-level response to taxes depends on the covariance between θ_i and $\frac{\partial D_i}{\partial p}$ across i.

Second, 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 $tCov\left(\theta_i, \frac{dq_i}{dt}\right)$.[6] In this case, one has to account for behavioral responses to the tax since the envelope theorem does not apply when consumers misoptimize in the baseline equilibrium. The first term, $-(1 - \mathbb{E}(\theta_i))t\frac{dq}{dt}$, resembles the "self-control adjustment" term in equation (10) in Gruber and Kőszegi (2004). It enters $\frac{dCS}{dt}$ positively whenever $\frac{dq}{dt} < 0$ and we see 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 optimal choice. The second term, $tCov\left(\theta_i, \frac{dq_i}{dt}\right)$, represents a misallocation term. When Assumption 1 holds, Corollary 1 shows that this term collapses to $tVar\left(\theta_i\right)\frac{\partial D}{\partial p}$ which

[5]It is analytically convenient to express the pass-through formula this way where ω is the pass-through rate when consumers fully optimize with respect to taxes, as it will facilitate a comparison between the different cases (perfect competition, monopoly, and imperfect competition).

[6]Chetty, Looney and Kroft (2009) fully characterized the effects of a tax on consumer surplus in terms of the pass-through rate, ρ. Our results show that pass-through is not sufficient for incidence when there are both salience effects and pre-existing taxes in the market.

mirrors the expression in Taubinsky and Rees-Jones (2018) and Bernheim and Taubinsky (2018). If consumers are fully attentive to the tax so that $\theta_i = 1$ for all i, we see that $\frac{dCS}{dt}$ and I are characterized purely by the price effect, even with a pre-existing tax.

Third, independent of the baseline tax or the degree of inattention to the tax, when supply is perfectly elastic ($\epsilon_S = \infty$), $\rho = 1$ and the full burden of the tax is on consumers so that $I = \infty$.

Lastly, the marginal excess burden of the tax is scaled by the degree of inattention to the tax, $\mathbb{E}(\theta_i)$, and includes an additional term reflecting the dispersion in inattention, as in Taubinsky and Rees-Jones (2018). In the case where $\theta_i = 0$ for all i, taxes are not distortionary since with quasilinear utility, the consumption allocation is the same as the consumption allocation with a lump-sum tax, as shown in Chetty, Looney, and Kroft (2009). In the case where there is full pass-through ($\rho = 1$) and homogeneous θ_i ($\mathbb{E}(\theta_i) = \theta$), the marginal excess burden is a quadratic function of θ, $\frac{dW}{dt} = \theta^2 t \frac{\partial D}{\partial p}$. In the case where there are no pre-existing taxes, introducing a small tax into a market with or without salience effects has no first-order effect on welfare.

3 Imperfect Competition

3.1 Monopoly

In this section, we depart from the benchmark case of perfect competition and consider a model of imperfect competition. In order to develop intuition, we begin with the special case of monopoly. We assume that the monopolist's cost of production is given by $c(q)$, with marginal cost $mc(q) \equiv c'(q)$, and we define $\epsilon_S \equiv \frac{c'(q)}{c''(q)q}$, which follows the definition in Weyl and Fabinger (2013). When consumers are identical, the monopoly problem is particularly simple since in this case, $\theta(p,t) = \theta t$ and $D(p+\theta t, 0) = D(p,t)$ and we may express the inverse demand function facing the firm as $P(q,t) = wtp(q) - \theta t$, where $wtp(q)$ is the inverse of $D(\cdot, 0)$. The monopolist's problem in this case can be stated as:

$$\max_q \left(wtp(q) - \theta t\right) q - c(q)$$

The first-order condition for the monopoly problem is $wtp'(q)q+wtp(q)-\theta t = mc(q)$. In this case, $mr(q) = wtp'(q)q+wtp(q)$ is shifted down by θt. If the tax was fully non-salient so that $\theta = 0$, then consumer demand is not affected by taxes.

In the general case with consumer heterogeneity, we follow the setup from last section where for each i, $D_i(p+\theta_i t, 0) = D_i(p,t)$, and $D(p,t) \equiv \int D_i(p,t)di$. As before, the market demand elasticity is defined as $\epsilon_D \equiv -\frac{\partial D(p,t)}{\partial p}\frac{p+t}{D}$. We now introduce several new definitions which are relevant for characterizing incidence and efficiency under imperfect competition. First, we define the representative agent's willingness to pay $wtp(q)$ as the inverse of $D(\cdot, 0)$.[7] Next, define the marginal willingness to pay as $mwtp(q) \equiv wtp'(q)$. Then $ms(q) \equiv -mwtp(q)q$ is marginal consumer surplus and the elasticity of inverse marginal surplus is given by $\epsilon_{ms} \equiv \frac{ms(q)}{ms'(q)q}$. Furthermore, define $MS(q,t) = -\frac{q}{\frac{\partial D}{\partial p}(p(t),t)} = \frac{ms(q)}{mwtp(q(t)) * \frac{\partial D}{\partial p}(p(t),t)}$. Note that $MS(q,0) = ms(q)$, and define $MS_t \equiv \frac{\partial MS}{\partial t}$.

Finally, we assume that tax revenue $R = tq$ and profits π are redistributed to the consumers as a lump-sum transfer. The consumer treats 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, the consumer 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).

$$W(p,t) = \underbrace{\int u_i(q_i(p,t))di - (p+t)q(p,t)}_{CS} + \underbrace{pq - c\,(q)}_{PS} + \underbrace{tq}_{R}$$

Given these definitions, we can now characterize the incidence and marginal excess burden of taxes for monopoly.

Proposition 2. *The incidence on consumers, producers, government, the pass-through rate and the*

[7] 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.

*marginal excess burden of a unit tax, t, under **monopoly** may be expressed as:*

$$\frac{dCS}{dt} = -\rho q - (1-\mathbb{E}(\theta_i))t\frac{dq}{dt} + tCov\left(\theta_i, \frac{dq_i}{dt}\right)$$

$$\frac{dPS}{dt} = -q\left(\mathbb{E}(\theta_i) + \frac{Cov\left(\theta_i, \frac{\partial D_i}{\partial p}\right)}{\frac{\partial D}{\partial p}}\right)$$

$$\frac{dR}{dt} = q + t\frac{dq}{dt}$$

$$\rho = 1 - (1-\omega)\left(\mathbb{E}(\theta_i) + \frac{Cov\left(\theta_i, \frac{\partial D_i}{\partial p}\right)}{\frac{\partial D}{\partial p}}\right) + \omega MS_t,\ \textit{where}\ \omega = \frac{1}{1 + \frac{\epsilon_D \frac{p}{p+t} - 1}{\epsilon_S} + \frac{1}{\epsilon_{ms}}}$$

$$I = \frac{\epsilon_D}{\frac{p+t}{q}\mathbb{E}\left(\theta_i \frac{\partial D_i}{\partial p}\right)}\left(\rho + (1-\mathbb{E}(\theta_i))\frac{t}{p+t}\epsilon_{Dt} - \frac{t}{q}Cov\left(\theta_i, \frac{dq_i}{dt}\right)\right)$$

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

Proof. See Appendix. □

Corollary 2. *Under Assumption 1, the effect of the tax on consumer surplus, producer surplus, pass-through, incidence and welfare can be expressed as:*

$$\frac{dCS}{dt} = -\rho q - (1-\mathbb{E}(\theta_i))t\frac{dq}{dt} + tVar\left(\theta_i\right)\frac{\partial D}{\partial p}$$

$$\frac{dPS}{dt} = -q\mathbb{E}(\theta_i)$$

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

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

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

Several interesting insights emerge from the analysis of salience and taxation under monopoly. First, we note that the formula characterizing the effects of the tax on consumer surplus is identical to the formula in the case of perfect competition. Note, however, that the *inputs* to the formula are different under monopoly as we discuss below.

Next, we see that the effects of a tax on producer surplus is $-q\left(\mathbb{E}(\theta_i) + \frac{Cov\left(\theta_i, \frac{\partial D_i}{\partial p}\right)}{\frac{\partial D}{\partial p}}\right)$. Consider first the case where $\theta_i = 1$ for all i. Since the monopolist sets the price (and 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 effect of taxes on producers since instead of consumer demand falling by the amount of the tax change, it falls by this amount scaled by the degree of inattention $\mathbb{E}(\theta_i)$. The covariance term $Cov\left(\theta_i, \frac{\partial D_i}{\partial p}\right)$ incorporates the correlation between θ_i and $\frac{\partial D_i}{\partial p}$ which determines the market-level demand response to the tax. When $Cov\left(\theta_i, \frac{\partial D_i}{\partial p}\right) > 0$, 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 and those who are fully inattentive to taxes. If those who optimize are price inelastic and those who 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. Finally, we note that the formula holds even when $mc(q)$ is constant so that $\epsilon_S = \infty$. This contrasts with perfect competition where $\frac{dPS}{dt} = 0$ when $\epsilon_S = \infty$.

Third, 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, unlike the case of perfect competition, the monopoly outcome may be associated with

[8]Since $\frac{\partial D_i}{\partial p} < 0$, this requires that consumers that are attentive to taxes are price inelastic; in other words, the absolute value of $\frac{\partial D_i}{\partial p}$ is negatively correlated with θ_i. Note that DellaVigna and Gentzkow (2019) find that high-income stores face less price elastic consumers and Taubinsky and Rees-Jones (2018) find that θ_i is higher for higher income individuals. This evidence suggests that $Cov\left(\theta_i, \frac{\partial D_i}{\partial p}\right) > 0$.

[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.

$\omega > 1$ which implies that an increase in $\mathbb{E}(\theta_i)$ raises incidence on consumers. To see this, consider the case of 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.[10] We also see that the expression for ρ in the case of monopoly depends additionally on MS_t. Up to first order this term can be approximated by $MS_t \approx \frac{-q}{\left(\frac{\partial D}{\partial p}\right)^2} Cov\left(\frac{\partial^2 D_i}{\partial p^2}, \theta_i\right)$ (see 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, the monopolist exploits this change in market power when re-optimizing prices. If more attentive consumers become more price elastic when taxes change, then $Cov\left(\frac{\partial^2 D_i}{\partial p^2}, \theta_i\right) < 0$ 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 $Cov\left(\frac{\partial^2 D_i}{\partial p^2}, \theta_i\right) < 0$, 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. Corollary 2 shows that this term vanishes under Assumption 1.

Finally, conditional on $\frac{dq}{dt}$, the effects of salience on the marginal excess burden of the tax operate in similar ways under perfect competition and monopoly through the terms $\mathbb{E}(\theta_i)t$ and $tCov\left(\theta_i, \frac{dq_i}{dt}\right)$. However, under monopoly 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 monopolistic markets as compared to perfectly competitive markets. To see this, note that we can express $\frac{dq}{dt} = \frac{\partial D}{\partial p}(\rho - 1 + \mathbb{E}(\theta_i))$ as shown in Corollary 2. Thus, salience enters linearly in the welfare formula $\frac{dW}{dt}$ and interacts directly with the markup. For example, with full pass-through and

[10] Note that even when $\epsilon_S = \infty$, the full incidence is not on consumers, unlike the case of perfect competition, although we note that this result holds independent of salience effects.

homogeneous consumers, $\frac{dW}{dt} = (p - mc(q))\,\theta\frac{\partial D}{\partial p} + \theta^2 t\frac{\partial D}{\partial p}$. 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.

To summarize, the analysis of the incidence and welfare consequences of a tax for the special case of monopoly 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. This motivates our analysis of tax salience in a more general model of imperfect competition.

3.2 Symmetric Imperfect Competition

We consider a differentiated product market (the "inside market") which is subject to a unit tax t on each product in the market. 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).[11] 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 the individual consumes $q^i = \frac{Q^i}{J}$ units of each product $j = 1, \ldots, J$, where Q^i

[11]We now use a superscript to index individuals since there is an additional dimension of heterogeneity coming from asymmetric products. We use the subscript j to index products.

is the aggregate quantity consumed by the individual.

Following Chetty, Looney and Kroft (2009), consumer i demand for product j is given by $q_j^i = q_j^i(p_1, \dots, p_J, t)$ 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 $q_j^i(p_1, \dots, p_J, t)$ to consumer willingness to pay. We thus make the following assumptions which mirror assumptions $A1$ and $A2$ in Chetty, Looney and Kroft (2009).

Assumption 2. *Taxes affect utility only through their effects on the chosen consumption bundle. Indirect utility is given by:*

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

Assumption 2 requires that taxes or salience have no impact on utility beyond their effects on consumption.

Assumption 3. *When tax-inclusive prices are fully salient, the agent chooses the same allocation as a fully-optimizing agent.*

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

Assumption 3 implies that when tax-inclusive prices are fully salient, agents maximize utility. As in Section 2 we allow 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 $q^i(p, t)$ demand 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 we assume $q^i(p, t) = q^i(p + \theta_i t, 0)$ for some $\theta_i > 0$; therefore, 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.

We define individual i's market demand as $Q^i(p, t) = Jq^i(p, t)$. Total market demand is then given by $Q(p, t) = \int Q^i(p, t) di$, from where we define the market demand elasticity $\epsilon_D \equiv -\frac{\partial Q(p,t)}{\partial p} \frac{p+t}{Q}$. 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. 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 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.[12] By letting $\nu_q = \frac{1}{mwtp(Q)} \times \frac{1}{\frac{dq_j}{dp_j}} = \frac{1}{mwtp(Q)} \times \frac{1}{\frac{\partial q_j}{\partial p_j} + \nu_p \sum_{k \neq j} \frac{\partial q_j}{\partial p_k}}$ 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

[12] 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.

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. The consumer treats 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, the consumer 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).

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

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

Proposition 3. *The incidence on consumers, producers, government, the pass-through rate and the marginal excess burden of a unit tax, t, under* ***symmetric imperfect competition*** *may be expressed as:*

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

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

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

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

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

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

Proof. See Appendix. □

Corollary 3. *Under Assumption 1, the effect of the tax on consumer surplus, producer surplus, pass-*

through, incidence and welfare can be expressed as:

$$\frac{dCS}{dt} = -\rho Q - (1-\mathbb{E}(\theta_i))t\frac{dQ(p(t),t)}{dt} + tVar\left(\theta_i\right)\frac{\partial Q}{\partial p}$$
$$\frac{dPS}{dt} = -Q\left[\left(1-\frac{\nu_q}{J}\right)(1-\rho) + \frac{\nu_q}{J}\mathbb{E}(\theta_i)\right]$$
$$\rho = 1-(1-\omega)\,\mathbb{E}(\theta_i),\ \textit{where}\ \omega = \frac{1}{1+\frac{\epsilon_D\frac{p}{p+t}-\frac{\nu_q}{J}}{\epsilon_S}+\frac{\frac{\nu_q}{J}}{\epsilon_{ms}}}$$

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

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

Proposition 3 leads to several additional insights. First, $\frac{dCS}{dt}$ has the same expression as perfect competition and monopoly. Thus, the change in consumer surplus does not depend directly on market conduct, except insofar as market conduct determines pass-through. Second $\frac{dPS}{dt}$ is a convex combination of the monopoly and perfect competition cases with weights $\frac{\nu_q}{J}$ and $1-\frac{\nu_q}{J}$, respectively. To understand this expression, note that when $\theta_i = 1$ for all i, $\frac{dPS}{dt} = -Q\left((1-\rho)+\rho\frac{\nu_q}{J}\right)$, similar to Weyl and Fabinger (2013). When firms have market power, they internalize the change in their own output on the market (given by $\frac{\nu_q}{J}$), and so we need to adjust the price effect by $\rho\frac{\nu_q}{J}$. Under monopoly, this effect becomes ρ and $\frac{dPS}{dt} = -Q$. Similar to monopoly, if firms have market power, then salience directly attenuates the reduction in demand due to taxes and thus reduces the tax burden on producers.

We again see that in the general case, there are interesting effects of salience on pass-through depending on the magnitudes of ν_q and ϵ_{ms}. As in the monopoly case, greater attention to taxes can increase the burden on consumers if there is overshifting of taxes, again illustrating that salience and the degree of competition *interact* in determining the relative incidence of taxes on consumers and producers. In the simpler case of Corollary 3, the only role that salience plays is attenuating the initial response of consumer demand to a change in taxes. The intuition for this result is that, conditional on the demand response to taxes, salience does not directly affect the equilibrium response of prices. Conditional on this response, the firm's equilibrium response to the tax is determined purely by standard forces such as the marginal cost function, the shape of the demand curve and market conduct. Mathematically, this is because ω depends on supply and demand fundamentals and the conduct parameter. Salience only affects the weighting on ω in the expression for ρ. Thus, conditional on ω, the role of salience in affecting pass-through will be similar whether firms have a lot of market power (as in Monopoly) or only a little bit of market power (as in Bertrand or Cournot). However, in Proposition 3 we see that the expression for ρ depends additionally on $\omega \frac{\nu_q}{J} MS_t$. Like in monopoly, 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 Appendix), but now it is weighted by $\frac{\nu_q}{J}$ — the relative distance of firm behavior to the monopoly benchmark. In the previous section we explain how this term captures that when taxes change, because consumers vary in their degree of inattention, there is a reallocation effect that effectively changes the elasticity of the demand curve.

Finally, we see that the marginal excess burden formula depends on the same set of sufficient statistics as in the monopoly case. In particular, the conduct parameter does not appear in the formula, and thus the intuition for welfare in the monopoly case carries over to the general model. Intuitively, the marginal excess burden of taxes is the lost social surplus that accrues from discouraging transactions in which the value of the product exceeds the cost of production. The marginal value of product with quasilinear utility is simply $p + \mathbb{E}(\theta_i)t$. The marginal cost of production is $mc(q)$. The discouraged transactions is represented by $dQ(p(t), t)/dt$ and we have the misallocation term which depends on $Var(\theta_i)$. Note that price effects do not show up in the formula since they are transfers between consumers and firms. Of course, the inputs to the formula, such as p and $mc(q)$, will depend on market conduct, but conditional on them, conduct does not independently affect the marginal excess burden.

4 Ad Valorem versus Unit Taxes

It is well known that ad valorem and unit taxes are not equivalent in imperfectly competitive markets (Delipalla and Keen 1992, Anderson, de Palma and Kreider 2001a, Adachi and Fabinger 2019). This section extends our results on incidence and excess burden in Proposition 3 to ad valorem taxes in the presence of salience effects. We consider the model of imperfect competition with both unit taxes and ad valorem taxes. The purpose of the model is to compare the incidence and welfare effects of these taxes and to forge a link with the empirical section which considers ad valorem taxes. For ease of exposition, we assume identical consumers and present the general expressions for ad valorem taxes in the presence of heterogeneous consumers that we calibrate in Section 6.

Let p denote the producer price and let $p(1+\tau)+t$ denote the price paid by consumers where τ is the ad valorem tax and t is the unit tax. Demand is given by $D(p,t,\tau)$ and assume that for $\tau > 0$ and $t > 0$, $D(p,0,0) > D(p,t,\tau) > D(p(1+\tau)+t,0,0)$. For any triple (p,t,τ) there exists $\theta_\tau(p,t,\tau)$ and $\theta_t(p,t,\tau)$ to be such that: $D(p,t,\tau) = D(p(1+\theta_\tau\tau)+\theta_t t,0,0)$. However following the literature and to simplify the setup assume θ_τ and θ_t are independent of the level of prices and tax rates. Equivalently we could define $\theta_\tau \equiv \frac{\frac{\partial D}{\partial \tau}}{\frac{\partial D}{\partial p}} \times \frac{1}{p}$ and $\theta_t \equiv \frac{\frac{\partial D}{\partial t}}{\frac{\partial D}{\partial p}}$ and assume they are constant with respect to prices and taxes.[13] Following the prior section, we extend the definition of willingness to pay to accommodate the ad valorem tax so that $wtp(Q) = p(1+\theta_\tau\tau)+\theta_t t$.

Let $\epsilon_D \equiv -\frac{\partial Q}{\partial p}\frac{p(1+\tau)+t}{Q}$, $\epsilon_D^* = \epsilon_D \frac{p}{p(1+\tau)+t}$ and define the pass-through rates for ad valorem and unit taxes respectively, as $\rho_\tau \equiv \frac{1}{p}\frac{\partial(p(1+\tau)+t)}{\partial \tau}$ and $\rho_t \equiv \frac{\partial(p(1+\tau)+t)}{\partial t}$. The following lemma shows how to identify θ_τ with commonly observable objects.

Lemma 2. *Let* $\epsilon_{D\tau} \equiv \frac{dQ}{d\tau}\frac{p(1+\tau)+t}{Q}$. *The following relationship holds:*

$$\epsilon_{D\tau} = -\epsilon_D * \frac{p}{1+\tau}\left((1+\theta_\tau\tau)\rho_\tau + \theta_\tau - 1\right)$$

and

$$\theta_\tau = \frac{(1-\rho_\tau)\, p\epsilon_D - \epsilon_{D\tau}(1+\tau)}{(1+\tau\rho_\tau)\, p\epsilon_D}$$

Proof. See Appendix. □

With Lemma 2 in hand, we can now state our main proposition for ad valorem taxes. Following the literature, we compare the pass-through rates and the marginal cost of public funds. A lower marginal

[13] Note that in the denominator of θ_τ and θ_t, the derivative is with respect to the first argument of D.

cost of public funds indicates greater efficiency. We begin with the characterization of pass-through rates.

Proposition 4. *In the symmetric model of imperfect competition, the pass-through rates for ad valorem and unit taxes are given respectively as:*

$$\rho_\tau = 1 - \frac{(1+\tau)\theta_\tau}{1+\theta_\tau \tau}\left(1 - \omega\frac{mc(q)}{p}\right)$$

$$\rho_t = 1 - \frac{(1+\tau)\theta_t}{1+\theta_\tau \tau}(1-\omega)$$

where $\omega = \frac{1}{1+\frac{(1+\theta_\tau\tau)\epsilon_D^* - \frac{\nu_q}{J}}{\epsilon_S} + \frac{\nu_q}{J}\frac{1}{\epsilon_{ms}}}$.

This implies that the two pass-through rates can be ranked based on the following:

$$\frac{\rho_\tau - 1}{\rho_t - 1} = \frac{\theta_\tau}{\theta_t}\frac{\omega\frac{mc}{p} - 1}{\omega - 1} = \frac{\theta_\tau}{\theta_t}\left(1 - \frac{\omega}{\omega - 1}\frac{\nu_q}{J\epsilon_D^*}\right)$$

Proof. See Appendix. □

A first observation is that when $\theta_\tau = \theta_t$, if $mc < p$ then $\rho_\tau < \rho_t$ which is consistent with the literature (Delipalla and Keen 1992; Adachi and Fabinger 2019). Thus, if consumers underreact to ad valorem and unit taxes similarly, the pass-through rate is lower for ad valorem taxes. A new observation is that even under perfect competition starting from $p = mc$, ad valorem taxes imply a higher pass-through than unit taxes $\rho_t < \rho_\tau$ if and only if the consumers are more responsive to ad valorem taxes than unit taxes $\theta_\tau > \theta_t$.[14] Most of the available empirical evidence in the literature applies to sales taxes and thus, θ_τ. Our results stress the need for additional evidence on θ_t.

Next, we derive the marginal cost of public funds for an ad valorem tax and a unit tax which are defined as $MC_\tau \equiv -\frac{dW/d\tau}{dR/d\tau}$ and $MC_t \equiv -\frac{dW/dt}{dR/dt}$, respectively.

Proposition 5. *Denote* $wtp = p(1+\theta_\tau\tau) + \theta_t t$ *the perceived price by the consumer and* $\epsilon_D^* = \epsilon_D \frac{p}{p(1+\tau)+t}$. *The marginal cost of public funds for an ad valorem tax,* τ, *and a unit tax,* t, *under*

[14] As a basic matter of tax administration, this is relatively unlikely. Indeed, it has been suggested to us that the relative saliency of unit taxes appears to have played an important role in dictating the implementation details of recently-adopted beverage taxes.

symmetric imperfect competition *may be expressed as:*

$$MC_\tau = \epsilon_D^* \frac{\frac{wtp-mc}{p}}{\frac{1+\tau\rho_\tau}{(1+\theta_\tau\tau)\rho_\tau+\theta_\tau-1} - \epsilon_D^*(\tau+\frac{t}{p})}$$

$$MC_t = \epsilon_D^* \frac{\frac{wtp-mc}{p}}{\frac{1+\tau\rho_t}{(1+\theta_\tau\tau)\rho_t+\theta_t-1} - \epsilon_D^*(\tau+\frac{t}{p})}$$

This implies the following:

$$\frac{MC_t}{MC_\tau} = \frac{\frac{1+\tau\rho_\tau}{(1+\theta_\tau\tau)\rho_\tau+\theta_\tau-1} - \epsilon_D^*(\tau+\frac{t}{p})}{\frac{1+\tau\rho_t}{(1+\theta_\tau\tau)\rho_t+\theta_t-1} - \epsilon_D^*(\tau+\frac{t}{p})}$$

In other words, the cost of ad-valorem taxes is lower than the cost of unit taxes ($MC_\tau < MC_t$) if and only if

$$\theta_\tau\left[1-\frac{1+\tau(1+\theta_\tau-\theta_t)}{1+\theta_\tau\tau}\left(1-\omega\frac{mc}{p}\right)\right] < \theta_t\left[1-\frac{1+\tau}{1+\theta_\tau\tau}(1-\omega)\right]$$

Proof. See Appendix. □

It is instructive to consider the benchmark case where $\theta_\tau = \theta_t$. In this case, $MC_\tau < MC_t$ if and only if $p > mc$. Thus, as long as consumers respond symmetrically to ad valorem and unit taxes, then salience does not affect the well-known result that ad valorem taxes are more efficient than unit taxes under imperfect competition. Of course, if consumers are sufficiently more attentive to ad valorem taxes than unit taxes, then this result shows that ad valorem taxes can be more distortionary than unit tax.

5 Data and Estimation

5.1 Data Description

Nielsen Retail Scanner Data We measure prices and quantity using the Nielsen Retail Scanner (RMS) data from $2006-2014$. This data set 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.

The UPCs are organized by Nielsen according to a hierarchical structure.[15] 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). 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 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. In terms of products, 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 modules account for almost 80 percent of the total sales in grocery stores in the Nielsen data.[16]

The key variables for our empirical analysis are price and quantity. 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

[15] 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.

[16] 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).

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.[17] 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.

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.[18] 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 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.[19] 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).[20]

[17]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.

[18]All data sources used to determine the exemption status of products are listed in Appendix Table OA.2.

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

[20]The Online Appendix shows the cross-sectional variation in sales tax rates and sales tax exemptions in our data. Appendix Figure OA.1 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. Appendix Figure OA.2 reports the food tax exemptions across states and shows that many of the states that tax food are located in

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 in our emprical 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 Sample As a last step in constructing our analysis sample, we merge the tax data onto the Retail Scanner data. The stores in the Nielsen 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

either the South or the Midwest. Appendix Figure OA.3 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. Appendix Figure OA.4 shows the number of tax changes within counties during the same time period.

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 sample includes 8,652 grocery stores, and includes price, quantity and tax rates for 198 modules in 1,460 counties over 36 year-quarters.

5.2 Estimation Strategy

The Effect of Taxes on Prices and Quantity Our main specification is a "constant effects" model which can be derived from the model above by assuming that consumers have identical demand functions. We estimate the effect of sales taxes on consumer prices and quantity using two complementary regression models. The first model uses the full set of counties from the Neilsen Retail Scanner data using the following estimating equation:

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

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 do not change within counties in ways that are correlated with changes in consumer demand (conditional on the fixed effects). This model allows for arbitrary trends across states and modules and thereby relies on within-county-over-time variation in tax rates. The estimate β^y can be interpreted as the elasticity of prices or quantity with respect to taxes ($\beta^{p(1+\tau)}$ and β^Q, respectively).

The second regression model uses a subsample of counties and a "county border pair" research design, following Holmes (1998) and Dube, Lester and Reich (2010). This alternative approach identifies the effects of sales taxes using pairs of counties on opposite sides of a state border and 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 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. One way this assumption could fail is if counties adjust their tax rates based on economic conditions within the border pair. To address this concern, we also report results in Appendix Table OA.3 in which we instrument the total tax rate with the state sales tax rate (and find similar results).

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.[21] 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 $\widehat{\beta^{p(1+\tau)}} = 0.961$ (s.e. 0.045) indicates a large but incomplete amount of pass-through of taxes onto consumer prices.[22] The next row reports the estimate $\widehat{\beta^Q} = -0.668$ (s.e. 0.185), indicating a meaningful quantity response to tax changes. The results in column (2) show similar results using the county border pair approach. The similar results across the columns is consistent with limited endogeneity bias in the full sample.[23]

[21] Clustering by state-module is more conservative than clustering by state-by-module-time. 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. The primary uncertainty we want to account for in our statistical inference is the design-based uncertainty from state and local policy changes that affect sales tax rates.

[22] Classical measurement error in effective tax rates biases our estimates of $\beta^{p(1+\tau)}$ towards 1. Instrumental variable estimates of the effect of sales taxes on prices presented in Appendix Table OA.1 are slightly smaller than their corresponding border-sample OLS estimates, suggesting a small amount of attenuation bias.

[23] The minimum wage literature has also tended to find that results using all counties and "traditional" fixed effects

Tax salience parameter Since the main specification is a "constant effects'" specification with no heterogeneity across consumers in terms of demand responses, there is no heterogeneity in the tax salience parameter, θ_τ. In this setup, 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. 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 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 that there are correlated demand shocks across stores 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, since these supply-side instruments will identify the average price elasticity of demand. Intuitively, this approach requires that chains select store locations based on overall demand factors (that are common across modules), but not module-specific demand

models are similar to results based on the county border pair approach (see, e.g., Dube, Lester, and Reich 2010).

factors. In Appendix Table OA.4, we report the reduced-form 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 insensitive to using this alternative choice of instrument.

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

$$\begin{aligned} \log(p(1+\tau)_{mrn}) &= \lambda z_{mrn} + \kappa'_{msn} + \kappa'_{mr} + \upsilon'_{mrn} \\ \log Q_{mrn} &= \alpha \log(p(1+\tau)_{mrn}) + \kappa_{msn} + \kappa_{mr} + \upsilon_{mrn} \end{aligned}$$

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 $\widehat{\alpha} = -1.202$ (s.e. 0.027), and for the border pair subsample the estimate is $\widehat{\alpha} = -1.223$ (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 formula in Lemma 2 evaluated at $t = 0$, which we re-arrange slightly to more closely line up with the empirical estimates:

$$\theta_\tau = \frac{(1-\rho_\tau)\,\tilde{\epsilon}_D + \tilde{\epsilon}_{D\tau}}{(1+\tau\rho_\tau)\,\tilde{\epsilon}_D} \tag{4}$$

Note that $\rho_\tau = d\log(p(1+\tau))/d\log(1+\tau)$ 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. Similar to other estimation approaches in the literature, our identification of θ_τ relies on consumers perceiving tax and price changes to be equally persistent, such that there is no difference in the degree of intertemporal substitution under full salience. Similarly, it requires that equivalent price and tax changes induce the same degree of substitution between product-modules.

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. We estimate $\widehat{\theta_\tau} = 0.575$ (s.e. 0.147) using the full sample and $\widehat{\theta_\tau} = 0.528$ (s.e. 0.130) using the border-pair subsample.[24] 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.

As a robustness test, we consider an alternative method for calculating the tax elasticity and the price elasticity, as well as the associated value of θ_τ, in Appendix Table OA.6. In Panels A and B, we report the effect of taxes and of the price instrument on total expenditures.[25] 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 $\widehat{\theta_\tau} = 0.552$ and $\widehat{\theta_\tau} = 0.491$ for the full sample and the border-pair subsample, respectively.

[24] Appendix Table OA.5 presents results based on alternative ways to account for spatial heterogeneity in consumption trends in our main sample. 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 slightly lower values of θ_τ, ranging between 0.376 and 0.507. 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.

[25] 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.

6 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 counterfactual 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 3. We assume constant marginal costs and a constant price elasticity of demand throughout this calibration exercise. Using our estimates of ρ_τ and θ_τ, along with the pass-through expression, we recover an estimate of $v_q/(J\epsilon_{ms}) = 0.041$ 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 $(p-mc)/p = 0.028$.[26] Our last intermediate step estimates $v_q/J = 0.034$.

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 in Table 2. Extending the incidence formula for ad valorem taxes to allow for heterogeneity in θ_τ results in the following[27]:

$$I = \frac{\rho_\tau(1+\tau) + (1-\theta_\tau)\tau\tilde{\epsilon}_{D\tau} + \tau(1+\tau)\tilde{\epsilon}_D(1/p)Var(\theta_\tau)}{\left(1-\frac{\nu_q}{J}\right)(1-\rho_\tau) + \frac{\nu_q}{J}\theta_\tau\left(1+\tau\rho_\tau\right)} \tag{5}$$

[26] 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/.

[27] While this is a slight abuse of notation, we omit the expectations operator when defining the mean inattention parameter to save on additional notation.

In Panel B of Table 3, column (1) assumes no heterogeneity in θ_τ, while column (2) illustrates sensitivity to heterogeneity in θ_τ by assuming that the variance of θ_τ is equal to 0.20.[28] In column (1), we calculate $I = 17.051$, 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 reduces the incidence on consumers to $I = 16.890$.

Turning to the marginal excess burden, we extend the ad valorem marginal excess burden formula to allow for heterogeneity in θ_τ and scale the expression so that it represents the change in welfare as a percentage of total revenue. This results in the following:

$$\frac{d\widetilde{W}}{d\tau} \equiv \frac{(1+\tau)}{pQ}\frac{dW}{d\tau} = \left(\frac{p-mc}{p} + \theta_\tau \tau\right)\tilde{\epsilon}_{D\tau} + \tau(1+\tau)\tilde{\epsilon}_D (1/p) Var(\theta_\tau) \tag{6}$$

Using the sample average tax rate of 3.6 percent for τ and assuming that $Var(\theta_\tau) = 0$, we find $d\widetilde{W}/d\tau = -0.033$ (column 1). This implies that the marginal excess burden is about 3.3 percent of total revenue. Allowing for heterogeneity in θ_τ across consumers increases the welfare cost of taxation to $d\widetilde{W}/d\tau = -0.042$ (column 2).

The final panel of Table 3 (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:

$$\frac{d\widetilde{W}^{Harberger}}{d\tau} \equiv \tau \tilde{\epsilon}_{D\tau} \tag{7}$$

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

[28] The empirical model assumed "constant effects" and thus ignored heterogeneity across consumers. The sensitivity analysis in column (2) can thus be interpreted as allowing for heterogeneity across consumers in tax salience (as accommodated by the theory), but continuing to assume that all consumers would respond similarly when fully optimizing (Assumption 1). To accommodate richer consumer heterogeneity in the empirical analysis requires more detailed individual-level data than the retail scanner data that we use in this paper. Also, in calibrating the incidence formula we assume a value for $(1/p)Var(\theta_\tau)$, which normalizes the consumer heterogeneity in θ_τ by price and avoids having to calibrate a value for the pre-tax price in any of our calibrations. To put the variance of θ_τ of 0.2 in context, consider the special case where $\theta_\tau \in \{0, 1\}$; i.e., some consumers fully optimize, while others are completely inattentive to taxes. Since θ_τ is binary, an average θ_τ of 0.575 (i.e., the share of consumers fully optimizing) implies a variance of 0.24. By comparison, Taubinsky and Rees-Jones (2018) report a lower bound estimate of the variance of θ_τ around 0.1. Since the variance of θ_τ enters the welfare and incidence formula linearly, using a lower value of the variance in our calibrations would naturally bring the incidence and marginal excess burden estimates reported in column (2) closer to the values in column (1) that assume no heterogeneity in tax salience across consumers.

of $\frac{d\widetilde{W}^{Harberger}}{d\tau} = -0.024$. To understand the sources of the difference between excess burden according to formula (6) (-0.042) and excess burden according to formula (7) (-0.024), 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:

$$\frac{d\widetilde{W}^{CLK}}{d\tau} \equiv \theta_\tau \tau \tilde{\epsilon}_{D\tau} \tag{8}$$

$$\frac{d\widetilde{W}^{TRJ}}{d\tau} \equiv \theta_\tau \tau \tilde{\epsilon}_{D\tau} + \tau(1+\tau)\tilde{\epsilon}_D(1/p)Var(\theta_\tau) \tag{9}$$

Assuming no heterogeneity gives an estimate of $\frac{d\widetilde{W}^{CLK}}{d\tau} = -0.014$ (column 1), which is smaller in magnitude than the excess burden according to the standard Harberger formula assuming full optimization (-0.024). Allowing for heterogeneity in θ gives an estimate of $\frac{d\widetilde{W}^{TRJ}}{d\tau} = -0.023$ (column 2), which is similar to the Harberger benchmark which ignores tax salience. 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):

$$\frac{d\widetilde{W}^{AH}}{d\tau} \equiv \left(\frac{p-mc}{p} + \tau\right)\tilde{\epsilon}_{D\tau} \tag{10}$$

Accounting for imperfect competition but assuming full optimization gives an estimate of $\frac{d\widetilde{W}^{AH}}{d\tau} = -0.043$. This is larger than the marginal excess burden estimate in column (1) and fairly similar to the estimate in column (2).

Overall, we find that tax salience accounts for a smaller percentage of the deviation between excess burden calculated using formula (6) and excess burden calculated using formula (7) in both columns.

Numerically, our findings in column (1) indicate that imperfect competition accounts for 219 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 -119 percent of the difference. In column (2) the corresponding percentages are 107 percent and -7 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. Appendix Table OA.7 reports all of the results in Table 3 using the county border pair subsample instead of the full sample of counties. Since the reduced-form effects are fairly similar, it is not surprising that the incidence and welfare results are broadly similar. The incidence continues to essentially all fall on consumers, and the excess burden is reduced somewhat (from 3.3 percent in column (1) and 4.2 percent in column (2) to 2.0 and 2.9 percent, respectively).

In Appendix Table OA.8, we show sensitivity to alternative values of the elasticity of inverse marginal surplus. For the main results in Table 3, 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.

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., $(1 - \theta_\tau)$.

6.1 Counterfactual Scenarios

The prior exercise shed 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 (see formula 4). Additionally, changes in either tax salience or market structure will change the pass-through of taxes into consumer prices, and we can use the analytical pass-through expression in Proposition 4 above to re-calculate the counterfactual pass-through of taxes into consumer prices in different scenarios.

Our main results are summarized in Table 4, 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

4 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 counterfactually set to zero (under perfect competition), or the conduct parameter is doubled (from the empirical estimate in Table 3) 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 3 allowing for salience effects and imperfect competition.[29] Column (2) holds constant the estimated 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 0.034 to 0.068 so that there is less competition in the market. This leads 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.[30] This reveals another interaction between tax salience

[29]For simplicity, we shut down heterogeneity in tax salience for these counterfactual exercises, and we report analogous counterfactual scenarios allowing for heterogeneity in tax salience in Online Appendix OA.9. The estimated pass-through and tax elasticity responses remain the same in each column (as compared to Table 4), which is a consequence of Assumption 1 that is used throughout this section.

[30]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.

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 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 4.7 percent of revenue to 9.7 percent across columns (3) and (6) under imperfect competition (going to full salience), compared to an increase of 1.5 percent to 4.3 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.

7 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 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 about 53 to 58 percent of the price elasticity. We use our formulas to calibrate a markup around 3 percent, which is consistent with grocery stores operating with fairly low profit margins.

We use these 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 suggests 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 (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.

4. 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.

5. 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.

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

7. 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.

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

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

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

11. 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.

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

13. Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer (2013). “Salience and Consumer Choice.” *Journal of Political Economy,* 121 (5): 803–43.

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. Caplin, Andrew, and Mark Dean (2015). “Revealed Preference, Rational Inattention, and Costly Infor-mation Acquisition,” *American Economic Review* 105 (7): 2183–2203.

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

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

19. 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.

20. 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.

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

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

23. 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.

24. 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.

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

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

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

28. 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.

29. 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.

30. Kőszegi, Botond, and Adam Szeidl (2013). "A Model of Focusing in Economic Choice," *Quarterly Journal of Economics* 128 (1): 53–104.

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

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

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

34. Nevo, Aviv (2001). "Measuring market power in the ready-to-eat cereal industry," *Econometrica*, 69: 307–42.

35. 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.

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

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

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

39. 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 Share	States taxing all food	States taxing module at reduced	States taxing module at full rate (but otherwise exempting
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 Expenditure Share	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,	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

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

Sample:	Full Sample	County Border Pair Subsample
	(1)	(2)
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]	0.961	0.986
	(0.045)	(0.016)
$d\log(Q)/d\log(1+\tau)$ [tax elasticity]	-0.668	-0.650
	(0.185)	(0.084)
Panel B: 2SLS Estimates of the Price Elasticity of Demand		
$d\log(Q)/d\log(p)$	-1.202	-1.223
	(0.027)	(0.027)
Panel C: "Plug-in" Estimate of Tax Salience Parameter		
θ_τ	0.575	0.528
	(0.147)	(0.130)
Specification:		
Store × Module fixed effects	y	y
Module × State × Year-Quarter fixed effects	y	
Module × Border Pair × Year-Quarter fixed effects		y
N	53,895,446	33,749,157

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 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 Retail Scanner data 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, where prices are instrumented with leave-self-out chain-level average prices. In Panel C, we report the estimate of the tax salience parameter. 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.

Table 3

Calibration of Incidence and Marginal Excess Burden Formulas

	(1)	(2)
Panel A: Inputs and Intermediate Estimates Needed in Calibration		
<u>Inputs:</u>		
Average tax rate, τ	0.036	
Price elasticity, $\tilde{\varepsilon}_D \equiv \partial\log(Q)/\partial\log(p)$	-1.202	
Tax pass-through, $\rho_\tau \equiv d\log(p(1+\tau))/d\log(1+\tau)$	0.961	
Tax elasticity, $\tilde{\varepsilon}_{D\tau} \equiv d\log(Q)/d\log(1+\tau)$	-0.668	
<u>Intermediate estimates:</u>		
Implied estimate of $v_q/(J\epsilon_{ms})$	0.041	
Implied markup $(p - mc)/p$	0.028	
Implied estimate of v_q/J	0.034	
($v_q/J = 0$ is perfect competition, $v_q/J = 1$ is perfect collusion)		
<u>Tax salience:</u>		
Tax salience parameter, θ_τ	0.575	
Heterogeneity in θ_τ, $(1/p)\mathrm{Var}(\theta_\tau)$	0.00	0.20
Panel B: Incidence and Marginal Excess Burden Formulas		
<u>Incidence (I)</u>		
$I \equiv (d\mathrm{CS}/d\tau)/(d\mathrm{PS}/d\tau)$	17.051	16.890
$= (\rho_\tau(1+\tau) + (1-\theta_\tau)\tau\tilde{\varepsilon}_{D\tau} - \tau(1+\tau)\tilde{\varepsilon}_D(1/p)\mathrm{Var}(\theta_\tau)) \,/\, ((1-v/J)(1-\rho_\tau)+(v/J)\theta_\tau(1+\tau\rho_\tau))$		
<u>Marginal Excess Burden ($d\tilde{W}/d\tau$)</u>		
$d\tilde{W}/d\tau = ((p-mc)/p + \theta_\tau\tau)\tilde{\varepsilon}_{D\tau} - \tau(1+\tau)\tilde{\varepsilon}_D(1/p)\mathrm{Var}(\theta_\tau)$	-0.033	-0.042
Panel C: Decomposition of the Deviation Between General Formula and Harberger Formula		
Harberger formula (assuming perfect competition and full salience), $d\tilde{W}/d\tau = \tau\tilde{\varepsilon}_{D\tau}$	-0.024	-0.024
Imperfect salience only, $d\tilde{W}/d\tau = \theta_\tau\tau\tilde{\varepsilon}_{D\tau} - \tau(1+\tau)\tilde{\varepsilon}_D(1/p)\mathrm{Var}(\theta_\tau)$	-0.014	-0.023
Decomposition of deviation (as % of difference b/w Harberger and general formula)	-119%	-7%
Imperfect competition only, $d\tilde{W}/d\tau = ((p-mc)/p + \tau)\tilde{\varepsilon}_{D\tau}$	-0.043	-0.043
Decomposition of deviation (as % of difference b/w Harberger and general formula)	219%	107%

<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 Table 2 column (1), 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 and a standard Harberger analysis. In column (1), we assume no heterogeneity in salience across consumers; in column (2) we allow for heterogeneity in salience parameter by calibrating the variance of θ_τ.

Table 4

Counterfactual Scenarios Adjusting Tax Salience and Market Structure

	Imperfect salience, $\theta_\tau = 0.575$	Counterfactual scenarios				
Tax salience parameter, θ_τ:		Imperfect salience, $\theta_\tau = 0.575$		Full salience, $\theta_\tau = 1.00$		
Market structure (conduct parameter, v_q/J):	Imperfect competition, $v_q/J = 0.034$	Perfect competition, $v_q/J = 0$	Imperfect competition, $v_q/J = 0.068$	Imperfect competition, $v_q/J = 0.034$	Perfect competition, $v_q/J = 0$	Imperfect competition, $v_q/J = 0.068$
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Inputs and Intermediate Estimates Needed to Calibrate Formulas						
Inputs: [constant parameters]						
Average tax rate, τ			0.036			
Price Elasticity, $\tilde{\varepsilon}_D \equiv \partial\log(Q)/\partial\log(p)$			-1.202			
Intermediate estimates:						
Implied estimate of $v_q/(J\epsilon_{ms})$	0.041	0.000	0.082	0.041	0.000	0.082
Implied markup $(p - mc)/p$	0.028	0.000	0.057	0.028	0.000	0.057
Panel B: Counterfactual Tax Responses and Implied Incidence and Marginal Excess Burden						
Counterfactual responses:	Baseline estimates	Counterfactual responses				
Tax Pass-Through, $\rho_\tau \equiv d\log(p(1+\tau))/d\log(1+\tau)$	0.961	1.000	0.924	0.933	1.000	0.869
Tax Elasticity, $\tilde{\varepsilon}_{D\tau} \equiv d\log(Q)/d\log(1+\tau)$ $= \tilde{\varepsilon}_D(\rho_\tau - 1 + \theta_\tau/(1+\tau\theta_\tau))(1+\tau)$	-0.668	-0.701	-0.607	-1.118	-1.202	-1.039
Incidence (I)						
$I \equiv (d\text{CS}/d\tau)/(d\text{PS}/d\tau)$ $= (\rho_\tau(1+\tau)+(1-\theta_\tau)\tau\tilde{\varepsilon}_{D\tau})/((1-v/J)(1-\rho_\tau)+(v/J)\theta_\tau(1+\tau\rho_\tau))$	17.051	∞	8.536	9.640	∞	4.674
Marginal Excess Burden ($d\tilde{W}/d\tau$)						
$d\tilde{W}/d\tau = ((p-mc)/p + \theta_\tau\tau)\tilde{\varepsilon}_{D\tau}$	-0.033	-0.015	-0.047	-0.072	-0.043	-0.097

Notes: This table reports calibrations of the tax incidence and marginal excess burden formulas for different counterfactual assumptions about tax salience and market structure. All columns assume no heterogeneity in salience parameter for simplicity. 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 2). 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 3, 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 3.