Banco de México

Working Papers

N° 2023-20

Efficiency and Incidence of Taxation with Free Entry and Love-of-Variety

Kory Kroft
University of Toronto and NBER

Jean-William Laliberté
University of Calgary

René Leal Vizcaíno
El Colegio de México

Matthew J. Notowidigdo
University of Chicago and NBER

December 2023

Documento de Investigación
2023-20

Working Paper
2023-20

Efficiency and Incidence of Taxation with Free Entry and Love-of-Variety*

Kory Kroft†
University of Toronto and NBER

Jean-William Laliberté‡
University of Calgary

René Leal Vizcaíno§
El Colegio de México

Matthew J. Notowidigdo**
University of Chicago and NBER

Abstract: We develop a theory of commodity taxation featuring imperfect competition along with love-of-variety preferences and endogenous firm entry and exit, and we derive new formulas for the efficiency and pass-through of specific and ad valorem taxes. These formulas unify existing canonical ones and feature a new term capturing the effect of variety on consumer surplus. Intuitively, if taxes reduce product varieties in the market, then the impact on social welfare depends on how much consumers value variety. As a proof-of-concept, we use the theoretical formulas to identify love-of-variety preferences in an empirical application. Our welfare analysis shows that the marginal excess burden of taxation is very sensitive to the estimated love-of-variety, which can overturn classical results on the desirability of ad valorem versus specific taxation.
Keywords: Public Economics; Taxation; Tax Incidence; Sales Tax
JEL Classification: H20, H22, H71

Resumen: Se desarrolla una teoría de impuestos al consumo con competencia imperfecta, entrada y salida de empresas endógena y con preferencias por mayor variedad de bienes. Se derivan nuevas fórmulas de eficiencia y traspaso para impuestos específicos y ad valorem. Estas fórmulas generalizan las canónicas añadiendo un nuevo término que captura el efecto de las preferencias por variedad en el excedente del consumidor. Intuitivamente, si los impuestos reducen la variedad de productos, el impacto en bienestar dependerá de las preferencias por variedad. Como prueba de concepto, a partir de las fórmulas teóricas se implementa un modelo para identificar empíricamente las preferencias por variedad. El análisis de bienestar muestra que el exceso marginal de la carga impositiva es muy sensible a las preferencias por variedad, lo cual puede revertir el resultado clásico de preferir impuestos ad valorem sobre impuestos específicos.
Palabras Clave: Economía pública; Impuestos; Incidencia fiscal; Impuestos sobre ventas

*We acknowledge funding from the Social Sciences and Humanities Research Council (SSHRC). This research used data from The Nielsen Company (US). 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, Nielsen or Banco de México. This paper was written while René Leal was working at Banco de México.

† Department of Economics. Email: kory.kroft@utoronto.ca.

‡ Department of Economics. Email: jean-william.lalibert@ucalgary.ca.

§ Centro de Estudios Económicos. Email: rleal@colmex.mx.

** Booth School of Business. Email: noto@chicagobooth.edu.

1 Introduction

This paper develops a unifying theory of commodity taxation featuring imperfect competition along with love-of-variety preferences and endogenous firm entry and exit. Our framework encompasses a wide range of market conduct – including both quantity and price competition. We derive new formulas for the marginal excess burden and pass-through of specific or excise taxes (fixed dollar amount) and ad valorem taxes (percentage of price), and as a proof-of-concept we implement the formulas in an empirical application.

Our theoretical tax formulas connect the efficiency cost and pass-through of a tax, in the presence of firm entry and exit, with sufficient statistics, as in Chetty (2009) (efficiency) and Weyl and Fabinger (2013) (pass-through). These sufficient statistics include classical ones emphasized in the tax literature – the elasticity of market demand, the elasticity of the firm's marginal cost, the curvature of the firm's own demand, and the market conduct parameter which characterizes the degree of competition (zero under perfect competition and one under monopoly) –and a new one that stems from consumers' love-of-variety preferences.[1]

The value of our framework is three-fold. First, we see our tax formulas as useful pedagogical tools since they unify existing theoretical results in a single framework and identify the key concepts that appear in more specialized settings.[2] Second, we use our framework to show how to identify and estimate the sufficient statistics in a theoretically consistent way. Lastly, the framework delivers several new insights into some key results in public economics on the economics of taxation with imperfect competition and free entry. In particular, it highlights the importance of modelling and identifying consumers' love-of-variety for policy.

First, while it is well known in homogeneous product models that the marginal excess burden depends on the sensitivity of demand and producer prices in response to the tax (see, for example, Besley 1989 and Delipalla and Keen 1992), we show that when consumers value product variety, the canonical tax formula with free entry is modified (for both specific and ad valorem taxation): it depends additionally on the effect of taxes on variety scaled by the effect of variety on consumer surplus – hereafter referred to as the "variety effect". Intuitively, when firms decide to enter (or exit) the market in response to a change in taxes, they do not internalize the positive effects of product creation (or product destruction) on consumer surplus; they only internalize the effects of the tax change on profits, which generates an externality. We illustrate the connection between the welfare effects of a tax with free entry

[1] The tax formulas for specific and ad valorem taxes depend on the same sufficient statistics. This is very convenient in empirical settings where researchers only observe one type of tax but are interested in comparing the efficiency costs and incidence between the alternative (counterfactual) tax structures. For example, we only observe ad valorem taxation in our empirical setting, and we use this insight to empirically assess whether ad valorem or specific taxes are more desirable on efficiency grounds.

[2] As Keen (1998) remarked (and which remains relevant today): "The models of imperfect competition used in the [taxation] literature are ... special cases ... [T]here remains much to be done – for example, exploring richer models ... and examining empirically the impact of tax structure on product quality and variety."

and the welfare effects of product entry (Mankiw and Whinston 1986).[3] In particular, we show that if the "business-stealing" effect, which arises since firms do not internalize the effect of entry on other firms' profits, dominates the variety effect, then a ceteris paribus tax increase that leads to fewer product varieties increases welfare.

Second, we extend the pass-through formulas in Weyl and Fabinger (2013) (specific taxes) and Adachi and Fabinger (2017) and Kroft et al. (2020*a*) (specific and ad valorem taxes) to the case where there is free entry and consumers have love-of-variety preferences. In a long-run equilibrium, prices are determined jointly by firms' first-order condition and a free-entry condition. An increase in taxes lowers firm profits and leads to exit. This puts upward pressure on prices and the resulting change in prices depends on the strength of love-of-variety preferences. We show that under a standard regularity condition, a greater love-of-variety raises the pass-through of both types of taxes. When consumers value variety, a reduction in the number of varieties leads to a larger loss in profits, which in turn leads to more exit, and thus a higher price.

Third, we compare the welfare effects and incidence of specific versus ad valorem taxation which follows a long tradition in public economics. The classical result in the literature is that when products are homogeneous, ad valorem taxes are more efficient and lead to lower pass-through than specific taxes (Keen 1998). We show that if the variety effect is sufficiently strong, both of these results can be overturned. In particular, if consumers have a strong enough preference for variety, then consumer prices can be higher and welfare can be lower under ad valorem taxation as compared to specific taxation. Additionally, if consumer prices are higher under ad valorem taxation, then welfare is higher under specific taxation, but the converse need not be true.

We illustrate the usefulness of each of these new results through an empirical application studying grocery stores selling consumer products in the United States. We combine Nielsen Retail Scanner data with detailed product-level and county-level sales tax data and use a cross-sectional "county border pair" research design (Holmes 1998, Dube, Lester and Reich 2010). This design exploits sales tax rate differentials between taxed and tax-exempt products across nearby stores located in contiguous counties across state borders. Guided by our theoretical formulas, we focus on identifying the causal effects of sales taxes on consumer prices, quantity demanded, and product variety. We find that sales taxes are slightly overshifted onto consumer prices, have a large effect on market demand, and meaningfully reduce the variety of products available to consumers, with the magnitude of the variety response being about one-third of the magnitude of the overall effect of sales taxes on market demand.

We show how to use these reduced-form estimates, along with our new formulas for ad

[3]For papers that examine the efficient provision of product variety, see Spence (1976), Dixit and Stiglitz 1977, Vives (2001), Parenti, Ushchev and Thisse (2017) and Dhingra and Morrow (2019).

valorem taxes, to identify the market conduct parameter, the variety effect, and the business-stealing effect. We estimate a conduct parameter of 0.092, where 0 is perfect competition and 1 is perfect collusion. This suggests a high degree of competition, as would be expected in our setting of retail grocery stores. Second, we estimate the variety effect to be 0.125. This can be given a willingness-to-pay interpretation: an exogenous 10 percent reduction in variety reduces average willingness-to-pay by 1.25 percent.[4] Third, we estimate a business-stealing effect of -0.060. Since the estimated business-stealing effect is smaller in magnitude than the variety effect, our results imply that the privately optimal product variety is insufficient relative to the socially optimal product variety.

Finally, we use these model-based parameter estimates to calibrate our new tax formulas. We estimate the marginal excess burden of ad valorem taxes to be about 8 percent of total firm revenues, which is about 60 percent larger than an alternative benchmark that accounts for imperfect competition but ignores consumers' love-of-variety preferences, and several times larger than a standard Harberger benchmark. These results show that love-of-variety preferences are an economically significant factor in determining the efficiency costs of sales taxes, and our theoretical formulas provide the economic intuition for these results: when consumers exhibit love-of-variety preferences and product variety is insufficient relative to the social optimum, then there is an additional welfare consequence of tax changes when taxes affect product variety.

Relationship to the Literature –

Our main theoretical contribution is to build on the papers which consider the welfare effects and incidence of taxation (specific taxation and ad valorem taxation) in the long run when entry and exit are allowed. Besley (1989) considers the welfare effects of specific taxation with free entry in a model of Cournot quantity competition with homogeneous products. Delipalla and Keen (1992) consider a homogeneous product conjectural variations model and derive both incidence and welfare results for specific taxation and ad valorem taxation in the short run with a fixed number of firms and in the long run with free entry. However, these papers do not take into account the value to consumers from greater product variety when products are differentiated.

Anderson, De Palma and Kreider (2001*b*) demonstrate – through the use of counterexamples featuring Bertrand competition and differentiated products – the possibility that specific taxation can dominate ad valorem taxation with free entry and exit.[5] To focus on the dis-

[4] One way to benchmark this magnitude is to note that it is much smaller than what would be implied by CES/Logit preferences, where the variety effect is pinned down by the inverse of the price elasticity of demand. This would imply an estimate of the variety effect of around 0.855.

[5] Kay and Keen (1983) consider a model of monopolistic competition with Dixit-Stiglitz preferences and fixed aggregate demand and show that specific taxation can dominate ad valorem taxation as the taste for variety becomes infinitely strong.

tortion to product variety, they assume that market demand is completely inelastic to taxes and thus abstract from the "static" distortionary effect of taxation. Our theoretical analysis allows taxes to affect both output and variety in the market.[6]

Our empirical analysis builds on the literature estimating the reduced-form effects of sales taxes on prices and market demand (see Barzel 1976, Johnson 1978, Delipalla 1995, Poterba 1996, Besley and Rosen 1999, Delipalla and O'Donnell 2001, Chetty, Looney and Kroft 2009, Einav et al. 2014*a*, Baker, Johnson and Kueng 2021, and Kroft et al. 2020*a*). Our main innovation in this literature is to empirically estimate the effect of taxes on product variety for a broad range of products. Our empirical finding that sales taxes reduce product variety is consistent with Cawley et al. (2020), which finds that a beverage tax in Philadelphia resulted in fewer taxed products available in stores.

Lastly, our paper can be viewed more broadly as providing a new model-based approach for identifying and estimating consumers' love-of-variety preferences. Our approach uses a "parallel demands" assumption to establish identification of the variety effect, which complements existing approaches in Industrial Organization (Trajtenberg 1990, Petrin 2002, Bajari and Benkard 2003, Ackerberg and Rysman 2005, Berry and Pakes 2007), Macroeconomics (Benassy 1996), and International Trade (Alessandria and Choi 2007, Arkolakis, Costinot and Rodríguez-Clare 2012).[7]

The rest of the paper is organized as follows. Section 2 characterizes the objectives of consumers and firms. In Section 3, we derive formulas for the marginal excess burden and pass-through of specific taxation (Proposition 1) and ad valorem taxation (Proposition 2). We also derive a desirability condition and a comparison of pass-through rates of specific taxes and ad valorem taxes (Proposition 3). In Section 4, we use our tax formulas to organize our empirical analysis, which focuses on identification and estimation of the causal effects of ad valorem taxation on consumer prices, quantity demanded, and product variety. In Section 5, we recover the model parameters using our reduced-form estimates, and we use them to calibrate our new formulas for ad valorem taxes and consider several economic applications of the calibration results. Section 6 concludes.

[6]Hamilton (2009) considers a specific model of retailer competition and emphasizes the role of *multi-product* firms demonstrating that one can get overshifting of specific taxes even with log-concave demand which is ruled out in most models with single-product firms. Hamilton also demonstrates that with multi-product firms, specific taxation can dominate ad valorem taxation with free entry. In our setting, we show that specific taxation can dominate ad valorem taxation even with single-product firms depending on the strength of consumers' love-of-variety.

[7]The "parallel demands" assumption requires that the inverse aggregate demand curve shifts *in parallel* when there is an exogenous change in the number of product varieties in the market.

2 The Model

We consider a differentiated product market (the “inside market”) which is subject to a specific tax t and ad valorem tax τ on consumers that apply to each product in the market. We adopt a “conjectural variations” approach to modeling the inside market. This approach permits firms to form beliefs about how the other firms in the market respond strategically, taking the form of competition in the inside market as exogenous. We assume that markets for other goods are perfectly competitive and are not subject to taxation, implying that taxes in the inside market have no indirect welfare effects on other markets in the economy. Throughout our analysis, we assume that product quality is fixed with respect to taxation and thus abstract from the “upgrading effects” of specific taxation (Keen 1998). Important papers that consider endogenous product quality are Cremer and Thisse (1994), Delipalla and Keen (2006), and Gillitzer, Kleven and Slemrod (2017). We also abstract from uncertainty in prices and thus the tax base (Kay and Keen 1983) and externalities in consumption and production (Pirttilä 1997).

Consumers

Following Auerbach and Hines (2001), we abstract from population heterogeneity and consider a single representative individual with exogenous income Z.[8] Preferences when there are J varieties available are given by the quasilinear utility function $u^J(q_1, \ldots, q_J) + y$, where q_j is the quantity consumed of variety $j = 1, \ldots, J$ and $y \in \mathbb{R}$ is the numeraire (representing consumption in the outside market). . We assume that the subutility function, u^J, 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 post-tax or consumer price is given by $p_j\,(1+\tau)+t$ for all $j = 1, ..., J$. We define $u(Q, J) \equiv u^J(Q/J, \ldots, Q/J)$ to be the compact notation of utility for the symmetric case where the individual consumes $q = \frac{Q}{J}$ units of each variety $j = 1, \ldots, J$. Furthermore, we assume that $u(Q, J)$ is concave in J, so that variety has diminishing returns, which ensures that the planner’s problem defined below is well-behaved.[9]

[8] As Keen (1998) argues, this is a feature of much of the economic literature in this area, since it is generally thought that the distributional effects of commodity taxes are determined by their levels and not the overall structure of the tax.

[9] We work with the representative consumer with love-of-variety embedded in a utility function over quantity and variety ($u(Q, J)$) for tractability, but it is well-known that this kind of representative agent model can be given a discrete choice model microfoundation (see, e.g., Anderson et al. 1987 which shows how a specific discrete model aggregates to representative-agent CES model with love-of-variety preferences similar to our model here). Intuitively, an individual consumer chooses their most preferred variety in the discrete choice model, and an increase in product variety increases the chance that some consumers choose a variety that they like even more, which raises total consumer surplus even if the total quantity demanded does not change. Similarly, a reduction in variety will lead some consumers to switch to a less-preferred variety, which reduces their utility.

Consumer demand for product variety j is given by $q_j = q_j(p_1, ..., p_J, \tau, t)$ which is a function of both prices and taxes (τ, t). We allow for salience effects by considering the possibility that $q_j(p_1, ..., p_J, \tau, t) \neq q_j(p_1(1+\tau)+t, ..., p_J(1+\tau)+t, 0, 0)$.[10] In what follows, we assume that the "observed" demand function $q_j(\cdot)$ is symmetric and twice differentiable and denote by $q^J(p, \tau, t)$ demand corresponding to symmetric prices and J firms: $q^J(p, \tau, t) \equiv q_j(p, ..., p, \tau, t)$. We define market demand as $Q(p, \tau, t, J) = Jq^J(p, \tau, t)$ and the inverse market demand $wtp(Q, J) \equiv P(Q, J) \equiv Q^{-1}(p, 0, 0, J)$ which corresponds to willingness-to-pay when taxes are fully salient. For fully salient taxes, market demand is $Q(p\,(1+\tau)+t, 0, 0, J)$ and inverse market demand is $P(Q, J)$. For non-salient taxes, we define the degree of inattention to ad valorem and specific taxation respectively as $\theta_\tau \equiv \frac{\frac{\partial Q(p,0,0,J)}{\partial \tau}}{\frac{\partial Q(p,0,0,J)}{\partial p}} \times \frac{1}{p}$ and $\theta_t \equiv \frac{\frac{\partial Q(p,0,0,J)}{\partial t}}{\frac{\partial Q(p,0,0,J)}{\partial p}}$, the ratio of the demand responses to the tax and price starting from $\tau = t = 0$.

We assume that θ_τ and θ_t are constant over (p, τ, t, J) so that $Q(p, \tau, t, J)$ satisfies $wtp(Q, J) = p + p\theta_\tau \tau + \theta_t t$. In other words, consumer willingness-to-pay is equal to the producer price plus the "perceived" tax liability. We also make use of the following definitions. First, $mwtp(Q) = \frac{\partial wtp}{\partial Q}(Q, J)$ is the marginal willingness-to-pay. Second, we define $\epsilon_{ms} \equiv \frac{ms(Q)}{ms'(Q)Q}$ as the elasticity of inverse marginal surplus where $ms(Q) \equiv -mwtp(Q)Q$ is marginal consumer surplus. This elasticity measures the curvature of the market demand curve and is central to pass-through when there is imperfect competition in the product market (see Weyl and Fabinger 2013). Next, the price elasticity of demand with fully salient taxes is given by $\epsilon_D \equiv -\frac{p(1+\tau)+t}{Qmwtp(Q)}$, and we define the price elasticity of demand with perceived taxes as $\epsilon_D^* \equiv \frac{p(1+\theta_\tau \tau)}{p(1+\tau)+t}\epsilon_D$. Finally, we define the "variety effect" as the effect of a change in variety on consumer surplus (holding prices and output constant). In doing so, we treat the number of firms as a continuous variable, a standard assumption in this literature following Seade (1980), Besley (1989) and Delipalla and Keen (1992).

Definition 1. Let Λ be the variety effect which captures the effect of a change in varieties J on consumer surplus $CS = u(Q, J) - (p\,(1+\tau)+t)Q$, keeping p and Q fixed. In this case, since $P(Q, J)$ is the inverse of $Q(p, 0, 0, J)$ and $CS(Q, p, \tau, t, J) = u(Q, J) - (p\,(1+\tau)+t)Q = \int_0^Q P(s, J)ds - (p\,(1+\tau)+t)Q$ then

$$\Lambda(Q, J) \equiv \frac{\partial CS}{\partial J} = \int_0^Q \frac{\partial P(s, J)}{\partial J} ds$$

[10] Throughout, we implicitly assume that (1) taxes affect utility only through their effects on the chosen consumption bundle and (2) when tax-inclusive prices are fully salient, the agent chooses the same allocation as a fully-optimizing agent. These are Assumptions $A1$ and $A2$ in Chetty, Looney and Kroft (2009). Although salience effects are not the focus of this paper, we permit them to be able to empirically implement our tax formulas. Accounting for salience effects does not deliver any novel insights beyond what is already emphasized in Kroft et al. (2020*a*). In Kroft et al. (2020*a*), our focus is primarily on understanding how tax salience and imperfect competition interact in tax formulas *when there are a fixed number of firms.* This paper, by contrast, focuses on love-of-variety preferences and imperfect competition *when the number of firms is endogenous.*

Graphically, the variety effect can be represented as the vertical distance between inverse aggregate demand curves before and after a change in the number of varieties (Kroft et al. 2020*c*).

Firms

On the supply side, there is an infinite pool of identical potential entrants. Each firm has the cost function $c_j(q_j) = c(q_j) + F$, where $c(\cdot)$ is the variable cost of production, which is increasing and twice differentiable with $c(0) = 0$, and $F > 0$ is the fixed cost of production.[11] Each firm makes two decisions. First, each firm decides whether to produce given the fixed cost F. Second, each firm chooses p_j to maximize profits π_j:

$$\max_{p_j} \pi_j = p_j q_j(p_1 \dots, p_J, \tau, t) - c(q_j(p_1 \dots, p_J, \tau, t)) - F$$
$$\text{s.t. } \frac{\partial p_k}{\partial p_j} = \nu_p \text{ for } k \neq j$$

The term ν_p is the conjectural price variation of the other firms' price as a function of p_j. This is firm j's belief – or "conjecture" – of how other firms will react to a price change. Different forms of competition make different assumptions on these beliefs, as we explain in more detail below. The first-order condition for p_j is given by:

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

where $mc(q) \equiv c'(q)$, and we will make use of the following definition $\epsilon_S \equiv \frac{c'(q)}{c''(q)q}$ which reflects the shape of the firm's marginal cost function. In a symmetric equilibrium, $p_j = p$ solves:

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

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 $p(\tau, t)$ exists.[12] By letting $\nu_q \equiv \frac{1}{mwtp(Q)} \times \frac{1}{\frac{dq_j}{dp_j}} = \frac{1}{mwtp(Q)} \times$

[11] In our empirical application, we will measure the number of different products sold at retail grocery stores. The retail stores choose varieties produced by different manufacturers, so the fixed cost should not be thought of as the cost to manufacturers of developing new products to be marketed and sold nationally. Those costs are sunk from the perspective of the model. Instead, the fixed cost in the model should be thought of as the cost to retailers of allocating some of the space in the store for additional varieties in one category of products instead of another category of products – that is, does the grocery store want to sell more varieties of cookies (untaxed) or batteries (taxed) given the current relative tax rates on the two types of products?

[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 $p(\tau, t)$, the same results follow.

$\frac{1}{\frac{\partial q_j}{\partial p_j}+\nu_p \sum_{k\neq j}\frac{\partial q_j}{\partial p_k}}$ we can rewrite the first-order condition as a generalized Lerner index:[13]

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

Setting $\nu_q = J$ yields the monopoly (perfect collusion) outcome, and setting $\nu_q = 0$ gives the perfect competition (marginal cost pricing) solution. Setting $\nu_q = 1$ corresponds to Cournot competition when goods are homogeneous, and setting $\nu_p = 0$ yields the Bertrand-Nash equilibrium. The model thus captures a wide range of market conduct.[14] We will assume throughout this section that the conduct parameter ν_q is constant, which implies $\frac{d\nu_q}{dt} = 0$. This assumption rules out some pricing models such as discrete choice-based models of differentiated products Nash-in-prices competition (see, e.g., Weyl and Fabinger 2013).[15]

In the "long run", the number of firms $J(\tau, t)$ in the symmetric equilibrium is determined by the free-entry condition $\pi_j(p(J, \tau, t), J, \tau, t) = 0$:

$$p(J(\tau,t),\tau,t)q(p(J(\tau,t),\tau,t),\tau,t) - c(q(p(J(\tau,t),\tau,t),\tau,t)) - F = 0 \tag{2}$$

Moreover, we impose assumptions that ensure uniqueness for the firm pricing decision and entry problem, so that there is a unique solution $p(\tau,t)$ and $J(\tau,t)$. In particular, we assume that $\pi_j(p_j, p_{-j}, J, \tau, t)$ is concave in p_j and decreasing in J. We also define $\Delta \equiv \left[2 - \frac{\nu_q}{J} + \frac{\epsilon_D^* - \frac{\nu_q}{J}}{\epsilon_S \frac{\nu_q}{J}} + \frac{\frac{\nu_q}{J}}{\epsilon_{ms}}\right] - \left(1 + \frac{\epsilon_D^* - \frac{\nu_q}{J}}{\epsilon_S \frac{\nu_q}{J}} + \frac{1}{\epsilon_{ms}}\right)\epsilon_D \frac{J}{p(1+\tau)+t}\frac{\partial P}{\partial J} + \left(1 - \frac{\nu_q}{J}\right)\epsilon_D \frac{JQ}{p(1+\tau)+t}\frac{\partial^2 P}{\partial J \partial Q}$, which we assume is always greater than 0 in order for the entry decision to be unique. We show in the Appendix (after the proof of Proposition 2) that $\Delta = -\frac{J\epsilon_D^*}{pq}\frac{\partial \pi}{\partial J}$ so the assumption that $\Delta > 0$ is equivalent to $\frac{\partial \pi}{\partial J} < 0$.[16] Lastly, we define the long-run demand as $Q_L(\tau,t) = Q(p(\tau,t),\tau,t,J(\tau,t))$.[17] The effect of taxes on long-run demand, taking into account the en-

[13]Equivalently $\frac{p-mc(q)}{p} = \frac{\frac{\nu_q}{J}}{\epsilon_D^*}$.

[14]The conjectural variation term is a reduced-form version of a Nash equilibrium only when it corresponds to static solution concepts (e.g. $\nu_p = 0$ or $\nu_q = 1, J$) or are reduced-forms of truly dynamic models (Vives 2001, Riordan 1985) or supply function equilibria (Hart 1982). We do not take a stand on the dynamic model that ν_q captures in reduced-form, instead proving that our evaluation of welfare is robust to any of the specifications that can be modeled this way.

[15]Fully generalizing our results to allow for the conduct parameter to depend on taxation is outside the scope of this paper, but this is potentially important given the widespread use of empirical models featuring discrete choice-based models of differentiated products Nash-in-prices competition. We conjecture that all of our main results in Propositions 1 and 2 are likely to go through as first-order approximations when considering small changes in taxes that do not have first-order effects on the conduct parameter. We also note that the conduct parameter appears in the incidence and entry formulas but not the efficiency formula, suggesting that this assumption may not be necessary to characterize the welfare effects of taxes with endogenous entry and love-of-variety preferences.

[16]Formally, the two stability conditions for the firms' problem are $1 + \frac{\epsilon_D^* - \frac{\nu_q}{J}}{\epsilon_S} + \frac{\frac{\nu_q}{J}}{\epsilon_{ms}} > 0$ and $\Delta > 0$, and we evaluate both of these conditions in our calibrations below.

[17]Here we use the assumption that for each set of taxes t and τ there is a unique symmetric price equilibrium $p(\tau,t)$ where $J(\tau,t)$ firms enter the market. The notation Q_L serves to mathematically differentiate the

dogeneity of variety to taxes, is given by $\frac{dQ_L}{dt}\Big|_\tau = \frac{\partial Q}{\partial t} + \frac{\partial Q}{\partial J}\frac{dJ}{dt}\Big|_\tau + \frac{\partial Q}{\partial p}\frac{dp}{dt}\Big|_\tau$ for the specific tax (when keeping the ad valorem tax constant), and similarly $\frac{dQ_L}{d\tau}\Big|_t = \frac{\partial Q}{\partial \tau} + \frac{\partial Q}{\partial J}\frac{dJ}{d\tau}\Big|_t + \frac{\partial Q}{\partial p}\frac{dp}{d\tau}\Big|_t$ for the ad-valorem tax (when keeping the specific tax constant).

The next lemma uses the free-entry condition in order to connect the markup to the reduced-form effects of taxation. We will use this lemma in the next section when deriving the welfare and pass-through effects of taxation.

Lemma 1. *In the long run (when the free-entry condition* (2) *is satisfied), for any tax rate t or τ, the following envelope condition holds:*

$$\frac{\varepsilon_{p,t}}{\varepsilon_{q,t}} = -\frac{p - c'(q)}{p} \tag{3}$$

$$\frac{\varepsilon_{p,\tau}}{\varepsilon_{q,\tau}} = -\frac{p - c'(q)}{p} \tag{4}$$

where $\varepsilon_{p,t} \equiv \frac{t}{p}\frac{dp}{dt}$ and $\varepsilon_{q,t} \equiv \frac{t}{q}\frac{dq}{dt}$ are the long-run elasticities of producer prices and firm-level output with respect to specific taxes, respectively, and $\varepsilon_{p,\tau} \equiv \frac{\tau}{p}\frac{dp}{d\tau}$ and $\varepsilon_{q,\tau} \equiv \frac{\tau}{q}\frac{dq}{d\tau}$ are the long-run elasticities of producer prices and firm-level output with respect to ad valorem taxes, respectively.

Proof. See Appendix for all proofs. □

This condition follows by totally differentiating the zero-profit condition in equation (2) with respect to the tax, $\frac{d\pi}{dt} = 0$ and $\frac{d\pi}{d\tau} = 0$. In economic terms, this condition requires that entry is such that after the tax change, the zero-profit condition continues to hold.

3 Welfare, Pass-through and Entry Effects of Commodity Taxes

It is well known that under perfect competition, specific taxation and ad valorem taxation are equivalent. However, this equivalence breaks down with imperfect competition. Our baseline analysis, summarized in Proposition 1, considers the marginal welfare gain, pass-through, and entry effects associated with a small increase in the specific tax t which applies to all goods in the inside market. We also consider comparative statics associated with changes in the ad valorem tax τ in Proposition 2. Finally, we compare both types of taxation under imperfect competition in Proposition 3. We first present results for the general case and then consider several specialized cases, in order to connect our results with the existing literature on taxation under imperfect competition.

functions $Q(\cdot,\cdot,\cdot,\cdot) : \mathbb{R}^4 \to \mathbb{R}$ and $Q_L(\tau, t) : (\tau, t) \to Q(p(\tau, t), \tau, t, J(\tau, t))$.

We assume throughout that tax revenue $R = (p\tau + t)\,Q$ and profits $J\pi$ are redistributed to the representative consumer as a lump-sum transfer. We assume that the consumer treats profits and tax revenue as fixed when choosing consumption, failing to consider the external effects of 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. Total welfare, W, is given by the sum of consumer surplus (CS), profits ($J\pi$), and government tax revenues (R):

$$W(p,t,\tau,J) = \underbrace{u(Q,J) - (p\,(1+\tau)+t)Q}_{CS} + \underbrace{pQ - Jc\left(\frac{Q}{J}\right) - JF}_{J\pi} + \underbrace{(p\tau+t)\,Q}_{R} \tag{5}$$

We assume that the pre-existing tax τ_0 is constant, and we consider a small increase in the specific tax starting from t_0. A first-order approximation to the marginal excess burden of specific taxation is given by the following formula:

$$\frac{dW(p(\tau,t),\tau,t,J(\tau,t))}{dt} = \underbrace{(p_0(1+\theta_\tau\tau_0)+\theta_t t_0 - c'(q_0))\frac{dQ_L}{dt}}_{\text{Quantity effect}} + \underbrace{(\Lambda_0 + \pi_0 - [p_0 - c'(q_0)]q_0)\frac{dJ}{dt}}_{\text{Diversity effect}} \tag{6}$$

where $p_0, q_0, Q_0, J_0, \pi_0, \Lambda_0$ are all variables evaluated at the equilibrium corresponding to t_0 and τ_0.

Equation (6) shows that the marginal excess burden of specific taxation is a combination of two terms. The first term represents the standard distortionary effect of taxation on output. Intuitively, the social marginal value of output is given by the difference between willingness-to-pay $p_0(1+\theta_\tau\tau_0)+\theta_t t_0$ and the social marginal cost $c'(q_0)$. With no pre-existing taxes ($t_0 = \tau_0 = 0$) or when $\theta_t = \theta_\tau = 0$, the first term depends only on the markup which represents the distortionary "wedge" in output due to the presence of market power. The second term represents the distortion to product variety. To see the intuition for this expression, consider the case of constant marginal cost. The second term then becomes $(\Lambda_0 - F)\,\frac{dJ}{dt}$. Thus, whether the change in variety induced by taxes lowers (or increases) welfare along this margin depends on whether the love-of-variety exceeds (or is less than) the fixed cost. The term $\Lambda_0 - F$ thus represents the distortionary wedge on the entry margin due to free entry. This discussion shows that welfare is maximized when there are no wedges in the economy either due to taxation, market power, or free entry. As we show in the Appendix, this occurs when $t_0 = \tau_0 = 0$, $p_0 = c'(q^*)$ and $\Lambda_0 = -\pi_0$.

Equation (6) nests canonical formulas in public economics for homogeneous or differentiated products, perfect or imperfect competition, fixed or endogenous variety, and fully optimizing or behavioral agents. First, in the case of perfect competition where $p = c'(q)$, fully optimizing agents ($\theta_t = \theta_\tau = 1$), and J is fixed, equation (6) reduces to $\frac{dW}{dt} = (t_0 + p_0\tau_0)\frac{dQ_L}{dt}$

which corresponds to the classic Harberger (1964) formula. Next, Auerbach and Hines (2001) consider a model of homogeneous products ($\Lambda_0 = 0$) with imperfect competition and fixed variety ($\frac{dJ}{dt} = 0$) and fully optimizing consumers ($\theta_t = \theta_\tau = 1$). In this case, the marginal excess burden of taxation is given by $\frac{dW}{dt} = (p_0(1+\tau_0) + t_0 - c'(q_0))\frac{dQ_L}{dt}$. This formula also holds in the case of second-best variety: a central planner chooses J optimally considering that pricing decisions are subsequently left to firms.[18]

In the case where goods are homogeneous ($\Lambda_0 = 0$), consumers optimize ($\theta_t = \theta_\tau = 1$), the marginal cost is constant ($c'(q) = c_0$), and taxation may affect the equilibrium number of firms when J is determined by the free-entry condition as in Besley (1989), the tax formula collapses to $\frac{dW}{dt} = (p_0(1+\tau_0) + t_0 - c_0)\frac{dQ_L}{dt} - F\frac{dJ}{dt}$.[19] In this case, the direct entry effect enters as a negative. The intuition is easiest to see in the case where there is a reduction in taxes which induces entry of new firms. Since firms are symmetric and marginal cost is constant, it is more efficient to produce output with existing firms than to have new firms enter and incur the fixed cost of production.[20]

It is also useful to express equation (6) in terms of the responsiveness of *firm* output to taxation by substituting for aggregate demand using the relation $\frac{dQ_L}{dt} = J\frac{dq_L}{dt} + q\frac{dJ}{dt}$ to get the following expression: $\frac{dW}{dt} = (p_0(1+\theta_\tau\tau_0) + \theta_t t_0 - c'(q_0))J_0\frac{dq_L}{dt} + (\Lambda_0 + (\theta_t t_0 + p_0\theta_\tau\tau_0)\, q_0 + \pi_0)\,\frac{dJ}{dt}$. The first term resembles the "business stealing" term in Mankiw and Whinston (1986), which is $J_0(p_0 - c'(q_0))\frac{\partial q}{\partial J}$. In the economy without taxes, there may be "too much" entry in the market equilibrium relative to the social optimum since firms do not internalize the effect of entry on other firms' profits. This is precisely the force in Besley (1989) that creates the possibility for taxes to increase social welfare. To see the formal connection to business stealing and inefficient entry, we note that $\frac{dq_L}{dt} = \frac{\partial q}{\partial J}\frac{dJ}{dt} + \frac{dq}{dt}\Big|_J$, where $\frac{dq}{dt}\Big|_J \equiv \frac{\partial q}{\partial p}\frac{dp}{dt} + \frac{\partial q}{\partial t}$. Assuming for simplicity that $\pi_0 = t_0 = \tau_0 = 0$, we may decompose the formula for marginal excess burden into the following:

$$\frac{dW}{dt} = \underbrace{\left(J_0(p_0 - c'(q_0))\frac{\partial q}{\partial J} + \Lambda_0\right)\frac{dJ}{dt}}_{\lesseqgtr 0} + \underbrace{(p_0 - c'(q_0))J\,\frac{dq}{dt}\Big|_J}_{<0} \tag{7}$$

[18] The proof is the following: the planner seeks to maximize $\max_J W(Q,J) = u(Q,J) - Jc\left(\frac{Q}{J}\right) - JF$ taking the pricing decisions of firms as given. When the planner solves for the second-best variety, she chooses J to set $\frac{dW(Q,J)}{dJ} = \frac{\partial W}{\partial Q}\frac{dQ}{dJ} + \frac{\partial W}{\partial J} = 0$. Then:

$$\frac{dW}{dt} = \left(\frac{\partial W}{\partial Q}\frac{dQ}{dJ} + \frac{\partial W}{\partial J}\right)\frac{dJ}{dt} + \frac{\partial W}{\partial Q}\frac{dQ_L}{dt} = \frac{\partial W}{\partial Q}\frac{dQ_L}{dt} = (p_0(1+\tau_0) + t_0 - c'(q_0))\frac{dQ_L}{dt}$$

[19] Besley (1989) assumes Cournot competition, but our results show that the same formula is valid for other types of competition that can be modelled using conjectural variations.

[20] One can also re-arrange this formula to show that in this case $\frac{dW}{dt} = (t_0 + p_0\tau_0)\,\frac{dQ_L}{dt} + (p_0 - c'(q_0))J\frac{dq_L}{dt}$.

The first term in parentheses in equation (7) is the marginal welfare gain (or loss) of additional variety. It is negative if the business-stealing effect $\left(J_0(p_0 - c'(q_0))\frac{\partial q}{\partial J} < 0\right)$ dominates the variety effect ($\Lambda_0 > 0$), in which case there is excessive entry. The second term is the standard distortionary output effect of the tax and is always negative. This formula shows that if the business-stealing effect is sufficiently strong so that $\frac{dq_L}{dt} > -\frac{\Lambda_0}{J_0(p_0-c'(q_0))}\frac{dJ}{dt}$, then taxes can actually increase welfare. Intuitively, if consumers do not value variety and new varieties primarily steal consumers away from existing varieties, then there is excessive entry and taxes serve a corrective function by discouraging entry at the margin.

In practice, equations (6) or (7) may be difficult to implement empirically since it is challenging to measure marginal cost $c'(q_0)$, and hence the markup $p_0 - c'(q_0)$. We use Lemma 1 to provide a remarkably simpler representation for the marginal excess burden that maps more easily to empirically estimable objects. We also define the pass-through rate as $\rho_t \equiv 1 + \frac{dp}{dt}$ and present a sufficient statistics formula for it and the entry effects in our model.

Proposition 1. *Assume $\nu_q \in (0, J]$. Under the free-entry condition (and therefore $\pi_0 = 0$), the marginal excess burden, pass-through, and entry effects of a small change in the* ***specific tax*** *t are given by the following:*

$$\frac{dW}{dt} = \Lambda_0 \frac{dJ}{dt} - Q_0 \frac{dp}{dt} + (\theta_t t_0 + p_0 \theta_\tau \tau_0)\frac{dQ_L}{dt} \tag{8}$$

$$\rho_t = 1 - (1 - \omega_t)\,\theta_t \tag{9}$$

$$\frac{dJ}{dt} = -\frac{\theta_t J \epsilon_D}{(1+\tau_0)p_0 + t_0}\left[\frac{1 + \frac{\epsilon_D^* - \frac{\nu_q}{J_0}}{\frac{\nu_q}{J_0}\epsilon_S} + \frac{1}{\epsilon_{ms}}}{\triangle}\right] \tag{10}$$

where $\omega_t = \frac{\triangle + \frac{\nu_q}{J_0}\left[1 - \frac{1}{\epsilon_{ms}}\right]}{\triangle}$ is the pass-through formula when there is full optimization ($\theta_t = 1$) .

Equation (8) extends the welfare formulas of Harberger (1964), Besley (1989), Auerbach and Hines (2001), Chetty (2009), Chetty, Looney and Kroft (2009), Taubinsky and Rees-Jones (2018), Farhi and Gabaix (2020), and Kroft et al. (2020*a*), and equation (9) the pass-through formulas of Weyl and Fabinger (2013), Adachi and Fabinger (2017) and Kroft et al. (2020*b*) to the case of love-of-variety preferences and endogenous entry with imperfect competition. In order to develop intuition for these formulas, we consider the special case where $\theta_t = \theta_\tau = 1$, $\Lambda_0 = 0$, and $t_0 = \tau_0 = 0$.

Corollary 1. *Assume $\nu_q \in (0, J]$. Consider the case of full-optimization ($\theta_\tau = \theta_\tau = 1$), homogeneous products ($\Lambda_0 = 0$) and no pre-existing taxes ($t_0 = \tau_0 = 0$). The marginal excess* *burden, pass-through, and entry formulas for a* ***specific tax*** t *are given by the following:*

$$\frac{dW}{dt} = -Q_0 \frac{dp}{dt} \tag{11}$$

$$\rho_t = \frac{2 + \frac{\epsilon_D^* - \frac{\nu_q}{J_0}}{\frac{\nu_q}{J_0}\epsilon_S}}{2 + \frac{\epsilon_D^* - \frac{\nu_q}{J_0}}{\frac{\nu_q}{J_0}\epsilon_S} - \frac{\nu_q}{J_0}\left(1 - \frac{1}{\epsilon_{ms}}\right)} \tag{12}$$

$$\frac{dJ}{dt} = -\frac{J\epsilon_D}{p_0}\left[\frac{1 + \frac{\epsilon_D^* - \frac{\nu_q}{J_0}}{\frac{\nu_q}{J_0}\epsilon_S} + \frac{1}{\epsilon_{ms}}}{2 + \frac{\epsilon_D^* - \frac{\nu_q}{J_0}}{\frac{\nu_q}{J_0}\epsilon_S} - \frac{\nu_q}{J_0}\left(1 - \frac{1}{\epsilon_{ms}}\right)}\right] \tag{13}$$

Corollary 1 shows that the welfare cost of taxation depends only on the price effect, as in Besley (1989) and Delipalla and Keen (1992).[21] Intuitively, firm profits are always 0 so whether social welfare increases depends on the effects on consumer surplus and government revenue. The mechanical effect of a dollar increase in taxes is a loss of $1 for consumers but a gain of $1 for the government and so is neutral for welfare. Therefore, the net effect on consumers and social welfare depends on whether producer prices rise or fall with the tax increase. In the case where producer prices fall, consumers are better off, and social welfare increases. The formula in Proposition 1 generalizes the marginal excess burden formula in these papers to allow for a love-of-variety and pre-existing taxes. As discussed above, when consumers have a preference for variety, there is an additional effect on consumer surplus since varieties are affected by the tax change. Additionally, when there are pre-existing taxes, one must account for the fiscal externality on government revenue.

Turning to pass-through, we see that similar to Weyl and Fabinger (2013), the formula for ρ_t in Corollary 1 depends on the conduct parameter $\left(\frac{\nu_q}{J_0}\right)$, the elasticities of demand and supply $(\epsilon_D^*, \epsilon_S)$, and the elasticity of the inverse marginal surplus (ϵ_{ms}).[22] In Kroft et al. (2021), we demonstrate that under a standard regularity condition, there is more pass-through in the long-run compared to the short-run. Intuitively, if a tax increase leads to firm exit, there is less competition in the market and thus higher prices. It is also immediate from the

[21]Note that in this case, $\frac{dW}{dt} = -Q_0 \frac{dp}{dt} = J(p_0 - c'(q_0))\frac{dq}{dt}$ which is the formula in Besley (1989). Since $\frac{dq}{dt} = \frac{\partial q}{\partial J}\frac{dJ}{dt} + \frac{dq}{dt}|_J$, whether welfare rises or falls with the tax depends on whether the improvement in welfare along the entry and exit margin due to less business stealing ($\frac{\partial q}{\partial J}\frac{dJ}{dt} > 0$) dominates the standard distortionary quantity response to the tax ($\frac{dq}{dt}|_J < 0$) so that the total output effect is positive, $\frac{dq}{dt} > 0$. The formula for $\frac{dq}{dt}$ is provided in the Appendix. This is because behavioral responses do not have a first-order effect on consumer surplus and do not have a "fiscal externality" on the government's budget since there are no pre-existing taxes in the baseline economy.

[22]Besley (1989) and Delipalla and Keen (1992) considered comparative statics with respect to local changes in taxes. Their focus was primarily on deriving conditions for overshifting and undershifting of taxes onto prices (which we consider below) as opposed to deriving interpretable formulas in terms of sufficient statistics, as in Weyl and Fabinger (2013) and Kroft et al. (2020*a*). In the Appendix, we formally establish the connection between the pass-through formulas in our paper and Delipalla and Keen (1992).

pass-through expression that there is overshifting of taxes ($\rho_t \geq 1$) whenever $1 - \frac{1}{\epsilon_{ms}} \geq 0$. Proposition 1 generalizes the expression by accounting for the love-of-variety which enters through $\Delta = -\frac{J\epsilon_D^*}{pq}\frac{\partial \pi}{\partial J}$. Since the effect of love-of-variety on pass-through operates partially through firms' entry and exit decision, it is instructive to first examine how this margin is affected by taxation. More formally, this can be seen by noting that the denominator for both ρ_t and $\frac{dJ}{dt}$ is equal to $\Delta = -\frac{J\epsilon_D^*}{pq}\frac{\partial \pi}{\partial J}$, showing that similar forces are at play on both the pricing and entry margin in equilibrium.

The effect of taxes on entry is derived by using the implicit function theorem on the long-run entry condition $\pi(q(J,t,\tau_0),J,t,\tau_0) = 0$, and the first-order condition of the firm $\frac{\partial \pi}{\partial q} = 0$, so that $\frac{dJ}{dt} = -\frac{\frac{\partial \pi}{\partial t}}{\frac{\partial \pi}{\partial J}}$. Entry is more responsive to taxes when either taxes lower short-run profits by a lot or when a change in the number of firms has a small effect on profits. In the extreme case of perfectly inelastic demand $\epsilon_D = 0$ (or $\theta_t = 0$ and $\epsilon_D > 0$), $\frac{dJ}{dt} = 0$. Intuitively, if consumers are completely inelastic with respect to the tax, then firms do not bear any burden of the tax and so there is no change in profits and thus no entry or exit.[23]

Next, we characterize the welfare, pass-through, and entry effects of an ad valorem tax.

Proposition 2. *Assume $\nu_q \in (0, J]$. Under the free-entry condition (and therefore $\pi_0 = 0$), the marginal excess burden, pass-through, and entry effects of a small change in the **ad valorem tax** τ are given by the following:*

$$\frac{dW}{d\tau} = \Lambda_0 \frac{dJ}{d\tau} - Q_0 \frac{dp}{d\tau} + (\theta_t t_0 + p_0 \theta_\tau \tau_0)\frac{dQ_L}{d\tau} \tag{14}$$

$$\rho_\tau = 1 - (1-\omega_\tau)\frac{(1+\tau_0)\theta_\tau}{1+\theta_\tau \tau_0} \tag{15}$$

$$\frac{1}{p_0}\frac{dJ}{d\tau} = -\frac{\theta_\tau J \epsilon_D}{(1+\tau_0)p_0 + t_0}\left[\frac{1 + \frac{\epsilon_D^* - \frac{\nu_q}{J_0}}{\frac{\nu_q}{J_0}\epsilon_S} + \frac{1}{\epsilon_{ms}} + \frac{1-\frac{\nu_q}{J_0}}{\epsilon_D^*}}{\triangle}\right] \tag{16}$$

where $\omega_\tau = \frac{\Delta + \frac{\nu_q}{J}\left(1 - \frac{1}{\epsilon_{ms}} + \frac{1}{\epsilon_D^*}\left(\frac{J\frac{\partial P}{\partial J}}{p(1+\tau)+t}\epsilon_D - 1\right)\right)}{\Delta}$ is the pass-through formula when there is full optimization ($\theta_\tau = 1$).

The formulas derived in Proposition 2 are straightforward extensions of the case of specific taxes. The Appendix considers the special case of the model where there is full-optimization

[23] For additional intuition, roughly speaking ϵ_D characterizes the strength of the effect of taxes on variety and Λ_0 characterizes the strength of changes in variety on prices. When $\epsilon_D > 0$ and $1 - \frac{1}{\epsilon_{ms}} > 0$, a higher Λ_0 leads to higher pass-through whenever $\frac{\nu_q}{J_0} + \frac{\epsilon_D^* - \frac{\nu_q}{J_0}}{\epsilon_S} + \frac{\frac{\nu_q}{J_0}}{\epsilon_{ms}} > 0$. Corollaries 6 and 7 in Kroft et al. (2021) show that this condition is satisfied whenever taxes lower short-run profits and formally establish the relationship between the short-run and long-run effects of taxation. Intuitively, when consumers value variety, a marginal change in J leads to a greater loss in profits.

($\theta_t = \theta_\tau = 1$), homogeneous products ($\Lambda_0 = 0$) and no pre-existing taxes ($t_0 = \tau_0 = 0$), which can be compared to Corollary 1 above. In this case, the pass-through expression is the same as the one in Corollary 1 for the specific tax except for the extra term $-\frac{\frac{\nu_q}{J_0}}{\epsilon_D^*}$ in the numerator, which implies that $\rho_\tau < \rho_t$ whenever $\Lambda_0 = 0$.[24]

To compare the welfare effects of specific taxation to ad valorem taxation, we define the marginal cost of public funds for each form of taxation as:

$$MCPF_t \equiv -\frac{\frac{dW}{dt}}{\frac{dR}{dt}}$$

$$MCPF_\tau \equiv -\frac{\frac{dW}{d\tau}}{\frac{dR}{d\tau}}$$

Proposition 3. *Let $\theta_t = \theta_\tau \in [0,1]$ and $\nu_q \in (0, J]$. Consider the long-run free-entry condition $\pi = 0$. Then:*

1. Starting from $\tau_0 = t_0 = 0$, the ranking of pass-through for both forms of taxation is given by:

$$\rho_\tau > \rho_t \Leftrightarrow \epsilon_D \frac{J}{p_0} \frac{\partial P}{\partial J} > 1$$

2. Independent of (τ_0, t_0), the effect of each form of taxation on entry is ranked by:

$$\frac{1}{p_0} \frac{dJ}{d\tau} < \frac{dJ}{dt}$$

3. Starting from $\tau_0 = t_0 = 0$, the marginal costs of public funds satisfy:

$$MCPF_t = -\frac{\Lambda_0}{Q_0} \frac{dJ}{dt} + \rho_t - 1$$

$$MCPF_\tau = -\frac{\Lambda_0}{p_0 Q_0} \frac{dJ}{d\tau} + \rho_\tau - 1$$

4. Assume parallel inverse aggregate demands so that $\frac{\partial P}{\partial J}(Q, J) = \frac{\Lambda(Q,J)}{Q}$.[25] Starting from

[24] This is a standard result in the public finance literature. To see the intuition for why ad valorem taxes lead to lower consumer prices, consider a marginal increase in the producer price. In the case of a specific tax, the firm gets the full amount of the price increase on the inframarginal units; however, with an ad valorem tax, the firm only keeps the part of the extra revenue on the inframarginal units that do not go to the government. Thus, firms face less of an incentive to increase prices at the margin under an ad valorem tax.

[25] The parallel demands assumption has been thoroughly analyzed in Kroft et al. (2020*c*) for discrete choice models, and in the Appendix we present a microfoundation for parallel demands in our continuous choice model. This assumption is not needed for any of the results in Propositions 1 and 2, and without this assumption

$\tau_0 = t_0 = 0$, the desirability condition ranking $MCPF_\tau$ and $MCPF_t$ is given by:[26]

$$MCPF_\tau > MCPF_t \Leftrightarrow \frac{\Lambda_0 \epsilon_D}{p_0 q_0} > \frac{\nu_q}{J_0} \Leftrightarrow \Lambda_0 > (p_0 - mc(q_0))\, q_0$$

This analysis permits a comparative evaluation between ad valorem and specific taxes. There are several noteworthy results from Proposition 3. First, we see that when $\frac{\partial P}{\partial J} = 0$ so that love-of-variety $\Lambda_0 = 0$, pass-through for ad valorem taxes is always less than pass-through for specific taxes, and ad valorem taxes are more efficient than specific taxes which is well-known in the literature (see Delipalla and Keen 1992 and Anderson, De Palma and Kreider 2001*a*;*b*). Second, we see that if the variety effect is sufficiently strong, then both of these results can be overturned. In particular, if consumers have a strong preference for variety, then consumer prices might be higher and welfare might be lower under ad valorem taxation. Additionally, we see that if consumer prices are higher under ad valorem taxation, then welfare is higher under specific taxation, but the converse need not be true.[27]

Our desirability condition in Proposition 3 accounts for both the static distortion to output and the distortion to product variety and highlights the role of the variety effect (Λ_0) and the markup ($p_0 - mc(q_0)$). Intuitively, if the variety effect is large relative to the markup, there is insufficient entry. Given that ad valorem taxes cause more of a distortion on the entry margin due to the multiplier effect (Keen 1998) (result 2 in Proposition 3), it follows that when the variety effect is large, specific taxation has more favorable welfare effects than ad valorem taxation.[28] Our results thus imply that in markets with stronger love-of-variety and more elastic demand, specific taxation is more likely to be a preferable form of taxation than ad valorem taxation.

---

one can still show that $MCPF_\tau > MCPF_t \Leftrightarrow \frac{\Lambda_0 \epsilon_D}{p_0 q_0} + \frac{\nu_q}{J_0}\epsilon_D \left(\frac{\partial P}{\partial J}\frac{J_0}{p_0} - \frac{\Lambda_0}{p_0 q_0}\right) > \frac{\nu_q}{J_0} \Leftrightarrow \Lambda_0 + \frac{\nu_q}{J_0}\left(Q_0 \frac{\partial P}{\partial J} - \Lambda_0\right) > (p_0 - mc(q_0))\, q_0$.

[26] It is worth noticing that in the knife edge case where $\epsilon_D = 0$, the Lerner condition implies $\frac{\nu_q}{J_0} = 0$, and so the expression in the middle should not be taken into account; instead one might apply the last condition $MCPF_\tau > MCPF_t \Leftrightarrow \Lambda_0 > (p_0 - mc(q_0))\, q_0$. Moreover, $\epsilon_D = 0$ and equation (6) imply $\frac{dW}{dt} = (\Lambda_0 - [p_0 - c'(q_0)]q_0)\frac{dJ}{dt}$ where $\frac{\partial W}{\partial J} = \Lambda_0 - [p_0 - c'(q_0)]q_0$; therefore ad valorem taxation is to be preferred if and only if there is over-entry of firms (or excessive product variety).

[27] Note that if variety is efficient, then when demand is perfectly inelastic (so that there is no static distortionary effect of taxation) the welfare cost of taxation and hence the marginal cost of public funds is 0 for *both* specific and ad valorem taxation. Kroft et al. (2021) contains a more detailed summary of the differences between our results and the results in Anderson, De Palma and Kreider (2001*b*).

[28] In particular, starting from $\tau_0 = t_0 = 0$, for a given initial $J(0)$ it can be shown that for revenue equivalent t and τ, result 2 in Proposition 3 implies that $J(\tau) < J(t)$. To see the intuition for this result, consider an increase in the fixed cost of production (ΔF) and suppose that profits are initially 0. In order for the firm to break even with a specific tax (holding constant the demand response so that $\Delta q = 0$), it has to increase the consumer price by $\Delta p = \Delta F/q$. However, with an ad valorem tax, the firm has to increase consumer prices by $\Delta p = \Delta F/(q(1-\tau))$. Thus ad valorem taxation leads to less entry than specific taxation.

4 Estimation of Reduced-Form Effects of Taxation

This section discusses the estimation of the reduced-form effects of sales taxes on (pre-tax) prices, quantity demanded, and product variety. We first describe the data used in the empirical analysis, and then the empirical strategy. Lastly, we discuss the reduced-form estimates that we will use as inputs into our calibrations in Section 5.

4.1 Data Description

We combine the Nielsen Retail Scanner (RMS) data for the years $2006 - 2014$,with data on state- and county-level sales tax rates and tax exemptions. The RMS data records weekly prices and quantities by product at the barcode (UPC) level for 35,000 stores in the United States. Products are organized in a hierarchical structure:UPCs are categorized into approximately 1,200 *product-module*s (e.g., fresh eggs, milk, aluminum foil, batteries, frozen desserts).

We define all of our variables at the level of module (m), store (r), county (c), and year (n), which requires aggregating the data over time and across products. We use the RMS data to generate measure of prices, quantity demanded and product variety. Details on the construction of these variables are provided in the Data Appendix.

We then assign each module-store-year observation a tax rate based on the rate effective on September 1st in the county in which a store is located. For each module, the effective tax rate, τ_{mcn}, depends on county and state sales tax rates and product-specific exemptions.[29]Grocery stores typically sell both food and nonfood products, and effective tax rates on food and nonfood products differ in most states.

To address the concern that sales taxes are spatially correlated across regions of the United States in ways that may endogenously reflect the geographic distribution of consumer preferences, we restrict our sample to grocery stores located in contiguous counties on opposite sides of a state border to implement a "county border pair" research design following Holmes (1998) and Dube, Lester and Reich (2010). For computational reasons, we further restrict our analysis sample to the top 20 percent of modules in terms of total sales. Table 1 presents the summary statistics for our final sample, which includes more than 11 million module-store-year observations covering 3,822 grocery stores, 198 modules, and 543 counties over 9 years. 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 counties in our sample are part of 497 different county border pairs.[30]

The sample average tax rate in our sample is 3.4 percent with a standard deviation of

[29]Tax rates may be measured with some error if we misclassified products and thereby mis-assigned exemptions. We show in Online Appendix E.4 that our results are unlikely to be affected by such measurement error.

[30]For estimation purposes, the original dataset is rearranged by stacking all pairs. For instance, a module-store cell located in county c appears as many times as the number of counties county c is paired with.

3.2 percentage points.[31] The average is considerably below average legislated sales tax rates since our sample includes a large number of tax-exempt products for which the tax rate is 0. Table 1 also reports summary statistics split by above-median and below-median residual tax rates (net of module-by-county-border-pair-by-year and store-by-year fixed effects). Average pre-tax prices are slightly higher while quantity and product variety are slightly lower in above-median tax rate cells compared to below-median tax rate cells.[32]

4.2 Estimation Strategy

Our empirical approach mimics a cross-sectional difference-in-differences strategy, where the first difference is across products within stores and the second difference is across stores within county border pairs.[33] We estimate the effects of sales taxes using the following "county border pair" regression model:

$$\log y_{mrdn} = \sum_{z=2006}^{2014} \beta^{y,z} \left[\log(1+\tau_{mcn}) \times 1\{z = n\}\right] + \delta_{mdn} + \delta_{rn} + \varepsilon_{mrdn} \tag{17}$$

where the outcome y_{mrnd} is either pre-tax prices p, quantity Q, or product variety J for module m, store r, and year n, and each county belongs to one or more county border pairs indexed by d. The term τ_{mcn} is the sales tax rate that applies to module m in county c in calendar year n. All regressions are weighted by the inverse of the number of county pairs that a store is part of.

From equation (17), we obtain one coefficient estimate per year. We then summarize the estimated effect of sales taxes on outcome y, β^{y}, by taking a simple average of all the coefficient estimates $\beta^{y,z}$, putting equal weight on all 9 yearly cross-sectional estimates (i.e.,

[31] The tax rate variation in our sample is across modules, counties, stores, and time. Some of that variation is absorbed by our fixed effects. Since identification comes from comparing across modules within stores, and across stores that are located in different counties within the same county border pair at a point in time, we residualize the tax rate variable and calculate that this residualized tax rate has a sample standard deviation of 1.1 percentage points (Appendix Table OA.2). This means that a two-standard-deviation change in the residualized tax rate is 2.2 percentage points, which is a meaningful enough tax change to lead to fairly precise estimates of the effect of taxes on prices, quantity, and variety.

[32] While the differences in residual prices, quantity, and variety between above-median and below-median cells are small, the difference in average residual tax rates between the above-median and below-median cells is also small (4.1 percent average tax rate in above-median cells versus 2.8 percent average tax rate in below-median cells. Scaling the difference in average prices, quantity, and variety by the difference in average tax rates, we calculate implied effects of taxes of 0.065 (prices), -1.098 (quantity), and -0.488 (variety). These are the same sign and the same order of magnitude as the estimates we recover from the full regression model described in the next section.

[33] We use cross-sectional variation to approximate the steady-state, "long-run" adjustments to tax policy in prices, quantity demanded, and product variety. Atkin, Faber and Gonzalez-Navarro (2018) similarly use cross-sectional variation in store-level prices to estimate long-run elasticities of substitution across stores.

$\beta^y = \frac{1}{9}\sum \beta^{y,z}$ for $y \in \{p, Q, J\}$).[34] To account for spatial auto-correlation as well as the fact that some counties border multiple states and therefore appear multiple times in the data, standard errors are two-way clustered by border-pair-by-module and state-module in all specifications (Boone et al. 2016, Cameron, Gelbach and Miller 2011).

The terms δ_{mdn} and δ_{rn} are module-by-border-pair and year-specific store fixed effects, respectively. The effects of sales taxes are therefore identified under the assumption that tax rates are uncorrelated with product-specific unobservable determinants of demand conditional on these fixed effects.[35] Intuitively, the identifying assumption is that state and local governments do not set local tax rates and product-specific exemptions based on their relative market shares. For example, our estimate of the effect of taxes on quantity demanded would be biased upwards if jurisdictions where the consumption share of unhealthy food products (e.g., candy, soft drinks) is relatively high responded to this high demand by specifically choosing not to tax exempt these goods that would otherwise qualify as food. The inclusion of module-by-border pair fixed effects accounts for broad spatial differences in tastes and tax rates for specific modules, effectively restricting comparisons to nearby stores that are in different states and therefore face different sales tax rates and exemptions for reasons that are plausibly unrelated to demand factors. Any store and county-level differences that do not vary across modules are absorbed by the store fixed effects. Yet, our results may not generalize to other settings as cross-border shopping is likely an important margin of substitution in a border-design approach.

One concern with interpreting the OLS estimates as the causal effect of sales taxes is that there is existing evidence that border counties adjust local sales tax rates strategically to compensate for cross-border differences in state-level sales tax rates (Agrawal 2015). We assess whether this is an important source of bias in our empirical setting by instrumenting the statutory tax rate τ_{mcn} with the state-level average $\bar{\tau}_{mc(s)n}$, which we calculate as the average tax rate across stores in the same state excluding all stores located in county c, separately for each module and year.

Another concern is how the presence of online shopping options affects the interpretation of our estimates. If sales taxes reduce measured variety in the Nielsen data but consumers can substitute online, this would effectively mitigate any loss in utility from fewer varieties in the inside market. There are several reasons why "offline-online" substitution may be limited in

[34] This econometric approach is numerically equivalent to estimating the model separately for each year and then averaging the coefficients.

[35] To assess the validity of our research design, in Appendix Figure OA.4 we report correlations between a number of different county characteristics (%rural, median age, %college, median income, unemployment rate, %white, %Black) and county tax rates across counties within the same county border pair. While county characteristics are correlated with county-level tax rates in our full sample of all counties, we find very little correlation once we condition on county border pair fixed effects, suggesting that within county border pair county tax rates are not correlated with other determinants of consumer demand.

our setting. First, in-state purchases online are also subject to sales taxes. Thus, when local sales taxes increase, this is likely to affect the number of varieties available online, as well. Second, the monetary cost of searching online has been estimated to be quite large (see, e.g., Hong and Shum 2006 and Brynjolfsson, Dick and Smith 2010). Consistent with this, Einav et al. (2014*b*) estimate that a one percentage point increase in a state's sales tax leads to an increase of just under 2 percent in online shopping. Finally, over the time period we study, online purchases represented only 5 percent of overall spending by households in Nielsen's Consumer Panel (Baker, Johnson and Kueng 2021).

4.3 Reduced-form Estimates

The main results from estimating equation (17) are reported in Table 2. The dependent variable is the average pre-tax price in column (1), quantity demanded in column (2), and product variety in column (3). Since we use pre-tax prices in column (1), a coefficient estimate of zero corresponds to full pass-through of sales taxes to consumers (i.e., a pass-through rate equal to one).

We report OLS estimates of the model in equation (17) in Panel A of Table 1. We find a small amount of overshifting of taxes onto pre-tax prices with a coefficient of $\hat{\beta}^p = 0.038$ (s.e. 0.016), implying a pass-through rate slightly greater than one. We find that the elasticity of quantity demanded with respect to sales taxes is $\hat{\beta}^Q = -0.677$ (s.e. 0.154). This estimate is very similar to previous work studying tax salience using the same scanner data; for example, Kroft et al. (2020*a*) estimate an output elasticity of -0.649 (s.e. 0.084) using a different source of tax variation.

Lastly, we find that the estimated elasticity of product variety with respect to sales taxes is equal to $\hat{\beta}^J = -0.236$ (s.e. 0.074). We are not aware of existing estimates of the effects of taxes on product variety for a broad cross-section of product-modules, so we cannot benchmark this estimate to the previous literature. One way to interpret the magnitude is to note that the overall effect of taxes on quantity demand can be decomposed as $\frac{d\log(Q)}{d\log(1+\tau)} = \frac{d\log(J)}{d\log(1+\tau)} + \frac{d\log(q)}{d\log(1+\tau)}$; i.e., the total elasticity is the sum of the variety elasticity and the "quantity demanded per variety" elasticity. Using this decomposition, we find the variety elasticity accounts for about one-third of the overall effect of taxes on quantity demanded. Panel B reports 2SLS estimates using the state-level average tax rate as an instrument, and we find results that are very similar to the OLS results, suggesting that product-specific endogenous sales tax rates within county border pairs are not a substantial source of bias in this setting.

Our theoretical model assumes symmetry across products so that consumers only care about the total number of products. In reality, of course, some products have broader appeal to consumers than others, and as a result, the actual change in product variety following a change in taxes could either overstate or understate love-of-variety depending on which products enter and exit the market. To investigate this, we modify our definition of product variety to be a weighted sum of the number of products available using national market shares as weights. This alternative measure puts relatively more weight on popular products relative to low-market-share products. Table 3 compares the results using the two different variety measures across columns. Both the OLS and 2SLS estimates show somewhat smaller magnitudes of the effects of taxes on weighted product variety compared to unweighted product variety (between -0.17 and -0.19 instead of -0.24). This implies that the "marginal varieties" that exit following an increase in sales taxes tend to be more "niche" products that have lower than average national market shares.

Our estimates of $\hat{\beta}^p$, $\hat{\beta}^Q$, and $\hat{\beta}^J$ represent averages across all of the modules in our sample, which cover a wide range of food and non-food products. We investigate whether the effects of taxes vary across types of products in Online Appendix E.2, and find broadly similar pass-through across the categories, and we find some suggestive evidence of larger effects of taxes on quantity and variety in health and beauty products compared to food products.

We assess the reliability and robustness of our main results in several ways in Online Appendix E.3. First, we show that the county border pair estimates are very stable across years. Second, we show that our results do not rely on specific county border pairs by dropping each state (one at a time).

Lastly, we implement placebo tests that exploit the difference in average tax rates between food and nonfood products that arise due to local tax exemptions. While nonfood tax rates generally exceed food tax rates in most counties, the gap between average food and nonfood tax rates varies substantially within county border pairs. We examine whether prices, quantity, and product variety vary with the tax rate that applies to *other* products in the same jurisdiction, conditional on the tax rate they are actually subject to (on the basis of whether they are food or nonfood products). These results are reported in Appendix Table OA.1. For both quantity and variety, the "placebo" tax rate has no residual explanatory power. Prices appear to increase slightly in response to both food and nonfood tax rate increases, though the coefficient on the "own" tax rate is a bit larger. This overshifting is small in magnitude, however, and the results for quantity and variety are fairly similar to the preferred estimates reported in Table 2.[36] We conclude from these placebo tests that our results reflect genuine causal effect of taxes rather than confounding factors that are correlated with both food and nonfood tax rates within a county.

Overall, we conclude that taxes have a clear effect on overall quantity demanded, and we also find some evidence of a small amount of overshifting of taxes and a modest reduction in

[36]Given our estimate of the demand elasticity, any "indirect" effect of taxes on other products on own quantity operating via price increases would be equal to -0.14, a value that falls within the confidence interval of the estimated "indirect" quantity response.

product variety.

5 Model calibrations

This section discusses identification and estimation of the deeper model parameters and calibrates the main welfare formula in equation (14) using the model-based parameters. We also show how estimating the deeper parameters allows us to learn about whether or not variety is socially optimal at current tax rates, but our main goal is to illustrate how the reduced-form effects of taxes on prices, quantity, and variety can be used to estimate consumers' love of variety using the full structure of our model.

5.1 Identifying and Estimating Model Parameters from the Reduced-Form Results

We use the formulas in Proposition 2 in Section 3 above that define each of the reduced-form effects of taxes (on prices, variety, and total quantity) in terms of the model parameters. We assume constant marginal costs (so that $\varepsilon_S = \infty$), and we calibrate the baseline tax rate of $\tau_0 = 0.034$ based on the average tax rate in our data (Table 1). We calibrate the price elasticity of demand to be $\epsilon_D = 1.170$ and the tax salience parameter to be $\theta_\tau = 0.556$ based on the estimates in Kroft et al. (2020*a*). The price elasticity of demand is estimated by instrumenting store-level prices with the average prices of other stores in the same retail chain, exploiting the tendency of chains to choose uniform prices across stores (DellaVigna and Gentzkow 2019). The salience parameter is estimated using changes in sales taxes within US counties over time and is approximately equal to the ratio of the tax elasticity to the price elasticity. We then make the following technical assumption that is necessary for identification. Specifically, we assume that $P(Q, J)$ shifts in parallel when variety J changes.

Assumption 1. *Inverse aggregate demands shift in parallel. Then* $\frac{\partial P}{\partial J}(Q, J) = \frac{\partial P}{\partial J}(Q', J)$ *for all* $Q, Q' > 0$.

In particular, under Assumption 1, $P(Q, J)$ is linearly separable in Q and J. In this case $\frac{\partial P}{\partial J}(Q, J) = \frac{\Lambda(Q,J)}{Q}$ and $\frac{\partial^2 P}{\partial J \partial Q} = 0$. This assumption has been thoroughly analyzed in Kroft et al. (2020*c*) for discrete choice models, and in the Appendix we present a microfoundation for parallel demands in our continuous choice model. As discussed in Kroft et al. (2020*c*), this assumption is satisfied by a range of discrete choice random utility models including standard Logit and nested Logit models, as well as random utility models with certain types of unobserved product heterogeneity. This assumption does rule out some other models of demand, however, such as the linear market demand system used in (Melitz and Ottaviano 2008) and the discrete choice model with random coefficients on consumer price used in Berry, Levinsohn, Pakes (1995), but we emphasize that this assumption is not strictly necessary for any of our welfare, pass-through, and entry formulas in Section 3. This assumption is only used for identification and estimation of the variety effect parameter, Λ_0.

Formally, this assumption allows us to replace the $\frac{\partial^2 P}{\partial J \partial Q}$ term that appears in both the pass-through and entry formulas (inside the Δ expression) with the variety effect parameter. As a result, this leaves three remaining unknown parameters – the love-of-variety parameter ($\tilde{\Lambda}_0 \equiv \Lambda_0 \frac{J}{p(1+\tau)Q}$), the elasticity of inverse marginal surplus (ε_{ms}), and the conduct parameter (v_q/J) – which we can then solve for using the three reduced-form estimates ($\hat{\beta}^p$, $\hat{\beta}^Q$, and $\hat{\beta}^J$). Specifically, we can solve for the three unknown parameters that make the model-implied effects of taxes exactly match the reduced-form estimates in Panel B of Table 2. The calibrated parameters and the parallel demands are therefore sufficient to be exactly identified.

We can build on the theoretical analysis in Section 3 to provide an intuitive discussion of how each parameter is identified. The conduct parameter is identified by the long-run free-entry condition (Lemma 1), which implies that the markup is equal to the ratio of the effects of taxes on pre-tax prices to the effect of taxes on the quantity demanded per firm. In terms of the reduced-form estimates, this means that the markup is equal to $-\hat{\beta}^p/(\hat{\beta}^Q - \hat{\beta}^J)$. Using the reduced-form estimates in Table 2, Panel C of Table 4 reports the markup estimate of $(p - c'(q))/p = 0.080$. To get a sense of how precise this estimate is, we bootstrap the entire estimation procedure and obtain 100 sets of estimates for $\hat{\beta}^p$, $\hat{\beta}^Q$ and $\hat{\beta}^J$, and we calculate the markup (as well as all other model parameters) for each of these replications. We obtain a standard error of 0.042.

Given this estimate of the markup, we can then use the generalized Lerner index defined in equation (1) to recover the conduct parameter using the (calibrated) values of the tax rate, the tax salience parameter, and the price elasticity of demand. This results in an estimate of $v_q/J = 0.092$ (s.e. 0.049), where 0 is perfect competition and 1 is perfect collusion. This implies that there is a high degree of competition in our sample of retail grocery stores, which is plausible given that retail stores are typically thought to operate under fairly small margins (see Kroft et al. (2020*a*) for a more detailed discussion).[37]

With the conduct parameter in hand, we can identify the elasticity of inverse marginal surplus using the formula for the pass-through of taxes into consumer prices (see Proposi-

[37] Kroft et al. (2020*a*) estimates a somewhat smaller mark-up of around 3 percent, but that analysis does not allow for endogenous entry, and the estimates are based on relative short-run responses to quarterly variation in sales taxes. The present paper instead uses cross-sectional variation, which one can loosely interpret as corresponding to steady-state, "longer run" effects of taxes allowing for endogenous entry. As a result, our preferred interpretation of the mark-up in the present paper is a "longer run" mark-up that would need to cover additional fixed costs that would be relevant for product entry decisions. We, therefore, view it as fairly reassuring that the reduced-form effects of taxes on quantity are similar across the two papers, but the implied mark-up is larger given that the mark-up in the "longer run" has to cover additional (fixed) costs that would plausibly be interpreted as "sunk" in in the shorter run (quarterly) analysis in Kroft et al. (2020*a*).

tion 2). Corollary $A1$ shows that ignoring love-of-variety, the parameter ϵ_{ms} is identified by the pass-through rate given knowledge of the conduct parameter (recovered in the previous step). Importantly, Corollary $A1$ shows that varying the demand elasticity does not affect what we infer about the curvature parameter holding constant the pass-through estimate and conduct parameter. This is consistent with previous work that emphasizes that, under constant marginal costs, it is the demand curvature rather than the demand elasticity that determines pass-through (Bulow and Pfleiderer 1983, Weyl and Fabinger 2013). Our analysis shows that this logic extends to allowing for love-of-variety preferences and free entry. To see this more formally, we can use the definition of Δ and the expression for $dJ/d\tau$ to solve for the following expression of the love-of-variety parameter in terms of ϵ_{ms}, the conduct parameter, the reduced-form effects of taxes on variety and total quantity demanded, and the other (calibrated) parameters:

$$\begin{aligned}
\tilde{\Lambda}_0 \quad &= \quad k_1 + k_2 * (1/\epsilon_{ms}) \qquad (18)\\
& \quad k_1 \equiv \left(\frac{1+\theta_\tau \tau}{1+\tau}\right)\left(\frac{\hat{\beta}^Q}{\hat{\beta}^J}\left(1+\frac{1-\frac{\nu_q}{J}}{\epsilon_D^*}\right)+\frac{\frac{\nu_q}{J}}{\epsilon_D^*}-2\right)\\
& \quad k_2 \equiv \left(\frac{1+\theta_\tau \tau}{1+\tau}\right)\frac{\hat{\beta}^Q}{\hat{\beta}^J}
\end{aligned}$$

The formula above shows there is an affine mapping from $1/\epsilon_{ms}$ to $\tilde{\Lambda}_0$, and so we can substitute this expression into the pass-through expression, leaving only ϵ_{ms} as the remaining unknown parameter (as a function of the three reduced-form effect estimates and the other calibrated parameters). We can then use this expression to solve for ϵ_{ms}, and Panel C of Table 4 reports $\epsilon_{ms} = -0.903$ (s.e. 0.189).[38] According to our theoretical analysis, this parameter must be negative whenever there is overshifting of ad valorem taxes, which is what we find empirically (see Table 2).

Lastly, given estimates of conduct parameter and curvature parameter, we can then use the affine mapping in equation (18) to identify and estimate $\widetilde{\Lambda}$. This shows that the identification of the love-of-variety parameter comes primarily through the ratio $(\hat{\beta}^Q/\hat{\beta}^J)$. The larger the ratio, the larger the inferred love-of-variety. Intuitively, when β^Q increases relative to β^J (holding the price elasticity of demand fixed), this means that demand falls a lot for a given change in variety, which in turn reveals that consumers have a high willingness-to-pay

[38]Formally, there will always be two values of ϵ_{ms} since the pass-through formula can be re-written as a quadratic in $1/\epsilon_{ms}$. One of the solutions to the quadratic violates the necessary condition that $dJ/d\tau \neq 0$. This condition is necessary for identification to ensure that $(\hat{\beta}^Q/\hat{\beta}^J)$ exists. The other solution to the quadratic will not violate this condition, and that is the solution we solve for numerically. Note that there is no guarantee the resulting solution will satisfy the stability condition (i.e., $\Delta > 0$). As discussed above in Section 3, the stability condition will be violated whenever the estimate of $\widetilde{\Lambda}$ is very large, which will generally be the case whenever $(\hat{\beta}^Q/\hat{\beta}^J)$ is very large.

for variety. Using equation (18), we estimate the variety effect to be $\tilde{\Lambda}_0 = 0.125$ (s.e. 0.339).[39] This parameter can be given a willingness-to-pay interpretation: an exogenous 10 percent reduction in variety reduces average willingness-to-pay by 1.25 percent. This magnitude is smaller than the reciprocal of the demand elasticity (0.855), which is the love-of-variety value that would occur in a Logit model of consumer demand (and which is often thought to substantially overstate consumers' true love-of-variety). While the variety effect is less precisely estimated than other parameters, the standard errors we obtain suggest we can reject the null that it is equal to the reciprocal of the demand elasticity at conventional levels.

5.2 Calibrating the Marginal Excess Burden of Ad Valorem Taxes

With the love-of-variety parameter estimate in hand, we can now calibrate the main welfare formula using the reduced-form empirical estimates presented in Section 4. Since the estimates above are based on ad valorem taxes, we calibrate the marginal excess burden formula for the case of an ad valorem tax (τ). To obtain an expression in terms of the reduced-form elasticities, we set $t = 0$ and normalize the welfare formula by total firm revenues,

$$\frac{d\widetilde{W}}{d\tau} \equiv \frac{dW}{d\tau}\frac{1+\tau}{pQ} = \theta_\tau \tau_0 \frac{d\log Q_L}{d\log(1+\tau)} - \frac{d\log p}{d\log(1+\tau)} + \tilde{\Lambda}_0 \frac{d\log J}{d\log(1+\tau)}.$$

The reduced-form estimates $\hat{\beta^Q}$, $\hat{\beta^p}$, and $\hat{\beta^J}$ reported in Table 2 (Panel B) can be used to calibrate $\frac{d\log Q_L}{d\log(1+\tau)}$, $\frac{d\log p}{d\log(1+\tau)}$, and $\frac{d\log J}{d\log(1+\tau)}$, respectively. Using these estimates, we calculate $d\widetilde{W}/d\tau = -0.083$ (see Panel D of Table 4). This is larger in magnitude than a standard Harberger benchmark adjusted for salience effects ($d\widetilde{W}/d\tau = -0.014$, s.e. 0.003), which is one useful benchmark for comparison. Ignoring love-of-variety (i.e., assuming $\tilde{\Lambda}_0 = 0$), but accounting for endogenous product variety through free entry leads to a marginal excess burden of $d\widetilde{W}/d\tau = -0.053$ (s.e. 0.019). This is the estimate one would calibrate based on our extension of the theoretical results in Besley (1989) to cover ad valorem taxes. Our estimate of the full marginal excess burden (MEB) is 58 percent larger than this benchmark, and the reason we find a larger negative effect of taxes on welfare is that we find a positive love-of-variety estimate and a negative effect of taxes on product variety, and our formula shows that the MEB is increasing in the product of these two terms. We should note that the relative lack of precision of our estimate of the variety effect may affect these quantitative comparisons. Obtaining a more precise estimate might be important for future research.

We assess the robustness and sensitivity of these results in Figure 1, which shows how the excess burden varies across different scenarios.[40] The first scenario (1) reproduces the main results from Table 4 for comparison. Scenario (2) and (3) show sensitivity to using

[39] When solving for $\tilde{\Lambda}_0$ in each bootstrap replication, we impose the constraint that $\tilde{\Lambda}_0 \geq 0$.

[40] The underlying numbers are reported in Appendix Table OA.7.

the alternative variety response that accounted for heterogenous market shares (see Table 3, column (2)). The smaller variety response using this alternative variety measure implied that the marginal varieties that exited following a tax increase had lower-than-average national market shares. Scenario (2) shows that holding constant the variety effect and all other model parameters leads to a smaller MEB using this smaller variety response. In the baseline scenario, the full MEB is 58 percent larger than the alternative benchmark that ignores love-of-variety preferences. In scenario (2), this falls to 46 percent because of the smaller variety response.[41] In scenario (3), we continue to use the alternate variety response estimate, but we re-estimate all of the model parameters. In this case, the variety effect actually increases, since all of the other reduced-form responses are being held constant. In this case, the MEB is very similar to the baseline scenario given these offsetting effects. This shows a certain robustness in our bottom-line welfare conclusions to accounting for product heterogeneity, since accounting for product heterogeneity and re-estimating the model parameters leads to a similar contribution of love-of-variety preferences to the MEB. Lastly, the remaining scenarios (4) through (7) show sensitivity to perturbing the demand elasticity and the tax salience parameter, since there is meaningful uncertainty about the appropriate value of each of these parameters, particularly since both the demand elasticity and the tax salience parameter vary across the population (Kroft et al. 2020*a*). Reassuringly, the results are fairly similar across these scenarios for modest perturbations.[42] Overall, we conclude that our baseline scenario is fairly robust to alternative assumptions, and that the MEB is meaningfully larger than an alternative benchmark that accounts for imperfect competition but not consumers' love of variety.

5.3 Welfare analysis of product variety

We can also use the full structure of the model to infer whether or not variety is socially optimal. To see how this is achieved, we first re-write the marginal excess burden formula

[41] Note that we are implicitly assuming that our main welfare formula with product heterogeneity has the same structure as our main welfare formula in Proposition 2, so that we can multiply the variety effect times the variety response in calculating the MEB. We conjecture that this is true, although we do not have a formal proof.

[42] The reason the calibration results are not too sensitive to the demand elasticity and tax salience parameters is that we are perturbing the parameters in a way that keeps the product of the two fairly constant. Appendix Table OA.4 shows that if we just perturb the demand elasticity, the variety effect is much more sensitive. This is intuitive because varying the price elasticity but not varying the tax salience parameter or the reduced-form effect of taxes on quantity is going to make it very hard to make sense of the results. Another way of thinking about this result is that the ratio of the reduced-form effect of taxes on quantity to the price elasticity of demand should not be too different from the tax salience parameter, since in our model this only comes through endogenous entry and exit and consumers' love of variety. Since the entry/exit effects we find are modest, then we will need a very large (or very small) variety effect to match our reduced-form results when we perturb the demand elasticity.

that connects to equation (7) as follows:

$$\begin{aligned}\frac{d\widetilde{W}}{d\tau} &= \underbrace{\left(\partial\widetilde{W}/\partial\log(J)\right)\frac{d\log(J)}{d\log(1+\tau)}}_{\lesseqgtr 0}+\underbrace{\frac{(p_0-c'(q_0))}{p_0}\frac{d\log(q)}{d\log(1+\tau)}|_J}_{<0}\\ &= \left(\frac{(p_0-c'(q_0))}{p_0}\frac{\partial\log(q)}{\partial\log(J)}+\tilde{\Lambda}_0\right)\frac{d\log(J)}{d\log(1+\tau)}+\frac{(p_0-c'(q_0))}{p_0}\frac{d\log(q)}{d\log(1+\tau)}|_J\end{aligned}$$

The first term in parentheses multiplying the effect of taxes on variety ($\partial\widetilde{W}/\partial\log(J)$) provides a direct test for whether or not variety is socially optimal because it balances the "business-stealing" social costs of additional variety against the love-of-variety benefits of additional variety. In order to calibrate the expression, we need an estimate of the love-of-variety parameter and the mark-up, along with an estimate of $\frac{\partial\log(q)}{\partial\log(J)}$. In the Appendix (Lemma $A1$), we show that this partial elasticity can be defined in terms of the conduct parameter, the demand elasticity, and the inverse elasticity of marginal surplus. Using all of the estimated parameters described in previous subsection and listed in Table 4, we estimate $\frac{\partial\log(q)}{\partial\log(J)} = -0.751$ (see Panel B of Table 5). Multiplying this partial elasticity by the average mark-up and adding the love-of-variety parameter leads to an estimate of $\partial\widetilde{W}/\partial\log(J) = 0.065$, which is greater than zero and thus implies that there is inefficient entry – i.e., that product variety is too low relative to social optimum.

In column (2) of Table 5, we hold all other parameters constant and calculate the value of the love-of-variety parameter such that $\partial\widetilde{W}/\partial\log(J) = 0$; this leads to $\tilde{\Lambda}_0 = 0.066$, or roughly half of the actual estimate. In this case, the business-stealing effect and variety effect exactly cancel out, which zeroes out the first term in the formula above. As a result, the marginal excess burden collapses to the standard formula for excess burden under imperfect competition, with the quantity elasticity scaled by the mark-up, just as in Auerbach and Hines Jr (2002). This shows that our main formula provides an alternative way to calibrate the marginal excess burden without using the mark-up (since the mark-up is not one of the sufficient statistics in the full welfare formula). Instead, our formula requires a pass-through estimate and a love-of-variety estimate, along with the effect of taxes on variety. This formula is robust to allowing for the level of product variety to be governed by free entry, and we do not need to assume that variety is socially optimal.

Lastly, column (3) in Table 5 reports results when there is no love-of-variety. In this case, variety is clearly above the social optimum because the remaining model parameters imply business-stealing costs of additional variety (which are social costs), but no social benefits of additional variety. In this scenario, there is excessive entry, as in Besley (1989).

5.4 Comparing ad valorem taxes to specific taxes

The last way we use our model-based estimates is to compare the incidence and efficiency costs of existing ad valorem taxes to the incidence and efficiency costs of a *counterfactual* specific tax. That is, suppose a given amount of revenue is raised by a specific tax instead of an ad valorem tax. Would that generate larger changes in pre-tax prices? Larger efficiency costs? And how do these conclusions vary with the magnitude of the love-of-variety (holding other parameters constant)? In Appendix Table OA.5, we use the formulas in Propositions 1 and 3 to calculate the counterfactual effects of specific tax and compare it to our results for existing ad valorem taxes. Key to this analysis is that the tax formulas for specific and ad valorem taxes depend on the same model parameters, so that once we have our parameter estimates, we can simulate what a counterfactual specific tax would do to prices, quantity, and variety, and calculate the resulting incidence and welfare effects.

We conclude from this analysis that the love-of-variety we estimate is large enough to "flip" the standard result that ad valorem taxes lead to lower efficiency costs than specific taxes. We instead find the opposite, and we show how the variety response to taxes and consumers' love-of-variety are the key forces that "flip" this standard result. We also use simulate the pass-through of a counterfactual specific tax (in the "short run" without entry and exit, and in the "long run" with free entry), using the formulas in this paper and in Kroft et al. (2021). Although it is difficult to make direct policy recommendations from these stylized calculations, we think they illustrate that even a relatively modest love-of-variety is enough to overturn the standard efficiency cost ranking of specific versus ad valorem taxes,[43] and these results imply that ad valorem taxes may be suboptimal tax instruments when applied to markets with substantial product differentiation.

6 Conclusion

In this paper, we develop new formulas to study the efficiency cost and incidence of ad valorem and specific taxes. On the firm side, our framework allows for both price and quantity competition and entry and exit. On the consumer side, our framework incorporates love-of-variety preferences. Our formulas are stated in terms of the relative elasticities of demand and supply, the curvature of the firm's own demand, market conduct, and the causal effect of a change in variety on consumer surplus. To obtain estimates of these parameters, we derive expressions for the causal effects of taxes on consumer prices, quantity demand, and product variety in terms of the full set of model parameters. We then combine retail scanner data from the United States with variation in (ad valorem) sales taxes to estimate these three reduced-

[43]Holding all other parameters constant, the efficiency costs of specific and ad valorem taxes coincide when $\tilde{\Lambda}_0 = 0.076$.

form terms and find values of the model parameters that cause the expressions to match the reduced-form empirical estimates. We use the resulting parameter estimates to calibrate the marginal excess burden of ad valorem taxes, and we carry out additional calibrations to assess whether or not variety is socially optimal (at current tax rates) and to determine the efficiency costs and incidence of existing ad valorem taxes relative to a counterfactual specific tax.

The theory comparing ad valorem to specific taxes has been applied to tariffs as well as taxes, and so we believe our modeling and calibration approach should also be useful in international trade contexts. Our identification approach can be implemented using large data sets covering a very large number of distinct products (like many existing approaches in international trade), and we believe our approach has a unique advantage in transparently connecting the reduced-form empirical estimates of the effects on prices, variety, and total quantity demanded to the underlying model parameters. Additionally, by separating the love-of-variety preference parameter from the demand elasticity, the curvature of demand, and the market conduct parameters, we believe our identification approach is particularly clear and intuitive. The average markup pins down the conduct parameter given the demand elasticity. The pass-through estimate pins down the curvature of demand, which is identified freely from the demand elasticity (as in Bulow and Pfleiderer 1983). The variety effect can be recovered given estimates of the other parameters and the relative magnitude of the reduced-form effects of taxes on variety and the total quantity demanded.

There are several natural extensions to the theoretical and empirical analysis in this paper that we leave for future work. On the theory side, we have worked with a stylized model that assumes away many potentially important dimensions of heterogeneity across consumers, products, and firms. For example, we did not model multi-product firms as in Hamilton (2009). If multi-product firms can make multiple product "entry" decisions in the long run and are governed by the same long-run free-entry condition (across their portfolio of products), then we conjecture that our efficiency formulas will still hold in this more general model. We expect incidence to be much more complicated with multi-product firms, however, since this will require a model of imperfect competition with multi-product firms. We also focused on consumers' love-of-variety and how taxes affect the number of products available, but there is existing empirical evidence that taxes also affect the quality of products sold to consumers (see, e.g., Harding, Leibtag and Lovenheim 2012). It would be valuable to extend our framework to allow for endogenous quality.

On the empirical side, our research design compares outcomes across modules sold in stores located in different counties within a given county border pair. As a result, our empirical approach is not well-suited for estimation of product-specific effects of taxes and product-specific variety effects. We have left this heterogeneity analysis for future work, which may require additional data and a different research design. We also expect there to be substantial consumer heterogeneity in love-of-variety, and it would be useful to gather individual-level data to study this type of preference heterogeneity. This consumer-level data can also address some limitations in both our theoretical analysis and our calibrations regarding the extent of "offline-online" substitution (e.g., consumers shopping online when their most preferred variety is no longer available locally) as well as cross-border shopping (see, e.g., Baker, Johnson and Kueng 2021). It would be important to extend our theoretical model to account for these responses if they turn out to be important empirically.

We conclude with our belief that our framework may also be used to study particular commodity markets subjected to "sin taxes" (such as soda taxes) as well as products that are currently subjected to specific taxes (such as gasoline, cigarettes, and alcohol). In all of these cases, there are often policy considerations that are outside the scope of our analysis (e.g., addressing externalities and "internalities"); however, we conjecture that many of the economic trade-offs we highlight in this paper will still be relevant for these settings, as well. For example, if the social cost of the externality scales with the responsiveness of total output to the tax, then it is separable from the variety effect. As a result, extended versions of our formulas that account for these kind of situations (where taxed products generate externalities from consumption) are still likely to point towards the relative efficiency of ad valorem versus specific taxes depending on the strength of love-of-variety preferences. Overall, we conclude from both our theoretical results and calibrations that consumers' love-of-variety and the degree of product differentiation should inform how policymakers think about the optimal design of tax structure in a broad range of settings.

References

Ackerberg, Daniel A, and Marc Rysman. 2005. "Unobserved product differentiation in discrete-choice models: estimating price elasticities and welfare effects." *The Rand Journal of Economics*, 36(4): 771.

Adachi, Takanori, and Michal Fabinger. 2017. "Multi-Dimensional Pass-Through and Welfare Measures under Imperfect Competition." *Working Paper.*

Agrawal, David R. 2015. "The tax gradient: Spatial aspects of fiscal competition." *American Economic Journal: Economic Policy*, 7(2): 1–29.

Alessandria, George, and Horag Choi. 2007. "Do sunk costs of exporting matter for net export dynamics?" *The Quarterly Journal of Economics*, 122(1): 289–336.

Anderson, Simon P, Andre De Palma, and Brent Kreider. 2001*a*. "Tax incidence in differentiated product oligopoly." *Journal of Public Economics*, 81: 173–192.

Anderson, Simon P, Andre De Palma, and Brent Kreider. 2001*b*. "The efficiency of indirect taxes under imperfect competition." *Journal of Public Economics*, 81: 231–251.

Anderson, Simon P, André De Palma, Jacques-Francois Thisse, et al. 1987. "The CES is a discrete choice model?" *Economics Letters*, 24(2): 139–140.

Arkolakis, Costas, Arnaud Costinot, and Andrés Rodríguez-Clare. 2012. "New trade models, same old gains?" *American Economic Review*, 102(1): 94–130.

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

Auerbach, Alan J, and James R Hines Jr. 2002. "Taxation and economic efficiency." In *Handbook of Public Economics.* Vol. 3, 1347–1421. Elsevier.

Auerbach, Alan J., and James R. Jr. Hines. 2001. "Perfect taxation with Imperfect Competition." *NBER Working Paper.*

Bajari, Patrick, and C Lanier Benkard. 2003. "Discrete Choice Models as Structural Models of Demand: Some Economic Implications of Common Approaches."

Baker, Scott R, Stephanie Johnson, and Lorenz Kueng. 2021. "Shopping for lower sales tax rates." *American Economic Journal: Macroeconomics*, 13(3): 209–50.

Barzel, Yoram. 1976. "An Alternative Approach to the Analysis of Taxation." *Journal of Political Economy*, 84(6): 1177–97.

Benassy, Jean-Pascal. 1996. "Taste for variety and optimum production patterns in monopolistic competition." *Economics Letters*, 52(1): 41–47.

Berry, Steven, and Ariel Pakes. 2007. "The pure characteristics demand model." *International Economic Review*, 48(4): 1193–1225.

Besley, Timothy. 1989. "Commodity Taxation and Imperfect Competition: A Note on the Effects of Entry." *Journal of Public Economics*, 40: 359–367.

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

Boone, Christopher, Arindrajit Dube, Lucas Goodman, and Ethan Kaplan. 2016. "Unemployment Insurance Generosity and Aggregate Employment."

Brynjolfsson, Erik, Astrid A Dick, and Michael D Smith. 2010. "A nearly perfect market? Differentiation vs. price in consumer choice." *QME*, 8: 1–33.

Bulow, Jeremy I, and Paul Pfleiderer. 1983. "A note on the effect of cost changes on prices." *Journal of Political Economy*, 91(1): 182–185.

Cameron, A Colin, Jonah B Gelbach, and Douglas L Miller. 2011. "Robust Inference With Multiway Clustering." *Journal of Business & Economic Statistics*, 29(2): 238–249.

Cawley, John, David Frisvold, Anna Hill, and David Jones. 2020. "The Impact of the Philadelphia Beverage Tax on Prices and Product Availability." *Journal of Policy Analysis and Management*, 39(3): 605–628.

Chetty, Raj. 2009. "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods." *Annual Review of Economics*, 1(1): 451–488.

Chetty, Raj, Adam Looney, and Kory Kroft. 2009. "Salience and Taxation: Theory and Evidence." *American Economic Review*, 99(4): 1145–1177.

Cremer, Helmuth, and Jacques-Francois Thisse. 1994. "Commodity Taxation in a Differentiated Oligopoly." *International Economic Review*, 35(3): 613–633.

Delipalla, S. 1995. "Specific versus ad valorem taxation: empirical evidence from the European cigarette industry." *mimeo, University of Kent.*

Delipalla, Sofia, and Michael Keen. 1992. "The comparison between ad valorem and specific taxation under imperfect competition." *Journal of Public Economics*, 49: 351–367.

Delipalla, Sofia, and Michael Keen. 2006. "Product quality and the optimal structure of commodity taxes." *Journal of Public Economic Theory*, 8(4): 547–554.

Delipalla, Sophia, and Owen O'Donnell. 2001. "Estimating tax incidence, market power and market conduct: The European cigarette industry." *International Journal of Industrial Organization*, 19(6): 885–908.

DellaVigna, Stefano, and Matthew Gentzkow. 2019. "Uniform Pricing in U.S. Retail Chains." *Quarterly Journal of Economics*, 134: 2011–84.

Dhingra, Swati, and John Morrow. 2019. "Monopolistic competition and optimum product diversity under firm heterogeneity." *Journal of Political Economy*, 127(1): 196–232.

Dixit, Avinash K., and Joseph E. Stiglitz. 1977. "Monopolistic Competition and Optimum Product Diversity." *American Economic Review*, 67(3): 297–308.

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.

Einav, Liran, Dan Knoepfle, Jonathan Levin, and Neel Sundaresan. 2014*a*. "Sales Taxes and Internet Commerce." *American Economic Review*, 104(1): 1–26.

Einav, Liran, Dan Knoepfle, Jonathan Levin, and Neel Sundaresan. 2014*b*. "Sales taxes and internet commerce." *American Economic Review*, 104(1): 1–26.

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

Gillitzer, Christian, Henrik Jacobsen Kleven, and Joel Slemrod. 2017. "A Characteristics Approach to Optimal Taxation: Line Drawing and Tax-Driven Product Innovation." *Scandinavian Journal of Economics*, 119(2): 240–267.

Hamilton, Stephen F. 2009. "Excise taxes with multiproduct transactions." *American Economic Review*, 99(1): 458–71.

Harberger, Arnold C. 1964. "The Measurement of Waste." *American Economic Review*, 54(3): 58–76.

Harding, Matthew, Ephraim Leibtag, and Michael F Lovenheim. 2012. "The heterogeneous geographic and socioeconomic incidence of cigarette taxes: evidence from Nielsen homescan data." *American Economic Journal: Economic Policy*, 4(4): 169–98.

Hart, Oliver. 1982. "Reasonable Conjectures." *Working Paper*.

Holmes, Thomas J. 1998. "The effect of state policies on the location of manufacturing: evidence from state borders." *Journal of Political Economy*, 106(4): 667–705.

Hong, Han, and Matthew Shum. 2006. "Using price distributions to estimate search costs." *The RAND Journal of Economics*, 37(2): 257–275.

Johnson, Terry R. 1978. "Additional Evidence on the Effects of Alternative Taxes on Cigarette Prices." *Journal of Political Economy*, 86(2): 325–8.

Kay, John A, and Michael J Keen. 1983. "How should commodities be taxed?: Market structure, product heterogeneity and the optimal structure of commodity taxes." *European Economic Review*, 23(3): 339–358.

Keen, Michael. 1998. "The balance between specific and ad valorem taxation." *Fiscal Studies*, 19(1): 1–37.

Kroft, Kory, Jean-William Laliberté, René Leal Vizcaíno, and Matthew Notowidigdo. 2020*a*. "Salience and Taxation with Imperfect Competition." *NBER Working Paper*.

Kroft, Kory, Jean-William P. Laliberté, René Leal-Vizcaíno, and Matthew J. Notowidigdo. 2020*b*. "A New Empirical Method for Valuing Product Variety." *Working Paper*.

Kroft, Kory, Jean-William P Laliberté, René Leal Vizcaíno, and Matthew J Notowidigdo. 2021. "Efficiency and Incidence of Taxation with Free Entry and Love-of-Variety Preferences." National Bureau of Economic Research.

Kroft, Kory, René Leal-Vizcaíno, Matthew J. Notowidigdo, and Ting Wang. 2020*c*. "Parallel Inverse Aggregate Demand Curves in Discrete Choice Models." *NBER Working Paper*.

Mankiw, N Gregory, and Michael D Whinston. 1986. "Free Entry and Social Inefficiency." *The RAND Journal of Economics, Spring*, 17(1): 48–58.

Melitz, Marc, and Gianmarco I. P. Ottaviano. 2008. "Market Size, Trade, and Productivity." *The Review of Economic Studies*, 75: 295–316.

Parenti, Mathieu, Philip Ushchev, and Jacques-François Thisse. 2017. "Toward a theory of monopolistic competition." *Journal of Economic Theory*, 167: 86–115.

Petrin, Amil. 2002. "Quantifying the benefits of new products: The case of the minivan." *Journal of Political Economy*, 110(4): 705–729.

Pirttilä, J. 1997. "Aspects of environmental taxation under imperfect competition." *mimeo, University of Essex.*

Poterba, James M. 1996. "Retail price reactions to changes in state and local sales taxes." *National Tax Journal*, 49(2): 169–179.

Riordan, Michael H. 1985. "Imperfect Information and Dynamic Conjectural Variations." *The RAND Journal of Economics*, 16(1): 41–50.

Seade, Jesus. 1980. "On the Effects of Entry." *Econometrica*, 48(2): 479–489.

Spence, Michael. 1976. "Product Differentiation and Welfare." *Source: The American Economic Review*, 66(2): 407–414.

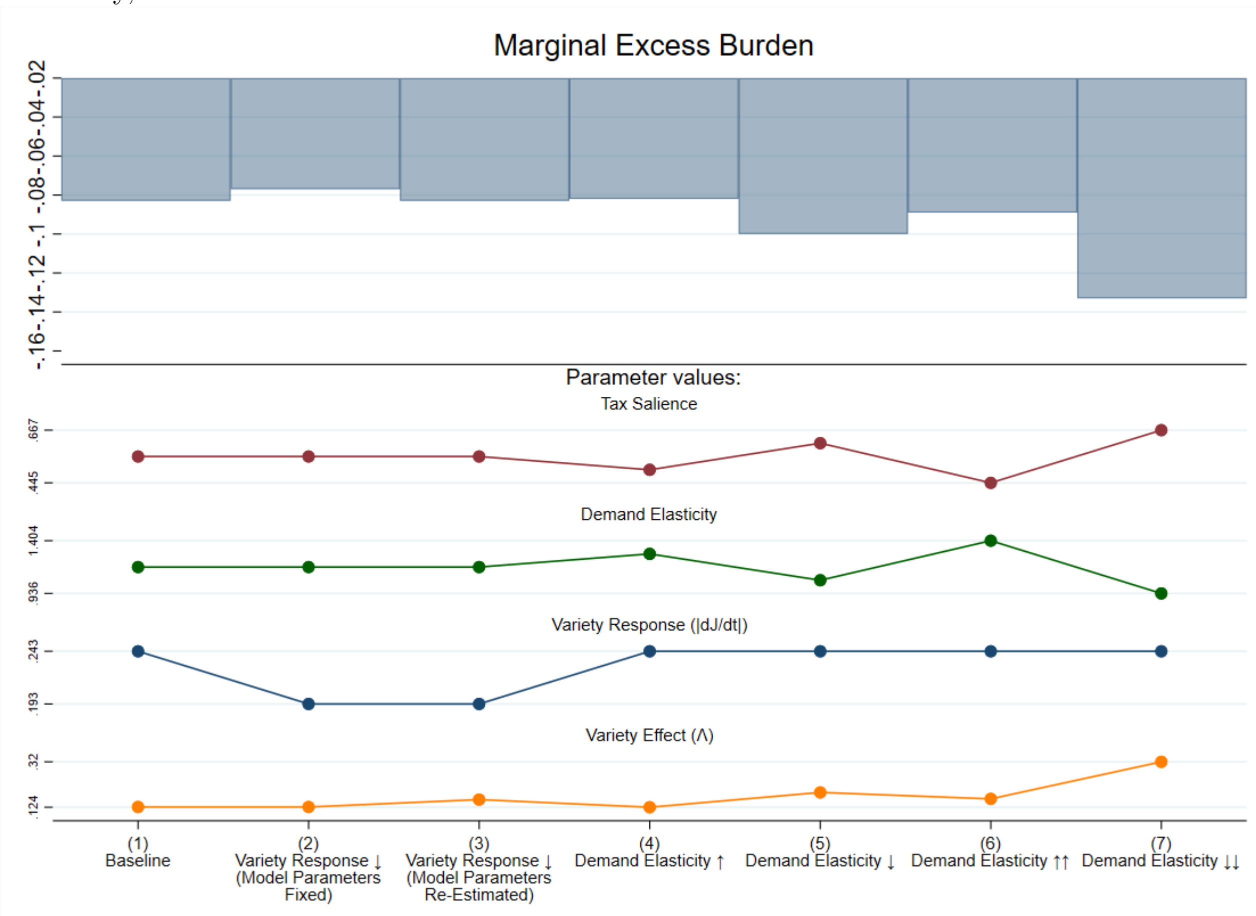
Taubinsky, Dmitry, and Alex Rees-Jones. 2018. "Attention variation and welfare: Theory and evidence from a tax salience experiment." *Review of Economic Studies*, 85(4): 2462–2496.

Trajtenberg, Manuel. 1990. *Economic analysis of product innovation: The case of CT scanners.* Harvard University Press.

Vives, Xavier. 2001. *Oligopoly Pricing: Old Ideas and New Tools.* MIT Press.

Weyl, E. Glen, and Michal Fabinger. 2013. "Pass-through as an economic tool: Principles of incidence under imperfect competition." *Journal of Political Economy*, 121(3): 528–583.

Figure 1: Sensitivity of Calibration Results to Alternative Values of Variety Response, Demand Elasticity, and Tax Salience Parameters



Notes: This figure reports sensitivity to different assumptions on the variety response, the demand elasticity and the tax salience parameter. Scenarios (2) and (3) use the alternative variety response to taxes, while scenarios (4) through (7) vary both the demand elasticity and tax salience parameters but hold the product of the tax salience parameter and demand elasticity constant in order to ensure that $d\log(Q)/d\log(1+\tau)$ is constant.

Table 1: Summary statistics

	Full sample		Above-median residual tax rate		Below-median residual tax rate	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Tax rate, τ_{mcn}	0.034	0.032	0.041	0.032	0.028	0.031
Key variables in reduced-form analysis (residualized)						
log Prices	0.004	0.055	0.005	0.055	0.004	0.055
log Quantity	10.240	0.396	10.235	0.394	10.245	0.398
log Product Variety	4.241	0.284	4.239	0.280	4.244	0.288
Sample size statistics:						
N (module-store-year observations)	11,624,918		5,811,227		5,813,691	
N (stores)	3,822		3,808		3,822	
N (modules)	198		198		198	
N (years)	9		9		9	
N (counties)	543		538		543	
N (border pairs)	497		494		497	

Notes: The sample is derived from the Nielsen Retail Scanner data covering the years 2006-2014 and is restricted to modules above the 80th percentile of the national distribution of sales. Sales tax rates are measured annually based on the rates that were effective on September 1. Prices, quantity, and variety are measured yearly. Median tax rates are calculated separately for each module-year-border pair cell. All outcomes are entered in logs, and the price variable is normalized to be mean zero in the broader sample of stores that include non-border counties. Statistics reported in this table are calculated for our border-county subsample only. See main text for more details.

Table 2: Effect of Sales Taxes on Prices, Quantity, and Product Variety

Dependent Variable:	Prices	Quantity	Variety
	(1)	(2)	(3)
Panel A: County Border Pair OLS Estimates			
$\log(1+\tau_{mcn})$	0.038	-0.677	-0.236
	(0.016)	(0.154)	(0.074)
Panel B: 2SLS Estimates Using State-Level Tax Rate as Instrument			
$\log(1+\tau_{mcn})$	0.039	-0.731	-0.243
	(0.016)	(0.153)	(0.075)
Specification:			
Store fixed effects	y	y	y
Module × County Border Pair fixed effects	y	y	y

Notes: The sample is derived from the Nielsen Retail Scanner data covering the years 2006-2014 and is restricted to modules above the 80th percentile of the national distribution of sales. Sales tax rates are measured annually based on the rates that were effective on September 1. Sales, prices, and variety are measured yearly. All reported coefficients are simple averages of nine estimated coefficients -- one for each year from 2006 to 2014. The sample is restricted to border counties and observations are weighted by the inverse of number of pairs a store belongs to. Standard errors are clustered two-way at the state-module level and at the border pair by module level. In panel B, the tax rate is instrumented with the state-level, leave-county-out, average tax rate.

Table 3: Robustness to Alternative Measure of Product Variety

Dependent variable:	Product varirety measure Number of products [Baseline] (1)	 Share-weighted sum of products (2)
Panel A: County Border Pair OLS Estimates		
$\log(1 + \tau_{mcn})$	-0.236	-0.172
	(0.074)	(0.063)
Panel B: 2SLS Estimates Using State-Level Tax Rate as Instrument		
$\log(1 + \tau_{mcn})$	-0.243	-0.193
	(0.075)	(0.064)
Specification:		
Store fixed effects	y	y
Module × County Border Pair fixed effects	y	y

Notes: Sales tax rates are measured annually based on the rates that were effective on September 1. Sales, prices, and variety are measured yearly. The Retail Scanner data is restricted to modules above the 80th percentile of the national distribution of sales. All reported coefficients are simple averages of nine estimated coefficients -- one for each year from 2006 to 2014. The sample is restricted to border counties and observations are weighted by the inverse of number of pairs a store belongs to. The first column reports results using the baseline measure of product variety: the number of products (i.e., number of distinct UPC codes) available during the year. The second column reports an alternative variety measure that is the weighted number of products available using each product's national (module-level) market share as a summation weight. As a result, if the varieties exiting have lower-than-average national market shares, this measure will have a smaller magnitude. Standard errors are clustered two-way at the state-module level and at the border pair by module level. In panel B, the tax rate is instrumented with the state-level, leave-county-out, average tax rate.

Table 4: Estimating Variety Effect and Calibrating Welfare Formulas

Panel A: Calibrated parameters	
Average tax rate, τ_0	0.034
Tax salience parameter, θ_τ	0.556
Demand elasticity, ϵ_D	1.170
Panel B: Reduced-form estimates	
Pass-through of taxes into pre-tax prices, $d\log(p)/d\log(1+\tau)$	0.039
	(0.016)
Quantity response, $d\log(Q)/d\log(1+\tau)$	-0.731
	(0.153)
Variety response, $d\log(J)/d\log(1+\tau)$	-0.243
	(0.075)
Panel C: Model parameters estimated by matching reduced-form estimates	
Markup, $(p - c'(q))/p$	0.080
	(0.042)
Implied conduct parameter, v_q/J	0.092
	(0.049)
Inverse elasticity of marginal surplus, ϵ_{ms}	-0.903
	(0.189)
Variety effect parameter, $\tilde{\Lambda}_0$	0.125
	(0.339)
Panel D: Calibrated welfare formulas	
Full marginal excess burden (MEB) formula, $d\tilde{W}/d\tau$	-0.083
	(0.108)
Alternative MEB formula benchmarks:	
Harberger (1964) / Chetty-Looney-Kroft (2009) benchmark, $\theta_\tau * \tau_0 * d\log(Q)/d\log(1+\tau)$	-0.014
	(0.003)
Besley (1989)-style benchmark; i.e., full MEB formula with $\tilde{\Lambda}_0 = 0$	-0.053
	(0.019)
% difference between full formula and Besley (1989)-style benchmark	57.5%

Notes: This table reports structural parameter estimates by finding parameters that allow the model to match the reduced-form estimates. The model parameters in Panel C are estimated by matching the reduced-form estimates of effects of taxes on prices, quantity, and variety by choosing the variety effect parameter, the inverse elasticity of marginal surplus, and the markup. These parameters can then be used to calibrate the main welfare formula. The final rows show the effect of taxes on welfare using the main welfare formula, and compare the results from the main formula with benchmarks from Harberger/Chetty-Looney-Kroft and Besley (1989). Standard errors in Panels C and D are based on 100 bootstrap replications of the 3 reduced-form parameters. We solve for the variety effect parameter, the inverse elasticity of marginal surplus, the markup and the conduct parameter for each of these replications, constraining the variety effect to be non-negative. 3 bootstrap relications are dropped because the stability conditions is violated in those cases. See text for details.

Table 5: Using Calibrations to Determine Whether Variety is Socially Optimal

	Baseline	Counterfactual scenarios	
	(1)	(2)	(3)
Panel A: Variety effect parameter, $\tilde{\Lambda}_0$			
Estimated variety effect parameter	0.125	0.066	0.000
Panel B: Socially optimal variety calculations			
$\partial \log(q)/\partial \log(J)$	-0.751	-0.827	-0.911
Business-stealing effect, $\partial \log(q)/\partial \log(J) * (p - c'(q))/p$	-0.060	-0.066	-0.073
$\partial \tilde{W}/\partial \log(J)$ = Variety effect ($\tilde{\Lambda}_0$) + business-stealing effect (>0 implies that variety is below the social optimum)	0.065	0.000	-0.073

Notes: This table reports results using the parameter estimates from Table 4 to calibrate whether or not variety is above or below the social optimum. Column (1) uses the baseline estimate of the variety effect and the other parameters in Table 4, while columns (2) and (3) report results using other values of the variety effect but hold other parameters constant. Column (2) finds the exact value such that the business-stealing effect and the variety effect are equal (so that variety is socially optimal at current tax rates), and column (3) sets the variety effect to zero so that consumers have no love of variety.

Online Appendix for "Efficiency and Incidence of Taxation with Free Entry and Love-of-Variety Preferences"

A Specific Taxation and Ad Valorem and Results

Proof. ***Marginal Excess Burden Formula for specific tax*** $\frac{dW}{dt}$**.**

Let the total welfare to be the sum of consumer surplus, profits and government tax revenues.

$$W(p(t), t, J(t)) = \underbrace{u(Q_L(t), J(t)) - (p(t)(1+\tau) + t)Q_L(t)}_{CS}$$
$$+ \underbrace{p(t)Q_L(t) - J(t)c\left(\frac{Q_L(t)}{J(t)}\right) - J(t)F}_{J\pi} + \underbrace{tQ_L(t) + p(t)\tau Q}_{R}$$

By totally differentiating $W_L(t) = W(p(t), t, J(t))$ with respect to t (and keeping τ constant) we obtain

$$\begin{aligned} \frac{dW_L}{dt} &= \quad \frac{\partial u}{\partial Q}(Q_0, J_0) - c'(q_0)\Bigg) \frac{dQ_L}{dt} + \quad \frac{\partial u}{\partial J}(Q_0, J_0) - c(q_0) - F + q_0 c'(q_0)\Bigg) \frac{dJ}{dt} \\ &= (p_0(1 + \theta_\tau \tau_0) + \theta_t t_0 - c'(q_0)) \frac{dQ_L}{dt} + (\Lambda_0 + \pi_0 - [p_0 - c'(q_0)] * q_0) \frac{dJ}{dt} \end{aligned} \tag{1}$$

where we used the first-order approximation from Chetty, Looney and Kroft (2009) $\frac{\partial u}{\partial J}(Q_0, J_0) = p_0(1 + \theta_\tau \tau_0) + \theta_t t_0$, we used our definition of variety effect $\Lambda_0 = \frac{\partial u}{\partial J}(Q_0, J_0)$ and profits $\pi_0 = p_0 q_0 - c(q_0) - F$. When $t_0 = 0$, $p_0 = c'(q^*)$ and $\Lambda_0 = -\pi_0$, we get $\frac{dW_L}{dt} = 0$ which is the first-best outcome. $\square$

Proof. ***Lemma 1.***

Let $\pi = pq - c(q) = 0$ be the free-entry condition of firms. When τ is constant, then $\frac{d\pi}{dt} = 0$ implies that$(p - mc)\frac{dq}{dt} = -q\frac{dp}{dt}$ and so $\frac{p-mc}{p} = -\frac{q/t}{p/t}\frac{\frac{dp}{dt}}{\frac{dq}{dt}}$.

If t is now constant, then $\frac{d\pi}{d\tau} = 0$ implies $(p - mc)\frac{dq}{d\tau} = -q\frac{dp}{d\tau}$ and so $\frac{p-mc}{p} = -\frac{q/\tau}{p/\tau}\frac{\frac{dp}{d\tau}}{\frac{dq}{d\tau}}$. $\square$

Proof. ***Proposition 1.***

Let $\Delta = \left[2 - \frac{\nu_q}{J} + \frac{\epsilon_D^* - \frac{\nu_q}{J}}{\epsilon_S \frac{\nu_q}{J}} + \frac{\frac{\nu_q}{J}}{\epsilon_{ms}}\right] - \frac{\epsilon_D \Lambda}{(p(1+\tau)+t)q}\left(1 + \frac{\epsilon_D^* - \frac{\nu_q}{J}}{\epsilon_S \frac{\nu_q}{J}} + \frac{1}{\epsilon_{ms}}\right) + \left(1 - \frac{\nu_q}{J}\right)\epsilon_D \frac{JQ}{p(1+\tau)+t}\frac{\partial^2 P}{\partial J \partial Q}$.[1] The firm stability conditions $\frac{\partial^2 \pi_j}{\partial p_j^2} < 0$ and $\frac{\partial \pi_j}{\partial J} < 0$, are respectively equivalent to $1 + \frac{\epsilon_D^* - \frac{\nu_q}{J}}{\epsilon_S} + \frac{\frac{\nu_q}{J}}{\epsilon_{ms}} > 0$ and $\Delta > 0$, where $\epsilon_D^* = \frac{p(1+\theta_\tau \tau)}{p(1+\tau)+t}\epsilon_D$. Here, Δ and ϵ_D^* are written in the general form that depends on both the specific tax rate t and the ad valorem tax rate τ.

By Lemma 1, we have $\frac{dPS}{dt} = 0$. Therefore substituting this into equation (1) we obtain:

$$\frac{dW}{dt} = \Lambda_0 \frac{dJ}{dt} - Q_0 \frac{dp}{dt} + (\theta_t t_0 + p_0 \theta_\tau \tau_0)\frac{dQ_L}{dt}$$

From the behavioral equation of consumers $wtp(Q,J) \equiv P(Q,J) = p(1+\theta_\tau \tau) + \theta_t t$, we have

$$mwtp(Q,J)\frac{dQ}{dt} + \frac{\partial P}{\partial J}\frac{dJ}{dt} = \frac{dp}{dt}(1+\theta_\tau \tau) + \theta_t \tag{2}$$

In addition, from the free-entry condition, $(p - mc)\frac{dq}{dt} = -q\frac{dp}{dt}$, and firm's first-order condition, $p - mc = -mwtp(Q,J)Q\frac{\nu_q}{J(1+\theta_\tau \tau)}$, we have

$$mwtp(Q,J)\nu_q \frac{dq}{dt} = (1+\theta_\tau \tau)\frac{dp}{dt} \tag{3}$$

Combining this with the behavioral equation above, and letting $mwtp(Q,J) = mwtp(Q)$ for simplicity, we have

$$\begin{aligned} mwtp(Q)\nu_q \frac{dq}{dt} &= mwtp(Q)\frac{dQ}{dt} + \frac{\partial P}{\partial J}\frac{dJ}{dt} - \theta_t \\ &= mwtp(Q)\left(J\frac{dq}{dt} + q\frac{dJ}{dt}\right) + \frac{\partial P}{\partial J}\frac{dJ}{dt} - \theta_t \end{aligned} \tag{4}$$

where the second line follows from substituting $\frac{dQ}{dt} = J\frac{dq}{dt} + q\frac{dJ}{dt}$. Therefore,

$$\frac{dq}{dt} = \frac{\theta_t - \left(\frac{\partial P}{\partial J} + q * mwtp(Q)\right)\frac{dJ}{dt}}{mwtp(Q)(J - \nu_q)} \tag{5}$$

[1] This becomes $\Delta = \left[2 - \frac{\nu_q}{J} + \frac{\epsilon_D^* - \frac{\nu_q}{J}}{\epsilon_S \frac{\nu_q}{J}} + \frac{\frac{\nu_q}{J}}{\epsilon_{ms}}\right] - \frac{\Lambda \epsilon_D}{(p(1+\tau)+t)q}\left(1 + \frac{\epsilon_D^* - \frac{\nu_q}{J}}{\epsilon_S \frac{\nu_q}{J}} + \frac{1}{\epsilon_{ms}}\right)$ under parallel demands.

Using now $\frac{dq}{dt} = \frac{\partial q}{\partial t} + \frac{\partial q}{\partial J}\frac{dJ}{dt}$ (note that $\frac{\partial q}{\partial t} = \frac{dq}{dt}\Big|_J$) we can get

$$\frac{dJ}{dt} = \frac{\theta_t - (J - \nu_q)mwtp(Q)\frac{\partial q}{\partial t}}{\frac{\partial P}{\partial J} + q * mwtp(Q) + (J - \nu_q)mwtp(Q)\frac{\partial q}{\partial J}} \tag{6}$$

From Kroft et al. (2020), we have

$$\frac{\partial q}{\partial t} = \frac{dq}{dt}\Big|_J = \frac{1}{Jmwtp(Q)}\left(\rho_t^{SR} + \theta_t - 1\right) = \frac{\omega_t^{SR}\theta_t}{Jmwtp(Q)} \tag{7}$$

where $\rho_t^{SR} = 1 - (1 - \omega_{SR})\theta_t$ and $\omega_{SR} = \frac{1}{1 + \frac{\epsilon_D^* - \frac{\nu_q}{J}}{\epsilon_S} + \frac{\frac{\nu_q}{J}}{\epsilon_{ms}}}$, where $\epsilon_D^* = \frac{p(1+\theta_\tau\tau)}{p(1+\tau)+t}\epsilon_D$ (short-run passthrough is taken from Kroft et al. (2020)).

Finally, fix t, and differentiate the first-order condition $(p-mc)(1+\theta_\tau\tau)+mwtp(Q,J)Q\frac{\nu_q}{J} = wtp(Q,J) - \theta_t t - mc(1+\theta_\tau\tau) + mwtp(Q,J)Q\frac{\nu_q}{J} = 0$ with respect to J to get:

$$\frac{\partial P}{\partial J} + mwtp(Q)\left(q + J\frac{\partial q}{\partial J}\right) - c''(q)(1+\theta_\tau\tau)\frac{\partial q}{\partial J} + \frac{\partial q}{\partial J}mwtp(Q)\nu_q + q\nu_q mwtp'(Q)\left(q + J\frac{\partial q}{\partial J}\right) + \frac{\partial^2 P}{\partial J\partial Q}q\nu_q = 0$$

where we have assumed that $\frac{\partial \nu}{\partial J} = 0$. Further simplifying yields:

$$\frac{\partial q}{\partial J} = -\frac{\frac{\partial P}{\partial J} + \frac{\partial^2 P}{\partial J\partial Q}q\nu_q + mwtp(Q)q + q^2\nu_q mwtp'(Q)}{(J+\nu_q)mwtp(Q) - c''(q)(1+\theta_\tau\tau) + Jq\nu_q mwtp'(Q)} \tag{8}$$

Rearranging equation (8), the denominator is equal to $J * mwtp(Q) * \left(1 + \frac{\epsilon_D^* - \frac{\nu_q}{J}}{\epsilon_S} + \frac{\frac{\nu_q}{J}}{\epsilon_{ms}}\right)$, and so we get:

$$\frac{\partial q}{\partial J} = -\frac{\omega_{SR}}{J * mwtp(Q)}\left(\frac{\partial P}{\partial J} + \frac{\nu_q}{J}Q\frac{\partial^2 P}{\partial J\partial Q}\right) - \frac{q}{J}\omega_{SR}\left(1 - \frac{\nu_q}{J} + \frac{\frac{\nu_q}{J}}{\epsilon_{ms}}\right) \tag{9}$$

Note:

$$\begin{aligned}\omega_{SR} * \frac{\nu_q}{J} * q * mwtp(Q) * \Delta = {} & \frac{\partial P}{\partial J}\left(1 - \omega_{SR}\left(1 - \frac{\nu_q}{J}\right)\right) - \omega_{SR}\frac{\nu_q}{J}\left(1 - \frac{\nu_q}{J}\right)\frac{\partial^2 P}{\partial J\partial Q}Q \\ & + q * mwtp(Q)\left(1 - \omega_{SR}\left(1 - \frac{\nu_q}{J}\right)\left(1 - \frac{\nu_q}{J} + \frac{\frac{\nu_q}{J}}{\epsilon_{ms}}\right)\right)\end{aligned}$$

Substituting equation (9) and equation (7) into equation (6), we get:

$$\frac{dJ}{dt} = \theta_t \left(\frac{1 - \omega_{SR}\left(1 - \frac{\nu_q}{J}\right)}{\omega_{SR}\frac{\nu_q}{J}\Delta} \right)$$

and substituting $\frac{dJ}{dt}$ into equation (5), we obtain:

$$\frac{dq}{dt} = \frac{\theta_t}{J * mwtp(Q)} \left(\frac{1 - \frac{1}{\epsilon_{ms}}}{\Delta} \right)$$

Finally, from equation (3) and the expression for $\frac{dq}{dt}$ we have:

$$\begin{aligned}\rho_t &= 1 + mwtp(Q, J)\nu_q \frac{dq}{dt} \\ &= \frac{\Delta + \frac{\nu_q}{J}\theta_t \left(1 - \frac{1}{\epsilon_{ms}}\right)}{\Delta}\end{aligned}$$

□

Proof. ***Corollary 1.***

The proof is immediate by setting $\theta_t = \theta_\tau = 1$, $\Lambda_0 = 0$ and $t_0 = \tau_0 = 0$ into the conditions of Proposition 1. □

Proof. ***Proposition 2***

Consider a change in the tax from τ_0 to τ_1. A first-order approximation to the marginal excess burden of taxation is:

$$\frac{dW}{d\tau} = \underbrace{(p_0(1 + \theta_\tau \tau_0) + \theta_t t_0 - c'(q_0))\frac{dQ_L}{d\tau}}_{\text{Quantity effect}} + \underbrace{(\Lambda_0 + \pi_0 - [p_0 - c'(q_0)] * q_0)\frac{dJ}{d\tau}}_{\text{Diversity effect}} \tag{10}$$

Under Lemma 1, the marginal excess burden of taxation is given by:

$$\frac{dW}{d\tau} = \Lambda_0 \frac{dJ}{d\tau} - Q_0 \frac{dp}{d\tau} + (\theta_t t_0 + p_0 \theta_\tau \tau_0) \frac{dQ_L}{d\tau} \tag{11}$$

Willingness-to-pay with ad valorem taxes takes the form $wtp(Q) = p(1+\theta_\tau \tau)$, so $mwtp(Q)\frac{dQ}{d\tau} + \frac{\partial P}{\partial J}\frac{dJ}{d\tau} = \frac{dp}{d\tau}(1+\theta_\tau \tau) + p\theta_\tau$. We have the free entry-condition $(p - mc)\frac{dq}{d\tau} = -q\frac{dp}{d\tau}$, and the firm's first-order condition $p - mc = -\frac{\nu_q}{J(1+\theta_\tau \tau)} mwtp(Q)Q$. Therefore, we have:

$$\nu_q * mwtp(Q)\frac{dq}{d\tau} = (1+\theta_\tau \tau)\frac{dp}{d\tau} \tag{12}$$

which implies:

$$\frac{dq}{d\tau} = \frac{p\theta_\tau - \left(\frac{\partial P}{\partial J} + q * mwtp(Q)\right)\frac{dJ}{d\tau}}{mwtp(Q)\left(1 - \frac{\nu_q}{J}\right)} \tag{13}$$

Using now $\frac{dq}{d\tau} = \frac{\partial q}{\partial \tau} + \frac{\partial q}{\partial J}\frac{dJ}{d\tau}$ (Here $\frac{\partial q}{\partial \tau} = \frac{dq}{d\tau}\Big|_J$). we get

$$\frac{dJ}{d\tau} = \frac{p\theta_\tau + \left(\nu_q - J\right) mwtp(Q)\frac{\partial q}{\partial \tau}}{\frac{\partial P}{\partial J} + q * mwtp(Q) + \left(J - \nu_q\right)\frac{\partial q}{\partial J}} \tag{14}$$

We also have

$$\frac{\partial q}{\partial \tau} = \frac{dq}{d\tau}\Big|_J = \frac{1}{Jmwtp(Q)}\left(\theta_\tau mc * \omega_{SR}\right)$$

where $\rho_\tau^{SR} = 1 - \left(1 - \omega_{SR}\frac{mc}{p}\right)\theta_\tau$ and $\omega_{SR} = \frac{1}{1 + \frac{\epsilon_D^* - \frac{\nu_q}{J}}{\epsilon_S} + \frac{\frac{\nu_q}{J}}{\epsilon_{ms}}}$. Moreover,

$$\frac{\partial q}{\partial J} = -\frac{\omega_{SR}}{J * mwtp(Q)}\left(\frac{\partial P}{\partial J} + \frac{\partial^2 P}{\partial J \partial Q} q\nu_q\right) - \frac{q\omega_{SR}}{J}\left(1 - \frac{\nu_q}{J} + \frac{\frac{\nu_q}{J}}{\epsilon_{ms}}\right) \tag{15}$$

Therefore, substituting $\frac{\partial q}{\partial \tau}$ and $\frac{\partial q}{\partial J}$ into equation (14)we have

$$\begin{aligned}
\frac{dJ}{d\tau} &= \theta_\tau\left(\frac{p - mc * \omega_{SR}\left(1 - \frac{\nu_q}{J}\right)}{\omega_{SR} * \frac{\nu_q}{J} * q * mwtp(Q) * \Delta}\right) \\
&= p\theta_\tau\left(\frac{\left(1 + \frac{\epsilon_D^* - \frac{\nu_q}{J}}{\epsilon_S} + \frac{\frac{\nu_q}{J}}{\epsilon_{ms}}\right) - \left(1 - \frac{\frac{\nu_q}{J}}{\epsilon_D^*}\right)\left(1 - \frac{\nu_q}{J}\right)}{\frac{\nu_q}{J} * q * mwtp(Q) * \Delta}\right) \\
&= -\frac{p\theta_\tau J\epsilon_D}{p(1+\tau)+t}\left(\frac{1 + \frac{1}{\epsilon_D^*} - \frac{\frac{\nu_q}{J}}{\epsilon_D^*} + \frac{\epsilon_D^* - \frac{\nu_q}{J}}{\epsilon_S \frac{\nu_q}{J}} + \frac{1}{\epsilon_{ms}}}{\Delta}\right)
\end{aligned} \tag{16}$$

Recall $\Delta = \left[2 - \frac{\nu_q}{J} + \frac{\epsilon_D^* - \frac{\nu_q}{J}}{\epsilon_S \frac{\nu_q}{J}} + \frac{\frac{\nu_q}{J}}{\epsilon_{ms}}\right] - \frac{\epsilon_D J \frac{\partial P}{\partial J}}{(p(1+\tau)+t)} \left(1 + \frac{\epsilon_D^* - \frac{\nu_q}{J}}{\epsilon_S \frac{\nu_q}{J}} + \frac{1}{\epsilon_{ms}}\right) + \left(1 - \frac{\nu_q}{J}\right) \epsilon_D \frac{JQ}{p(1+\tau)+t} \frac{\partial^2 P}{\partial J \partial Q}$.
Substituting equation (16) into equation (13), then:

$$\begin{aligned}
\frac{dq}{d\tau} &= \frac{-\theta_\tau \omega_{SR}}{Jmwtp(Q)} \left(\frac{\frac{\partial P}{\partial J}(p - mc) + q * mwtp(Q)\left(p\left(1 - \frac{\nu_q}{J} + \frac{\frac{\nu_q}{J}}{\epsilon_{ms}}\right) - mc\right)}{\omega_{SR} \frac{\nu_q}{J} \Delta} \right) \\
&= \frac{-p\theta_\tau}{Jmwtp(Q)} \left(\frac{\frac{\frac{\nu_q}{J}}{\epsilon_D^*} - \frac{\nu_q}{J} + \frac{\frac{\nu_q}{J}}{\epsilon_{ms}} - \frac{J \frac{\partial P}{\partial J} \epsilon_D}{((1+\tau)p+t)} \frac{\frac{\nu_q}{J}}{\epsilon_D^*}}{\frac{\nu_q}{J} \Delta} \right)
\end{aligned}$$

Finally,

$$\begin{aligned}
\rho_\tau &= \frac{1}{p} \frac{1+\tau}{1+\theta_\tau \tau} \nu_q mwtp(Q) \frac{dq}{d\tau} + 1 \\
&= -\frac{\nu_q}{J} \frac{\theta_\tau (1+\tau)}{(1+\theta_\tau \tau)} \left(\frac{\frac{\partial P}{\partial J}\left(\frac{p-mc}{p}\right) + q * mwtp(Q)\left(\frac{p-mc}{p} - \frac{\nu_q}{J} + \frac{\frac{\nu_q}{J}}{\epsilon_{ms}}\right)}{\frac{\nu_q}{J} \Delta} \right) + 1 \\
&= \frac{\frac{\nu_q}{J} \Delta - \frac{\nu_q}{J} \frac{\theta_\tau(1+\tau)}{(1+\theta_\tau \tau)} \left(\frac{p-mc}{p} - \frac{\nu_q}{J} + \frac{\frac{\nu_q}{J}}{\epsilon_{ms}}\right) + \frac{J \frac{\partial P}{\partial J} \epsilon_D}{((1+\tau)p+t)} \left(\frac{\nu_q}{J} \frac{\theta_\tau(1+\tau)}{(1+\theta_\tau \tau)} \frac{p-mc}{p}\right)}{\frac{\nu_q}{J} \Delta}
\end{aligned}$$

Using $\frac{p-mc}{p} = \frac{\frac{\nu_q}{J}}{\epsilon_D^*}$, we obtain:

$$\rho_\tau = \frac{\Delta + \frac{\nu_q}{J} \frac{\theta_\tau(1+\tau)}{(1+\theta_\tau \tau)} \left[1 - \frac{1}{\epsilon_D^*} - \frac{1}{\epsilon_{ms}} + \frac{J \frac{\partial P}{\partial J}}{(1+\tau)p+t} \left(\frac{\epsilon_D}{\epsilon_D^*}\right)\right]}{\Delta}$$

$\square$

Derivation of Δ

Let $\Delta = \left[2 - \frac{\nu_q}{J} + \frac{\epsilon_D^* - \frac{\nu_q}{J}}{\epsilon_S \frac{\nu_q}{J}} + \frac{\frac{\nu_q}{J}}{\epsilon_{ms}}\right] - \left(1 + \frac{\epsilon_D^* - \frac{\nu_q}{J}}{\epsilon_S \frac{\nu_q}{J}} + \frac{1}{\epsilon_{ms}}\right) \epsilon_D \frac{J}{p(1+\tau)+t} \frac{\partial P}{\partial J} + \left(1 - \frac{\nu_q}{J}\right) \epsilon_D \frac{JQ}{p(1+\tau)+t} \frac{\partial^2 P}{\partial J \partial Q}$, we show that $\Delta = -\frac{J\epsilon_D^*}{pq} \frac{\partial \pi}{\partial J}$.

Proof. The effect of taxes on entry is derived by using the implicit function theorem on the long-run entry condition $\pi(q(J,t,\tau_0), J, t, \tau_0) = 0$, and the first-order condition of the firm $\frac{\partial \pi}{\partial q} = 0$, so that $\frac{dJ}{d\tau} = -\frac{\frac{\partial \pi}{\partial \tau}}{\frac{\partial \pi}{\partial J}}$. Therefore

$$\begin{aligned}
\frac{\partial \pi}{\partial J} &= -\frac{\frac{\partial \pi}{\partial t}}{\frac{dJ}{dt}} \\
&= \frac{\frac{p\theta_\tau q - mwtp(Q)*Q*\left(1-\frac{\nu_q}{J}\right)\frac{\partial q}{\partial \tau}}{1+\theta_\tau \tau}}{\frac{p\theta_\tau - mwtp(Q)*J*\left(1-\frac{\nu_q}{J}\right)\frac{\partial q}{\partial \tau}}{\frac{\partial P}{\partial J} + q*mwtp(Q) + mwtp(Q)*J*\left(1-\frac{\nu_q}{J}\right)\frac{\partial q}{\partial J}}} \\
&= \frac{q}{1+\theta_\tau \tau}\left(\frac{\partial P}{\partial J} + q*mwtp(Q) + mwtp(Q)*J*\left(1-\frac{\nu_q}{J}\right)\frac{\partial q}{\partial J}\right) \\
&= \frac{q}{1+\theta_\tau \tau} mwtp(Q) Q \frac{1}{J} (\Delta) \\
&= \frac{q(p(1+\tau)+t)}{1+\theta_\tau \tau} \frac{mwtp(Q)Q}{p(1+\tau)+t} \frac{1}{J} (\Delta) \\
&= -\frac{1}{\epsilon_D^*} \frac{pq}{J} (\Delta)
\end{aligned}$$

$\square$

Corollary. ***A1.*** *Consider the case of full-optimization ($\theta_\tau = \theta_t = 1$), homogeneous products ($\Lambda_0 = 0$) and no pre-existing taxes ($\tau_0 = t_0 = 0$). The marginal excess burden and pass-*

through formulas are given respectively by:

$$\frac{dW}{d\tau} = -Q_0 \frac{dp}{d\tau} \tag{17}$$

$$\rho_\tau = \frac{2 + \frac{\epsilon_D^* - \frac{\nu_q}{J_0}}{\epsilon_S \frac{\nu_q}{J_0}} - \frac{\frac{\nu_q}{J_0}}{\epsilon_D^*}}{2 - \frac{\nu_q}{J_0} + \frac{\epsilon_D^* - \frac{\nu_q}{J_0}}{\epsilon_S \frac{\nu_q}{J_0}} + \frac{\frac{\nu_q}{J_0}}{\epsilon_{ms}}} \tag{18}$$

$$\frac{1}{p_0}\frac{dJ}{d\tau} = -\frac{J\epsilon_D}{p_0}\left[\frac{1 + \frac{\epsilon_D^* - \frac{\nu_q}{J_0}}{\epsilon_S \frac{\nu_q}{J_0}} + \frac{1}{\epsilon_{ms}} + \frac{1 - \frac{\nu_q}{J_0}}{\epsilon_D^*}}{2 - \frac{\nu_q}{J_0} + \frac{\epsilon_D^* - \frac{\nu_q}{J_0}}{\epsilon_S \frac{\nu_q}{J_0}} + \frac{\frac{\nu_q}{J_0}}{\epsilon_{ms}}}\right] \tag{19}$$

Proof. ***Corollary A1.***

The proof is immediate by setting $\theta_\tau = \theta_t = 1$, $\Lambda_0 = 0$ and $\tau_0 = t_0 = 0$ into the conditions of Proposition 2. □

Lemma. ***A1.*** *For fixed* τ*. The effect of competition on prices and output is given respectively by:*

$$\frac{\partial p}{\partial J} = \left[\frac{\partial P}{\partial J} - \frac{p+t}{J\epsilon_D}\left(1 + \frac{J}{q}\frac{\partial q}{\partial J}\right)\right]$$
$$\frac{J}{q}\frac{\partial q}{\partial J} = -\omega_{SR}\left[1 - \frac{\nu_q}{J}\left(1 - \frac{1}{\epsilon_{ms}}\right) - \frac{J\epsilon_D}{(1+\tau)p + tq}\frac{\partial P}{\partial J}\right]$$

Thus, in the case of constant marginal cost ($\epsilon_S = \infty$), $\frac{\partial p}{\partial J} < 0$ *if and only if* $\frac{1}{\epsilon_{ms}}\frac{\Lambda\epsilon_D}{(p+t)q} < 1$ *and there is business stealing* $\left(\frac{\partial q}{\partial J} < 0\right)$ *whenever* $\frac{\Lambda\epsilon_D}{(p+t)q} + \frac{\nu_q}{J}\left(1 - \frac{1}{\epsilon_{ms}}\right) < 1$*.*

For fixed t. The effect of competition on prices and output is given respectively by:

$$\frac{\partial p}{\partial J} = \frac{1}{1+\theta_\tau \tau}\left[\frac{\partial P}{\partial J} - \frac{p\,(1+\tau)}{J\epsilon_D}\left(1 + \frac{J}{q}\frac{\partial q}{\partial J}\right)\right]$$
$$\frac{J}{q}\frac{\partial q}{\partial J} = -\omega_{SR}\left[1 - \frac{\nu_q}{J}\left(1 - \frac{1}{\epsilon_{ms}}\right) - \frac{J\epsilon_D}{(1+\tau)p + t}\frac{\partial P}{\partial J}\right]$$

Thus, in the case of constant marginal cost ($\epsilon_S = \infty$), $\frac{\partial p}{\partial J} < 0$ *if and only if* $\left(\frac{1}{\epsilon_{ms}}\right)\frac{\epsilon_D}{p(1+\tau)}J\frac{\partial P}{\partial J} < 1$ *and there is business stealing* $\left(\frac{\partial q}{\partial J} < 0\right)$ *whenever* $\frac{\epsilon_D}{(1+\tau)p}J\frac{\partial P}{\partial J} + \frac{\nu_q}{J}\left(1 - \frac{1}{\epsilon_{ms}}\right) < 1$*. Furthermore, assuming parallel demands* $\frac{\partial P}{\partial J} = \frac{\Lambda}{Q}$*.*

Proof. ***Lemma A1. Unit Taxes:***

From the behavioral equation $wtp(Q) = P(Q, J) = p + \theta t$, we can express price as a function of J and t. Then we have

$$p(J, t) = P(Q(J, t), J) - \theta t$$

Therefore,

$$\begin{aligned}
\frac{\partial p}{\partial J} &= \frac{\partial P}{\partial J} + mwtp(Q, J)\frac{\partial Q}{\partial J} \\
&= \frac{\partial P}{\partial J} + q * mwtp(Q, J) + mwtp(Q, J) * J * \frac{\partial q}{\partial J} \\
&= \left[\frac{\partial P}{\partial J} - \frac{p+t}{J\epsilon_D}\left(1 + \frac{J}{q}\frac{\partial q}{\partial J}\right)\right]
\end{aligned}$$

From the proof of Proposition 1, we also have that:

$$\begin{aligned}
\frac{\partial q}{\partial J} &= -\frac{\frac{\Lambda}{Q} + mwtp(Q)q + q^2\nu_q mwtp'(Q)}{(J+\nu_q)mwtp(Q) - c''(q) + Jq\nu_q mwtp'(Q)} \\
&= -\frac{\omega_{SR}\Lambda}{JQ * mwtp(Q)} - \frac{q}{J}\omega_{SR}\left(1 - \frac{\nu_q}{J} + \frac{\frac{\nu_q}{J}}{\epsilon_{ms}}\right)
\end{aligned}$$

Therefore,

$$\begin{aligned}
\frac{\partial p}{\partial J} &= \left[\frac{\Lambda}{Q} - \frac{p+t}{J\epsilon_D}\left(1 + \frac{J}{q}\frac{\partial q}{\partial J}\right)\right] \\
\frac{J}{q}\frac{\partial q}{\partial J} &= -\omega_{SR}\left[1 - \frac{\nu_q}{J}\left(1 - \frac{1}{\epsilon_{ms}}\right) - \frac{\Lambda\epsilon_D}{(p+t)q}\right]
\end{aligned}$$

Ad valorem:

The proof is analogous to Lemma 2. The only modification is that the behavioral equation for ad valorem taxation $p(J,t) = \frac{P(Q(J,t),J)}{1+\theta_\tau \tau}$ implies a rescaling is needed for $\frac{\partial p}{\partial J}$. □

B Comparison between Ad Valorem and Specific Taxation

We begin by considering the reduced-form effects of taxes in order to compare ad valorem to specific taxation. Throughout we will make use of the definitions $\epsilon_D = -\frac{p(1+\tau)+t}{Qmwtp(Q)}$, $\epsilon_D^* = \frac{p(1+\theta_\tau \tau)}{p(1+\tau)+t}\epsilon_D$, and $\Delta = \left[2 - \frac{\nu_q}{J} + \frac{\epsilon_D^* - \frac{\nu_q}{J}}{\epsilon_S \frac{\nu_q}{J}} + \frac{\frac{\nu_q}{J}}{\epsilon_{ms}}\right] - \left(1 + \frac{\epsilon_D^* - \frac{\nu_q}{J}}{\epsilon_S \frac{\nu_q}{J}} + \frac{1}{\epsilon_{ms}}\right)\epsilon_D \frac{J}{(p(1+\tau)+t)q}\frac{\partial P}{\partial J} + \left(1 - \frac{\nu_q}{J}\right)\epsilon_D \frac{JQ}{p(1+\tau)+t}\frac{\partial^2 P}{\partial J \partial Q} > 0$ for the stability condition:

$$
\begin{aligned}
\rho_t &= \frac{\Delta + \theta_t \frac{\nu_q}{J}\left(1 - \frac{1}{\epsilon_{ms}}\right)}{\Delta} \\
\rho_\tau &= \frac{\Delta + \frac{\nu_q}{J}\frac{\theta_\tau(1+\tau)}{(1+\theta_\tau \tau)}\left(1 - \frac{1}{\epsilon_{ms}} + \frac{1}{\epsilon_D^*}\left(\frac{J\epsilon_D}{(p(1+\tau)+t)}\frac{\partial P}{\partial J} - 1\right)\right)}{\Delta} \\
\frac{dq}{dt} &= -\frac{\theta_t q \epsilon_D}{p(1+\tau)+t}\left(\frac{1 - \frac{1}{\epsilon_{ms}}}{\Delta}\right) \\
\frac{dq}{d\tau} &= -\frac{\theta_\tau p q \epsilon_D}{p(1+\tau)+t}\left(\frac{1 - \frac{1}{\epsilon_{ms}} - \frac{1}{\epsilon_D^*} + \frac{J\epsilon_D}{(p(1+\tau)+t)}\frac{\partial P}{\partial J}\frac{1}{\epsilon_D^*}}{\Delta}\right) \\
\frac{dJ}{dt} &= -\frac{\theta_t J \epsilon_D}{p(1+\tau)+t}\left(\frac{1 + \frac{\epsilon_D^* - \frac{\nu_q}{J}}{\epsilon_S \frac{\nu_q}{J}} + \frac{1}{\epsilon_{ms}}}{\Delta}\right) \\
\frac{dJ}{d\tau} &= -\frac{\theta_\tau p J \epsilon_D}{p(1+\tau)+t}\left(\frac{1 + \frac{1}{\epsilon_D^*} - \frac{\frac{\nu_q}{J}}{\epsilon_D^*} + \frac{\epsilon_D^* - \frac{\nu_q}{J}}{\epsilon_S \frac{\nu_q}{J}} + \frac{1}{\epsilon_{ms}}}{\Delta}\right) \\
\frac{dQ}{dt} &= -\frac{\theta_t Q \epsilon_D}{p(1+\tau)+t}\left(\frac{2 + \frac{\epsilon_D^* - \frac{\nu_q}{J}}{\epsilon_S \frac{\nu_q}{J}}}{\Delta}\right) \\
\frac{dQ}{d\tau} &= -\frac{\theta_\tau p Q \epsilon_D}{p(1+\tau)+t}\left(\frac{2 + \frac{\epsilon_D^* - \frac{\nu_q}{J}}{\epsilon_S \frac{\nu_q}{J}} + \left(\frac{J\epsilon_D}{(p(1+\tau)+t)}\frac{\partial P}{\partial J} - \frac{\nu_q}{J}\right)\frac{1}{\epsilon_D^*}}{\Delta}\right) \\
\frac{dW}{dt} &= \Lambda\frac{dJ}{dt} + \theta_t t\frac{dQ}{dt} - Q\frac{dp}{dt} \\
\frac{dW}{d\tau} &= \Lambda\frac{dJ}{d\tau} + \theta_\tau \tau p\frac{dQ}{d\tau} - Q\frac{dp}{d\tau} \\
\frac{dR}{dt} &= Q + t\frac{dQ}{dt} \\
\frac{dR}{d\tau} &= pQ + \tau p\frac{dQ}{d\tau} + \tau Q\frac{dp}{d\tau}
\end{aligned}
$$

Proof. ***Proposition 3.*** Rewrite ρ_τ as:

$$\rho_\tau = \frac{\left[2 + \frac{\epsilon_D^* - \frac{\nu_q}{J}}{\epsilon_S \frac{\nu_q}{J}} - \left(1 - \frac{\theta_\tau(1+\tau)}{(1+\theta_\tau \tau)}\right)\left(\frac{\nu_q}{J} - \frac{\frac{\nu_q}{J}}{\epsilon_{ms}}\right)\right]}{\Delta}$$
$$- \frac{\frac{J\epsilon_D}{p(1+\tau)+t}\frac{\partial P}{\partial J}\left(1 + \frac{\epsilon_D^* - \frac{\nu_q}{J}}{\epsilon_S \frac{\nu_q}{J}} + \frac{1}{\epsilon_{ms}}\right) + \frac{\theta_\tau(1+\tau)}{(1+\theta_\tau \tau)}\frac{\frac{\nu_q}{J}}{\epsilon_D^*}\left[\frac{J\epsilon_D}{p(1+\tau)+t}\frac{\partial P}{\partial J} - 1\right]}{\Delta}$$

Then, observe that for $\theta_t = \frac{\theta_\tau(1+\tau)}{(1+\theta_\tau \tau)}$ (for example if $\theta_t = \theta_\tau$ and $\tau = 0$) then

$$\rho_\tau - \rho_t = \frac{\frac{\theta_\tau(1+\tau)}{(1+\theta_\tau \tau)}\frac{\frac{\nu_q}{J}}{\epsilon_D^*}\left[\frac{J\epsilon_D}{p(1+\tau)+t}\frac{\partial P}{\partial J} - 1\right]}{\Delta}$$

so

$$\rho_\tau > \rho_t \Leftrightarrow \frac{J\epsilon_D}{p(1+\tau)+t}\frac{\partial P}{\partial J} > 1 \Leftrightarrow \frac{\Lambda}{Q} + q * mwtp(Q) > 0$$

We now consider the marginal cost of public funds (MCPF) starting from zero initial taxes.

$$R = \tau p Q + t Q$$

$$\begin{aligned} MCPF_t &= -\frac{\Lambda \frac{dJ}{dt} + \theta_t t \frac{dQ}{dt} - Q\frac{dp}{dt}}{Q + t\frac{dQ}{dt}} \\ &= -\frac{\Lambda}{Q}\frac{dJ}{dt} + \frac{dp}{dt} \\ &= -\frac{\Lambda}{Q}\frac{dJ}{dt} + \rho_t - 1 \end{aligned}$$

$$\begin{aligned} MCPF_\tau &= -\frac{\Lambda \frac{dJ}{d\tau} + \theta_\tau \tau p \frac{dQ}{d\tau} - Q\frac{dp}{d\tau}}{pQ + \tau p\frac{dQ}{d\tau} + \tau Q \frac{dp}{d\tau}} \\ &= -\frac{\Lambda}{pQ}\frac{dJ}{d\tau} + \rho_\tau - 1 \end{aligned}$$

Furthermore,

$$\frac{dJ}{dt} = \frac{\theta_t}{\frac{\partial P}{\partial J} + q * mwtp(Q)} + \frac{1 - \frac{1}{\frac{\nu_q}{J}}}{\frac{\partial P}{\partial J} + q * mwtp(Q)} \frac{dp}{dt}$$

$$\frac{dJ}{d\tau} = \frac{p\theta_\tau}{\frac{\partial P}{\partial J} + q * mwtp(Q)} + (1 + \theta_\tau \tau) \frac{1 - \frac{1}{\frac{\nu_q}{J}}}{\frac{\partial P}{\partial J} + q * mwtp(Q)} \frac{dp}{d\tau}$$

and when taxes are zero, we get:

$$\frac{dJ}{dt} = \frac{\theta_t}{\frac{\partial P}{\partial J} + q * mwtp(Q)} + \frac{1 - \frac{1}{\frac{\nu_q}{J}}}{\frac{\partial P}{\partial J} + q * mwtp(Q)} (\rho_t - 1)$$

$$\frac{dJ}{d\tau} = \frac{p\theta_\tau}{\frac{\partial P}{\partial J} + q * mwtp(Q)} + \frac{1 - \frac{1}{\frac{\nu_q}{J}}}{\frac{\partial P}{\partial J} + q * mwtp(Q)} p(\rho_\tau - 1)$$

and so

$$\begin{aligned} MCPF_t &= -\frac{\Lambda}{Q} \frac{\theta_t}{\frac{\partial P}{\partial J} + q * mwtp(Q)} + (\rho_t - 1) \left(1 - \frac{\Lambda}{Q} \frac{1 - \frac{1}{\frac{\nu_q}{J}}}{\frac{\partial P}{\partial J} + q * mwtp(Q)} \right) \\ MCPF_\tau &= -\frac{\Lambda}{Q} \frac{\theta_\tau}{\frac{\partial P}{\partial J} + q * mwtp(Q)} + (\rho_\tau - 1) \left(1 - \frac{\Lambda}{Q} \frac{1 - \frac{1}{\frac{\nu_q}{J}}}{\frac{\partial P}{\partial J} + q * mwtp(Q)} \right) \end{aligned}$$

Assuming $\theta_t = \theta_\tau$ and $\tau = t = 0$, and $\frac{\partial P}{\partial J} = \frac{\Lambda}{Q}$, note that $1 - \frac{\Lambda}{Q} \frac{1 - \frac{1}{\frac{\nu_q}{J}}}{\frac{\Lambda}{Q} + q * mwtp(Q)} = \left(\frac{q * mwtp(Q) + \frac{\frac{\Lambda}{Q}}{\frac{\nu_q}{J}}}{\frac{\Lambda}{Q} + q * mwtp(Q)} \right)$.

Therefore:

$$\begin{aligned} sign(MCPF_\tau - MCPF_t) &= sign \left((\rho_\tau - \rho_t) * \frac{q * mwtp(Q) + \frac{\frac{\Lambda}{Q}}{\frac{\nu_q}{J}}}{\frac{\Lambda}{Q} + q * mwtp(Q)} \right) \\ &= sign \left(q * mwtp(Q) + \frac{\frac{\Lambda}{Q}}{\frac{\nu_q}{J}} \right) \end{aligned}$$

Finally, observe:

$$
\begin{aligned}
sign\left(\frac{1}{p}\frac{dJ}{d\tau} - \frac{dJ}{dt}\right) &= sign\left((\rho_\tau - \rho_t) * \frac{1 - \frac{1}{\frac{\nu_q}{J}}}{\frac{\partial P}{\partial J} + q * mwtp(Q)}\right) \\
&< 0
\end{aligned}
$$

$\square$

C Microfoundations for Demand

In this section, we provide the microfoundation for parallel demands. First, we introduce a class of continuous choice models that are nested by our utility function.

Preferences. Let the representative consumer's utility function given by

$$
u_J(q_1, \ldots, q_J, m) = h_J(q_1, \ldots, q_J) + m
$$

for any $h_J : \{1, \ldots, J\} \to \mathbb{R}$ which is symmetric in all its arguments, continuously differentiable, strictly quasi-concave and $h(0, \ldots, 0) = 0$ and where the linear good m is interpreted as money.

Demand. The consumer's problem is

$$
\begin{gathered}
\max u_J(q_1, \ldots, q_J, m) = h_J(q_1, \ldots, q_J) + m \\
\text{subject to } m + \sum_{j=1}^{J} p_j q_j = y.
\end{gathered}
\tag{20}
$$

When the consumer is facing symmetric prices $p_j = p$ for all j, we can transform the problem as follows. Define $H_J(Q) = h_J\left(\frac{Q}{J}, \ldots, \frac{Q}{J}\right)$ where we interpret Q as aggregate demand.

The new problem then is given by

$$u^*(p, J, y) = \max_Q H_J(Q) + y - pQ.$$

From the first-order condition, we obtain the family of inverse demands $P(Q, J) = H_J'(Q)$. Furthermore, it is easy to see that given the optimal aggregate quantity $Q(p, J)$ for price p, the strict quasi-concavity of h_J implies the consumer chooses symmetric quantities $q_j = \frac{Q}{J}$ for all j in the original problem.

Furthermore, none of the assumptions on utility are too restrictive. We show that for any family of downward sloping aggregate demands there exists a utility function $u_J : \mathbb{R}^{J+1} \to \mathbb{R}$ satisfying the conditions above that rationalize the aggregate demands. Let $P(Q, J)$ be continuously differentiable and strictly decreasing in Q. Let H be any antiderivative $\int P(Q, J)dQ$, which exists because $P(Q, J)$ is differentiable. Then, for some $\rho \in (0, 1)$, the following is a strictly quasi-concave direct utility function that rationalizes $P(Q, J)$ for integer J when all prices p_j in the market are equal:

$$u(q_1, \ldots, q_J, m) = H\left(\left(J^{\rho-1}\sum_{j=1}^{J} q_j^\rho\right)^{\frac{1}{\rho}}\right) + m.$$

Furthermore, we can make sense of J as a continuous variable if we permit a continuum of varieties $q : [0, J] \to \mathbb{R}$ and let

$$u_J(q, m) = H\left(\left(\int_0^J J^{\rho-1} q^\rho(j)dj\right)^{\frac{1}{\rho}}\right) + m.$$

We provide two examples in the following to further illustrate the idea of parallel demands and its applications.

Example 1. Bulow and Pfleiderer (1981) obtain the following three categories of inverse demands as the unique curves with the property of constant pass-through:

1. $P(Q, J) = \alpha_J - \beta_J Q^\delta$, for $\delta > 0$,
2. $P(Q, J) = \alpha_J - \beta_J log(Q)$,
3. $P(Q, J) = \alpha_J + \beta_J Q^{1/\eta}$, for $\eta < 0$, which is the constant elasticity inverse demand shifted by the intercept α_J.

An important case is when $\beta_J = \beta$ for all J, then the inverse aggregate demands are linearly separable in J and Q and they shift in parallel as J moves.[2] The fact that these are the only class of curves for which marginal costs are passed-on in a constant fraction makes them a tractable benchmark and therefore they have been popular in applied work. Fabinger and Weyl (2016) generalize Bulow and Pfleiderer (1983) and characterize a broader class of "tractable equilibrium forms" of the form $P(Q, J) = \alpha_J + \beta Q^t + \gamma Q^u$ which allow for greater modeling flexibility. Again, as long as β and γ are independent of J, then we say that the inverse demands shift in parallel.

Example 2. This example shows that our revealed-preference approach allows for rational preferences that display *hate-of-variety* ($a'(J) < 0$). Imagine there is a marginal cost of consumption cJ for each unit of some good that is consumed; that is, for each unit consumed, the agent faces a constant cost of evaluating each of J varieties before he chooses. Preferences are given by

$$U = H\left(\sum_{j=1}^{J} q_j\right) - cJ\sum_{j=1}^{J} q_j + m$$

where H is concave. The inverse demands are then $P(Q, J) = h(Q) - cJ$ with $h = H'$ decreasing, therefore aggregate demand shifts inward as the variety increases (the intercept being $h(0) - cJ$). We can interpret this as the agent displaying a strong degree of thinking aversion or attention costs. More generally, if the inverse demands are given by $P(Q, J) = a(J) - h(Q)$ then the sign of $a'(J)$ is unrestricted.

[2] For example, for the first class one possible family of utility functions, among many, that rationalize the inverse aggregate demands is given by

$$u_J(q_1, \ldots, q_J, m) = \alpha_J \left(J^{\rho-1}\sum_{i=1}^{J} q_i^\rho\right)^{\frac{1}{\rho}} - \beta_J \frac{\left(\sum_{i=1}^{J} q_i\right)^{\delta+1}}{\delta+1} + m.$$

D Formulas in Calibration

Taking logs and rescaling by $\frac{W}{pQ}$ equation (11) we obtain the following expression which we use in Section 5 of the paper:

$$\frac{dlog(W)}{dlog(1+\tau)}\frac{W}{pQ} = \tilde{\Lambda}_0 \frac{dlog(J)}{dlog(1+\tau)} - \frac{dlog(p)}{dlog(1+\tau)} + \theta_\tau \tau_0 \frac{dlog(Q_L)}{dlog(1+\tau)} \tag{21}$$

where $\tilde{\Lambda}_0 \equiv \frac{\Lambda_0}{pQ}$.

We now show the derivation equation (18) in the paper. Note that the Lerner condition $\frac{p-mc}{p(1+\tau)} = \frac{\frac{\nu_q}{J}}{(1+\theta_\tau\tau)\epsilon_D}$ and the long-run free entry condition $\frac{\frac{dlogp}{d\tau}}{\frac{dlogq}{d\tau}} = -\frac{p-mc}{p}$ we can identify

$$\frac{\nu_q}{J} = -\epsilon_D \frac{1+\theta_\tau \tau}{1+\tau} \frac{\frac{dlogp}{d\tau}}{\frac{dlogq}{d\tau}} \tag{22}$$

We have from Proposition 2, and assuming constant mc, that

$$\frac{dJ}{d\tau} = -\frac{\theta_\tau J \epsilon_D}{(1+\tau)} \left[\frac{1 + \frac{1}{\epsilon_{ms}} + \frac{1-\frac{\nu_q}{J}}{\epsilon_D^*}}{\triangle} \right]$$

and

$$\rho_\tau = \frac{\triangle - \frac{\theta_\tau(1+\tau)}{(1+\theta_\tau\tau)}\left(\frac{\frac{\nu_q}{J}}{\epsilon_D^*} - \frac{\nu_q}{J} + \frac{\frac{\nu_q}{J}}{\epsilon_{ms}}\right) + \frac{\Lambda\epsilon_D}{(p(1+\tau)+t)q}\left(\frac{\theta_\tau(1+\tau)}{(1+\theta_\tau\tau)}\frac{\frac{\nu_q}{J}}{\epsilon_D^*}\right)}{\triangle}$$

where $\triangle \equiv 1 + \left[1 + \frac{\epsilon_D^* - \frac{\nu_q}{J_0}}{\frac{\nu_q}{J}\epsilon_S}\right]\left[1 - \frac{\Lambda\epsilon_D}{(1+\tau)pq}\right] - \frac{1}{\epsilon_{ms}}\frac{\Lambda\epsilon_D}{(1+\tau)pq} - \frac{\nu_q}{J}\left[1 - \frac{1}{\epsilon_{ms}}\right]$. Then

$$\triangle = -\frac{\theta_\tau J \epsilon_D}{(1+\tau)}\left[\frac{1+\frac{1}{\epsilon_{ms}} + \frac{1-\frac{\nu_q}{J}}{\epsilon_D^*}}{\frac{dJ}{d\tau}}\right] = \frac{-\frac{\theta_\tau(1+\tau)}{(1+\theta_\tau\tau)}\left(\frac{\frac{\nu_q}{J}}{\epsilon_D^*} - \frac{\nu_q}{J} + \frac{\frac{\nu_q}{J}}{\epsilon_{ms}}\right) + \frac{\Lambda\epsilon_D}{(p(1+\tau)+t)q}\left(\frac{\theta_\tau(1+\tau)}{(1+\theta_\tau\tau)}\frac{\frac{\nu_q}{J}}{\epsilon_D^*}\right)}{\rho_\tau - 1}$$

And so, using $\rho_\tau - 1 = (1+\tau)\frac{dlog(p)}{d\tau}$, then

$$\frac{\Lambda\epsilon_D}{pq}\left(\frac{1}{(1+\theta_\tau\tau)}\frac{\frac{\nu_q}{J}}{\epsilon_D^*}\right) = -J\epsilon_D\frac{dlog(p)}{d\tau}\left[\frac{1+\frac{1}{\epsilon_{ms}}+\frac{1-\frac{\nu_q}{J}}{\epsilon_D^*}}{\frac{dJ}{d\tau}}\right] + \frac{1+\tau}{(1+\theta_\tau\tau)}\left(\frac{\frac{\nu_q}{J}}{\epsilon_D^*} - \frac{\nu_q}{J} + \frac{\frac{\nu_q}{J}}{\epsilon_{ms}}\right)$$

which implies

$$\frac{\Lambda}{pq} = -\frac{\epsilon_D^*}{\epsilon_D}(1+\theta_\tau \tau)\frac{\frac{\epsilon_D}{\frac{\nu_q}{J}}\frac{dlog(p)}{d\tau}}{\frac{dlog(J)}{d\tau}}\left(1+\frac{1}{\epsilon_{ms}}+\frac{1-\frac{\nu_q}{J_0}}{\epsilon_D^*}\right)+(1+\tau)\frac{\epsilon_D^*}{\epsilon_D}\left(\frac{1}{\epsilon_D^*}-1+\frac{1}{\epsilon_{ms}}\right)$$

Now, from $\frac{\epsilon_D^*}{\epsilon_D}=\frac{1+\theta_\tau\tau}{1+\tau}$ and equation (22) we get

$$\begin{aligned}
\frac{\Lambda}{p(1+\tau)q} &= -\frac{1+\theta_\tau\tau}{1+\tau}\left(\frac{1+\theta_\tau\tau}{1+\tau}\right)\frac{-\frac{1+\tau}{1+\theta_\tau\tau}\frac{dlog(q)}{d\tau}}{\frac{dlog(J)}{d\tau}}\left(1+\frac{1}{\epsilon_{ms}}+\frac{1-\frac{\nu_q}{J_0}}{\epsilon_D^*}\right)+\frac{1+\theta_\tau\tau}{1+\tau}\left(\frac{1}{\epsilon_D^*}-1+\frac{1}{\epsilon_{ms}}\right)\\
&= \left(\frac{1+\theta_\tau\tau}{1+\tau}\right)\frac{\frac{dlog(q)}{d\tau}}{\frac{dlog(J)}{d\tau}}\left(1+\frac{1}{\epsilon_{ms}}+\frac{1-\frac{\nu_q}{J_0}}{\epsilon_D^*}\right)+(1+\theta_\tau\tau)\left(\frac{1}{\epsilon_D^*}-1+\frac{1}{\epsilon_{ms}}\right)\\
&= \left(\frac{1+\theta_\tau\tau}{1+\tau}\right)\left[\frac{\frac{dlog(q)}{d\tau}}{\frac{dlog(J)}{d\tau}}\left(1+\frac{1}{\epsilon_{ms}}+\frac{1-\frac{\nu_q}{J_0}}{\epsilon_D^*}\right)+\frac{1}{\epsilon_D^*}+\frac{1}{\epsilon_{ms}}-1\right]\\
&= \left(\frac{1+\theta_\tau\tau}{1+\tau}\right)\left[\frac{1}{\epsilon_{ms}}\left(\frac{\frac{dlog(q)}{d\tau}}{\frac{dlog(J)}{d\tau}}+1\right)+\frac{\frac{dlog(q)}{d\tau}}{\frac{dlog(J)}{d\tau}}\left(1+\frac{1-\frac{\nu_q}{J_0}}{\epsilon_D^*}\right)+\frac{1}{\epsilon_D^*}-1\right]\\
&= \left(\frac{1+\theta_\tau\tau}{1+\tau}\right)\left[\frac{1}{\epsilon_{ms}}\left(\frac{\frac{dlog(Q)}{d\tau}}{\frac{dlog(J)}{d\tau}}\right)+\frac{\frac{dlog(Q)}{d\tau}-\frac{dlog(J)}{d\tau}}{\frac{dlog(J)}{d\tau}}\left(1+\frac{1-\frac{\nu_q}{J}}{\epsilon_D^*}\right)+\frac{1}{\epsilon_D^*}-1\right]\\
&= \left(\frac{1+\theta_\tau\tau}{1+\tau}\right)\left[\frac{1}{\epsilon_{ms}}\left(\frac{\hat{\beta^Q}}{\hat{\beta^J}}\right)+\frac{\hat{\beta^Q}}{\hat{\beta^J}}\left(1+\frac{1-\frac{\nu_q}{J}}{\epsilon_D^*}\right)+\frac{\frac{\nu_q}{J}}{\epsilon_D^*}-2\right]
\end{aligned}$$

E Data Appendix

E.1 Data Descriptiom

The RMS data records weekly prices and quantities by product at the barcode level, designated as Universal Product Codes (UPCs), for 35,000 stores in the United States (excluding Hawaii and Alaska). Products are organized in a hierarchical structure: there are over 2.5 million different UPCs, which are categorized into approximately 1,200 *product-module*s (e.g., fresh eggs, milk, aluminum foil, batteries, frozen desserts). In these data, we aggregate weekly revenue and quantities sold to the yearly level separately for each UPC. The average yearly price for product (UPC) j in store r is calculated by dividing the total yearly revenue (from the sales of that product) by the number of units sold.

To obtain a module-level price index (aggregating average yearly prices across all of the products in a module), we follow Handbury and Weinstein (2015) and regress log average yearly price on UPC fixed effects and store fixed effects, separately for each module and each year. The estimated store fixed effects serve as the pre-tax price. To measure quantity demanded, we re-calculate yearly revenue replacing the price of each product j in store r by the average national price (across all stores in our sample), and then aggregate across products within a module-by-store-by-year cell. This effectively constitutes a price-weighted quantity demanded index based on prices that are common across stores, an approach that is similar to the real consumption index developed by Kaplan, Mitman and Violante (2020). We measure product variety as the simple count of distinct UPCs with positive sales within a module-by-store-by-year cell, and we show robustness to alternative variety measure that weights each variety by its national market share.[3]

Our sales tax exemptions data is collected from a variety of sources, which includes state laws, state regulations, and online brochures. All sources are listed in Kroft et al. (2020), Online Appendix Table OA.2. In general, tax exemptions are set by U.S. states and are roughly module-specific. We therefore assign a tax exemption status to each state-module-year cell in our data. We then assign the appropriate tax rate given the inputed exemption status. In most cases, products are either taxed at the regular rate or fully exempt. In some states, some products are rather subject to a reduced rate. Our measures do take these features into account. We examine potential measurement error due to mis-assignment of exmption status in Section E.4 below.

[3]Note that when a product does not show up in the RMS data, we don't know if the product is not in the data because it wasn't purchased or because it wasn't available, so we aggregate to the year level to minimize the chance we are undercounting variety and avoiding a "mechanical" correlation between declining quantity demanded and low product variety.

E.2 Heterogeneous Effects

To investigate whether the reduced-form effects of taxes vary across types of products, we grouped modules into categories using Nielsen's product category codes, and Appendix Figure OA.5 presents the estimates of the effects of taxes on prices, quantity, and variety in each of the five broad categories (health and beauty care, dry grocery, other food, cleaning products, other nonfood). We also present estimates of the price elasticity of demand in each of these categories for comparison, using the methodology in Kroft et al. (2020).

E.3 Robustness

First, we report the yearly OLS and 2SLS estimates for each of the main outcomes reported in Table 2 (i.e., pre-tax prices, quantity, and variety) for each year in Appendix Figures OA.1 and OA.2. These figures show that the county border pair estimates are very stable across years and clustered around the across-year simple unweighted average. We gain precision by pooling the OLS and 2SLS estimates across years, and these figures show that our model-based estimates are not sensitive to the specific choice of years in the sample.

Second, we show that our results do not rely on specific county border pairs by dropping each state (one at a time) and re-running our reduced-form analysis dropping all of the county border pairs that have a county in the dropped state. Appendix Figure OA.3 shows that our main results are very stable as we drop each state one by one.[4]

[4] These results are consistent with the binscatter plots of regression residuals presented in Appendix Figure OA.6, which show that our estimated effects of taxes on prices, quantity, and variety do not appear to be driven by outliers. We also show robustness to dropping alcohol and tobacco products in Appendix Table OA.3, since these products have excise taxes in some states and counties. This would not cause bias in the reduced-form analysis if the variation in excise taxes is uncorrelated with the variation in ad valorem sales taxes that is our focus in this paper.

E.4 Measurement Error in Tax Rates

There are two potential sources of measurement error in how we code tax rates. First, we only consider state-level exemptions. That is, we do not incorporate county-level exemptions or county-specific sales surtaxes. Our understanding is that these cases are uncommon. Second, in practice taxability may vary within modules in some cases. For example, in the state of 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. We coded these products as tax exempt since we cannot readily identify the specific products that do not meet that threshold. 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 chose to code the statutory tax rate as missing. This results in excluding less than 3 percent of the observations in our sample.

To insure that our results are not contaminated by measurement error, we re-estimate our key reduced-form elasticities excluding modules for which we suspect there might be some within-module variation in exemption status in some states. These modules are listed below:

CANDY-NON-CHOCOLATE
CANDY-CHOCOLATE
FRUIT JUICE - APPLE
FRUIT DRINKS & JUICES-CRANBERRY
FRUIT DRINKS-OTHER CONTAINER
FRUIT JUICE-REMAINING
FRUIT JUICE - ORANGE - OTHER CONTAINER
VEGETABLE JUICE AND DRINK REMAINING
BAKERY - BREAD - FRESH
BAKERY-BUNS-FRESH
BAKERY-ROLLS-FRESH
BAKERY-MUFFINS-FRESH
BAKERY-CAKES-FRESH
BAKERY-BREAKFAST CAKES/SWEET ROLLS-FRESH
BAKERY-DOUGHNUTS-FRESH
BAKERY-BAGELS-FRESH
WATER-BOTTLED
FRUIT-DRIED AND SNACKS
PRECUT FRESH SALAD MIX
FRUIT-REFRIGERATED
COMBINATION LUNCHES
REMAINING-READY MADE SALADS
ENTREES-REFRIGERATED

SANDWICHES-REFRIGERATED/FROZEN
MAGAZINES SELECTED TITLES
DIETING AIDS-COMPLETE NUTRITIONAL
NUTRITIONAL SUPPLEMENTS

Results for models that exclude the modules listed above are presented in Appendix Table OA.6. All estimates are very similar to those reported in Table 2.

References

Bulow, Jeremy I, and Paul Pfleiderer. 1983. "A note on the effect of cost changes on prices." *Journal of Political Economy*, 91(1): 182–185.

Chetty, Raj, Adam Looney, and Kory Kroft. 2009. "Salience and Taxation: Theory and Evidence." *American Economic Review*, 99(4): 1145–1177.

Fabinger, Michal, and E. Glen Weyl. 2016. "Functional Forms for Tractable Economic Models and the Cost Structure of International Trade." *Working Paper*.

Handbury, Jessie, and David E Weinstein. 2015. "Goods prices and availability in cities." *The Review of Economic Studies*, 82(1): 258–296.

Kaplan, Greg, Kurt Mitman, and Giovanni L Violante. 2020. "Non-durable consumption and housing net worth in the great recession: Evidence from easily accessible data." *Journal of Public Economics*, 104176.

Kroft, Kory, Jean-William Laliberté, René Leal Vizcaíno, and Matthew Notowidigdo. 2020. "Salience and Taxation with Imperfect Competition." *NBER Working Paper*.

Online Appendix Table OA.1:
Effect of Food and Nonfood Sales Taxes **[Placebo Test]**

	(1)	(2)	(3)	(4)
Panel A: Dependent variable is log Prices				
Own tax rate differential	0.187		0.166	0.045
	(0.021)		(0.018)	(0.011)
Other tax rate differential		0.150	0.120	
		(0.021)	(0.018)	
Panel B: Dependent variable is log Quantity				
Own tax rate differential	-0.844		-0.850	-0.878
	(0.258)		(0.227)	(0.173)
Other tax rate differential		-0.125	0.029	
		(0.257)	(0.227)	
Panel C: Dependent variable is log Variety				
Own tax rate differential	-0.206		-0.216	-0.270
	(0.125)		(0.115)	(0.100)
Other tax rate differential		0.015	0.054	
		(0.106)	(0.093)	
Specification:				
Food dummy	y	y	y	y
Cell (border pair by year) fixed effects				y
N (observations)	8430	8430	8430	8430

Notes: This table reports regressions of prices, quantity, and product variety on average tax rates for food and nonfood products. For each border pair-by-year cell, there are two observations: one for food products and one for nonfood products. All variables are measured as within-cell differences between the two contiguous counties. Own tax rate is the average food tax rate differential for food observations and the average nonfood tax rate differential for nonfood observations. Other tax rate is the average food tax rate differential for nonfood observations and the average nonfood tax rate differential for food observations. Standard errors are clustered at the border pair-by-year cell-level. Each regression includes a dummy variable for food products. Observations are weighted to reflect the number of underlying module-by-store-by-year observations in each cell.

Online Appendix Table OA.2
Variance Decomposition of Tax Rates

Variance of log(1+τ)	0.0010
Standard deviation of log(1+τ)	0.0312
Standard deviation within:	
Store × Year cells	0.0269
Module × Border Pair × Year cells	0.0108
Fraction of variance within:	
Store × Year cells	74.6%
Module × Border Pair × Year cells	11.9%

Notes: This table reports variance decompositions of the tax rate variable in the RMS data.

Online Appendix Table OA.3:
Effect of Sales Taxes on Prices, Quantity, and Product Variety
[Robustness to Dropping Alcohol and Tobacco Product Modules]

Dependent Variable:	Prices	Quantity	Variety
	(1)	(2)	(3)
Panel A: County Border Pair OLS Estimates			
$\log(1+\tau_{mcn})$	0.008	-0.678	-0.261
	(0.011)	(0.137)	(0.060)
Panel B: 2SLS Estimates Using State-Level Tax Rate as Instrument			
$\log(1+\tau_{mcn})$	0.011	-0.736	-0.267
	(0.011)	(0.135)	(0.060)
Specification:			
Store fixed effects	y	y	y
Module × County Border Pair fixed effects	y	y	y

Notes: Sales tax rates are measured annually based on the rates that were effective on September 1. Sales, prices, and variety are measured yearly. The Retail Scanner data is restricted to modules above the 80th percentile of the national distribution of sales. All reported coefficients are simple averages of nine estimated coefficients -- one for each year from 2006 to 2014. The sample is restricted to border counties and observations are weighted by the inverse of number of pairs a store belongs to. Standard errors are clustered two-way at the state-module level and at the border pair by module level. In panel B, the tax rate is instrumented with the state-level, leave-county-out, average tax rate.

Online Appendix Table OA.4: Additional Sensitivity Analysis of Calibration Results

	Baseline calibration	Alternative demand elasticity and tax salience parameters			
Panel A: Calibrated parameters					
Average tax rate, τ_0	0.034	0.034	0.034	0.034	0.034
Tax salience parameter, θ_τ	0.556	**0.500**	**0.612**	**0.556**	**0.556**
Demand elasticity, ϵ_D	1.170	**1.170**	**1.170**	**1.287**	**1.053**
Panel B: Reduced-form estimates					
Pass-through of taxes into pre-tax prices, $d\log(p)/d\log(1+\tau)$	0.039	0.039	0.039	0.039	0.039
Quantity response, $d\log(Q)/d\log(1+\tau)$	-0.731	-0.731	-0.731	-0.731	-0.731
Variety response, $d\log(J)/d\log(1+\tau)$	-0.243	-0.243	-0.243	-0.243	-0.243
Panel C: Model parameters estimated by matching reduced-form estimates					
Markup, $(p - c'(q))/p$	0.080	0.080	0.080	0.080	0.080
Implied conduct parameter, v_q/J	0.092	**0.092**	**0.092**	**0.101**	**0.083**
Inverse elasticity of marginal surplus, ϵ_{ms}	-0.903	**-0.970**	**-0.846**	**-0.903**	**-0.903**
Variety effect parameter, $\tilde{\Lambda}_0$	0.125	**0.366**	**-0.108**	**-0.113**	**0.425**
Panel D: Calibrated welfare formulas					
Full marginal excess burden (MEB) formula, $d\tilde{W}/d\tau$	-0.083	-0.140	-0.028	-0.025	-0.156
Alternative MEB formula benchmarks:					
Harberger/CLK benchmark, $\theta_\tau * \tau_0 * d\log(Q)/d\log(1+\tau)$	-0.014	-0.012	-0.015	-0.014	-0.014
Besley(1989)-style benchmark; i.e., full MEB formula with $\tilde{\Lambda}_0 = 0$	-0.053	-0.051	-0.054	-0.053	-0.053
% difference between full formula and Besley(1989)-style benchma	57.5%	172.9%	-48.3%	-51.8%	195.3%

Notes: This table reports structural parameter estimates by finding parameters that allow the model to match the reduced-form estimates. The table reports sensitivity to different assumptions on the demand elasticity and the tax salience parameter.

Online Appendix Table OA.5: Counterfactual Scenarios Comparing Ad Valorem and Unit Tax Taxes

Variety effect parameter, $\tilde{\Lambda}_0$	Baseline variety effect estimate, $\tilde{\Lambda}_0 = 0.125$		No variety effect counterfactual, $\tilde{\Lambda}_0 = 0.000$		Large variety effect counterfactual, $\tilde{\Lambda}_0 = 1.000$	
	Ad valorem tax ($d\tau$)	Specific tax (dt)	Ad valorem tax ($d\tau$)	Specific tax (dt)	Ad valorem tax ($d\tau$)	Specific tax (dt)
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Pass-through of taxes into pre-tax prices						
$d\log(p)/d\log(1+\tau)$ or $d\log(p)/dt$	0.039	0.059	0.036	0.060	0.060	0.056
Difference b/w ad valorem and specific tax	-0.020		-0.024		0.004	
Panel B: Marginal cost of public funds ($MCPF$)						
$MCPF_\tau$ or $MCPF_t$	0.082	0.067	0.048	0.072	0.307	0.033
Difference between ad valorem and specific tax	0.015		-0.024		0.273	
Panel C: The effects of taxes on variety and profits						
$d\log(J)/d\log(1+\tau)$ or $d\log(J)/dt$	-0.243	0.037	-0.245	0.037	-0.230	0.035
$\partial\log(\pi)/\partial\log(1+\tau)$ or $\partial\log(\pi)/\partial t$	-0.041	0.006	-0.041	0.006	-0.041	0.006
Panel D: Competitive effects of entry						
$\partial\log(p)/\partial\log(J)$	-0.106	-0.105	-0.092	-0.091	-0.207	-0.204
$\partial\log(q)/\partial\log(J)$	-0.751	-0.740	-0.911	-0.897	0.368	0.363
Stability condition (must be >0)	1.822	1.822	1.806	1.806	1.927	1.927

Notes: This table reports counterfactual estimates of reduced-form effects of specific taxes under different assumptions on the variety effect based on using the model parameter estimates of Table 4. The difference between the ad valorem and specific tax *MCPF* estimates ($MCFP_\tau$ - $MCPF_t$) switches sign as the variety effect increases (comparing columns (1) and (2) to (3) and (4)). The difference between ad valorem and specific tax pass-through rate is less sensitive to the variety effect and only switches sign when the variety effect is large (columns (5) and (6)).

Online Appendix Table OA.6: Robustness to Measurement Error, Effect of Sales Taxes on Prices, Quantity, and Product Variety

Dependent Variable:	Prices	Quantity	Variety
	(1)	(2)	(3)
Panel A: County Border Pair OLS Estimates			
$\log(1+\tau_{mcn})$	0.028	-0.607	-0.198
	(0.017)	(0.170)	(0.080)
Panel B: 2SLS Estimates Using State-Level Tax Rate as Instrument			
$\log(1+\tau_{mcn})$	0.030	-0.671	-0.207
	(0.017)	(0.169)	(0.080)
Specification:			
Store fixed effects	y	y	y
Module × County Border Pair fixed effects	y	y	y

Notes: The sample is derived from the Nielsen Retail Scanner data covering the years 2006-2014. The sample excludes modules with potential variation in tax rate exemptions across products within the module. Sales tax rates are measured annually based on the rates that were effective on September 1. Sales, prices, and variety are measured yearly. All reported coefficients are simple averages of nine estimated coefficients -- one for each year from 2006 to 2014. The sample is restricted to border counties and observations are weighted by the inverse of number of pairs a store belongs to. Standard errors are clustered two-way at the state-module level and at the border pair by module level. In panel B, the tax rate is instrumented with the state-level, leave-county-out, average tax rate.

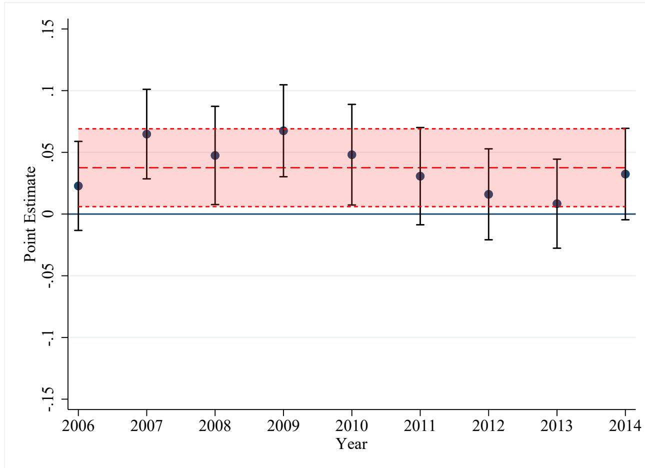
Online Appendix Table OA.7: Sensitivity of calibration results to alternative values of variety response, demand elasticity, and tax salience parameters

	Baseline calibration	Alternative measure of variety response: Hold model parameters fixed	Alternative measure of variety response: Re-estimate model parameters	Alternative demand elasticity and tax salience parameters			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Calibrated parameters							
Average tax rate, τ_0	0.034	0.034	0.034	0.034	0.034	0.034	0.034
Tax salience parameter, θ_τ	0.556	0.556	0.556	**0.500**	**0.612**	**0.445**	**0.667**
Demand elasticity, ϵ_D	1.170	1.170	1.170	**1.287**	**1.053**	**1.404**	**0.936**
Panel B: Reduced-form estimates							
Pass-through of taxes into pre-tax prices, $d\log(p)/d\log(1+\tau)$	0.039	0.039	0.039	0.039	0.039	0.039	0.039
Quantity response, $d\log(Q)/d\log(1+\tau)$	-0.731	-0.731	-0.731	-0.731	-0.731	-0.731	-0.731
Variety response, $d\log(J)/d\log(1+\tau)$	-0.243	**-0.193**	**-0.193**	-0.243	-0.243	-0.243	-0.243
Panel C: Model parameters estimated by matching reduced-form estimates							
Markup, $(p - c'(q))/p$	0.080	0.080	**0.072**	0.080	0.080	0.080	0.080
Implied conduct parameter, v_q/J	0.092	0.092	**0.084**	**0.101**	**0.083**	**0.110**	**0.074**
Inverse elasticity of marginal surplus, ϵ_{ms}	-0.903	-0.903	**-0.804**	**-0.970**	**-0.846**	**-1.047**	**-0.795**
Variety effect parameter, $\tilde{\Lambda}_0$	0.125	0.125	**0.157**	**0.124**	**0.188**	**0.160**	**0.320**
Panel D: Calibrated welfare formulas							
Full marginal excess burden (MEB) formula, $d\tilde{W}/d\tau$	-0.083	-0.077	-0.083	-0.082	-0.100	-0.089	-0.133
Alternative MEB formula benchmarks:							
Harberger/CLK benchmark, $\theta_\tau * \tau_0 * d\log(Q)/d\log(1+\tau)$	-0.014	-0.014	-0.014	-0.012	-0.015	-0.011	-0.017
Besley(1989)-style benchmark; i.e., full MEB formula with $\tilde{\Lambda}_0 = 0$	-0.053	-0.053	-0.053	-0.051	-0.054	-0.050	-0.056
% difference between full formula and Besley(1989)-style benchma	57.5%	45.6%	57.5%	58.8%	84.1%	77.6%	139.9%

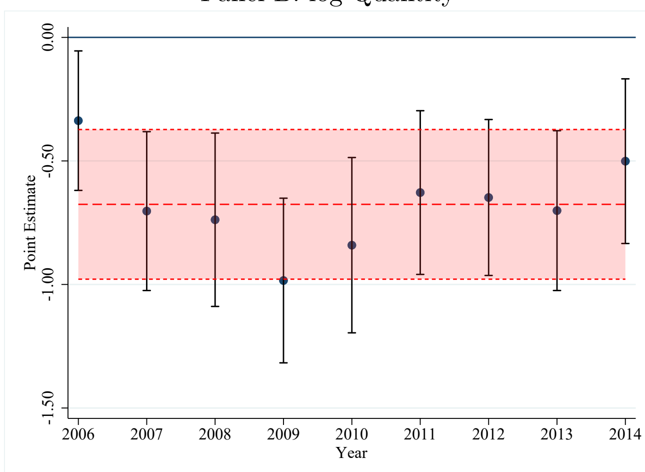
Notes: This table reports structural parameter estimates by finding parameters that allow the model to match the reduced-form estimates. The table reports sensitivity to different assumptions on the demand elasticity and the tax salience parameter. Columns (2) and (3) use the alternative variety response to taxes, while columns (4) through (7) vary both the demand elasticity and tax salience parameters but hold the product of the tax salience parameter and demand elasticity constant in order to ensure that dlog(Q)/dlog(1+τ) is constant.

Figure OA.1: Year-by-Year OLS Regression Coefficients

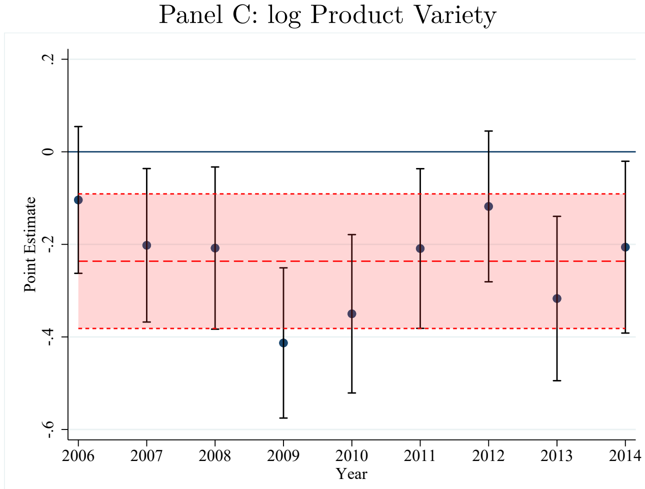
Panel A: log Prices



Panel B: log Quantity



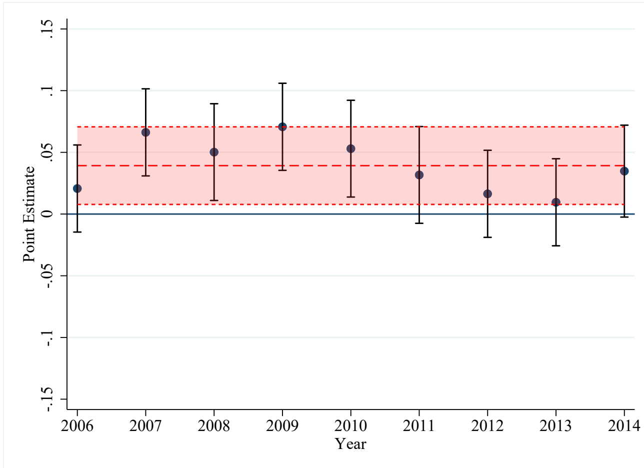
Panel C: log Product Variety



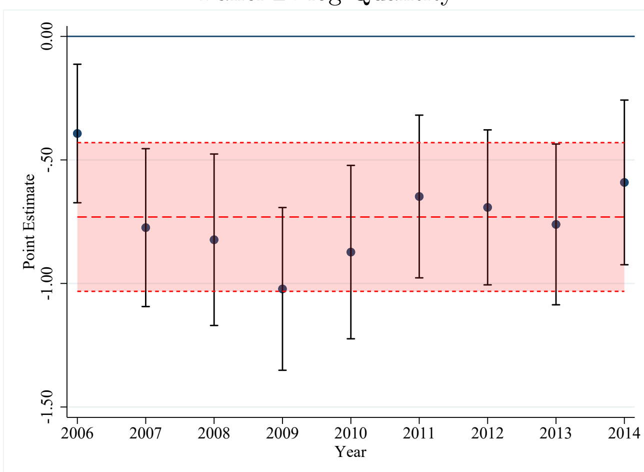
Notes: This figures shows yearly estimates of the effects of sales taxes on price (panel A), quantity (panel B) and product variety (C). All models are based on equation (17) and estimated by OLS. The black vertical bars indicate 95% confidence intervals. The dashed red horizontal line indicates the average coefficient estimate across all 9 years, and the red area denotes the 95% confidence interval around that average.

Figure OA.2: Year-by-Year 2SLS Regression Coefficients

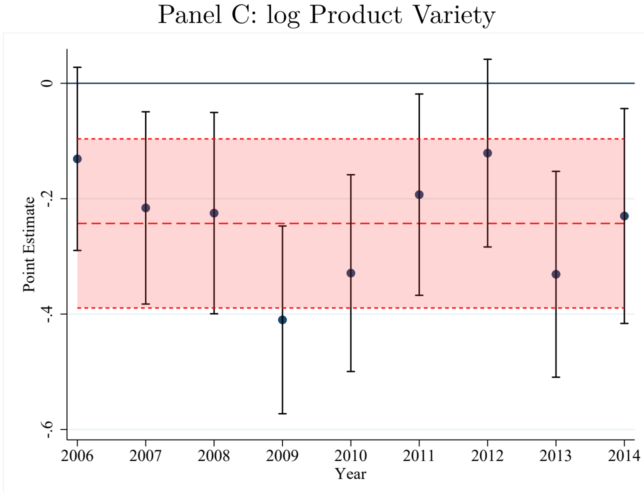
Panel A: log Prices



Panel B: log Quantity



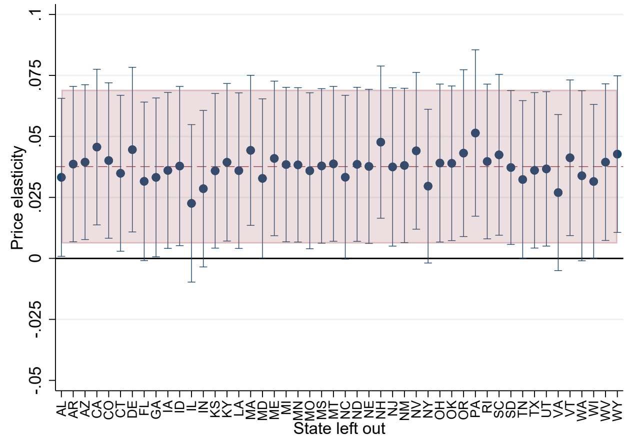
Panel C: log Product Variety



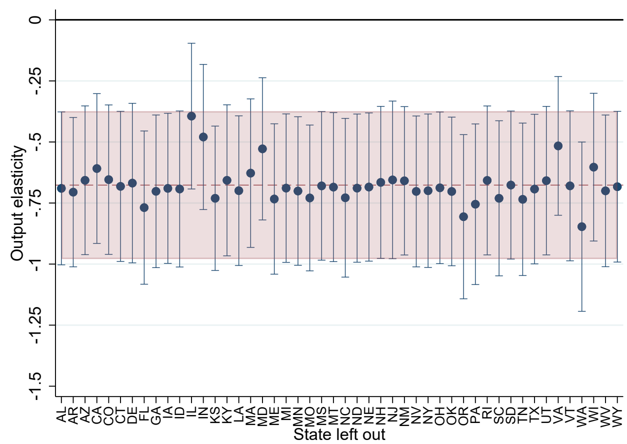
Notes: This figures shows yearly estimates of the effects of sales taxes on price (panel A), quantity (panel B) and product variety (C). All models are based on equation (17) and estimated by 2SLS. The instrument is the average state-level, leave-county-out average tax rate for each module-year cell. The black vertical bars indicate 95% confidence intervals. The dashed red horizontal line indicates the average coefficient estimate across all 9 years, and the red area denotes the 95% confidence interval around that average.

Figure OA.3: Leave-State-Out Regression Coefficients

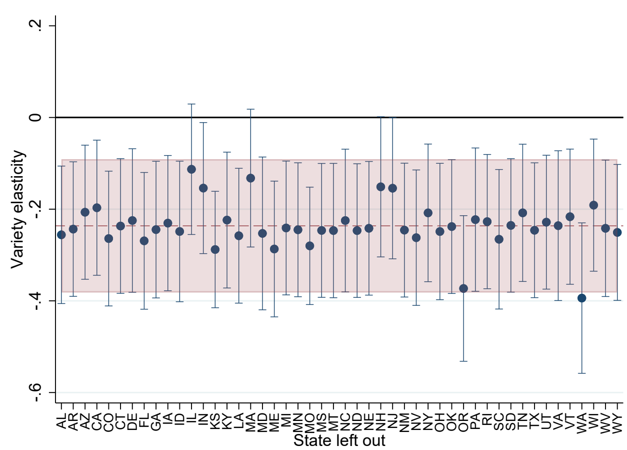
Panel A: log Prices



Panel B: log Quantity

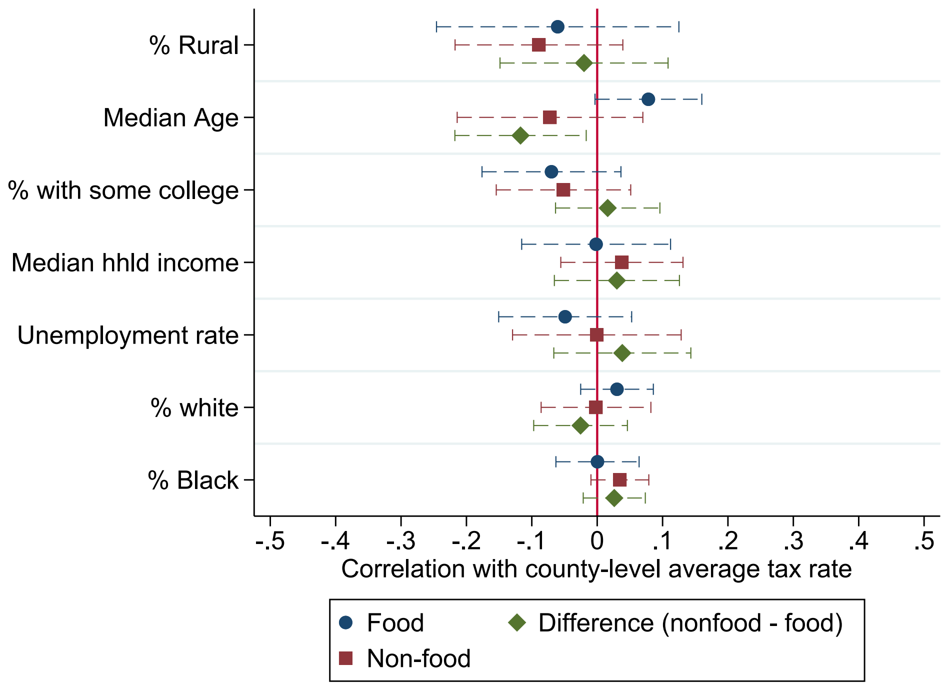


Panel C: log Product Variety



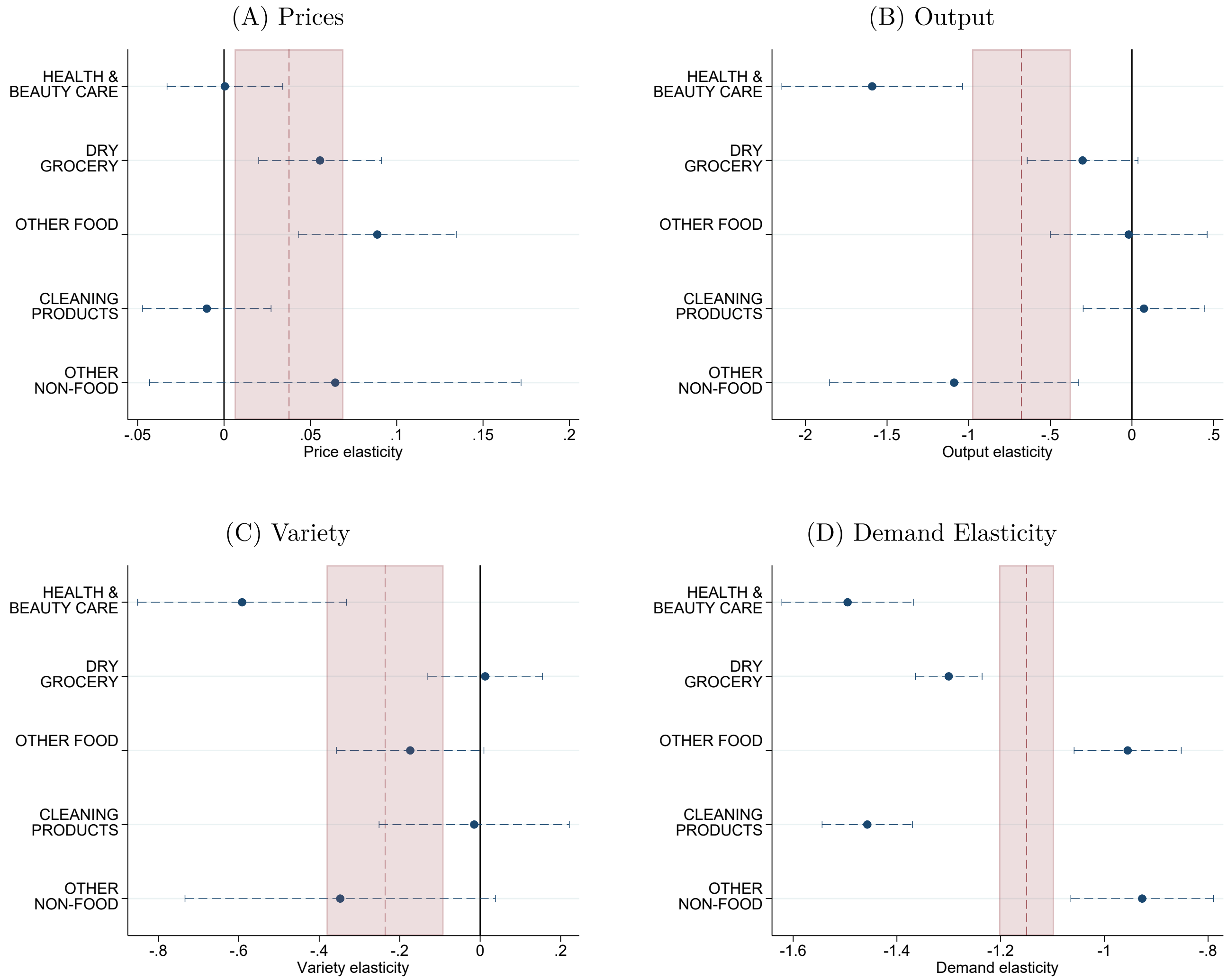
Notes: This figures shows yearly leave-state-out estimates of the effects of sales taxes on price (panel A), quantity (panel B) and product variety (C). All models are based on equation (17) and estimated by OLS. For each regression, all stores located in a given state or in a county adjacent to that state are dropped. The blue vertical bars indicate 95% confidence intervals. The dashed red horizontal line indicates the average coefficient estimate across all 9 years, and the red area denotes the 95% confidence interval around that average.

Figure OA.4: Correlations between County Demographics and Tax Rates



Notes: This figures shows correlation coefficients between county-level demographics (from the American Community Survey) and county-level average sales tax rates in 2008. Blue dots depict correlations with the average tax rate on food products. Red squares depict correlations with the average tax rate on non-food products. Green diamonds depict correlations with the county-specific difference between tax rates on non-food and food products. All correlations are estimated by OLS using a specification that includes border-pair fixed effects. The horizontal dashed bars indicate 95% confidence intervals. Standard errors are clustered at the state level.

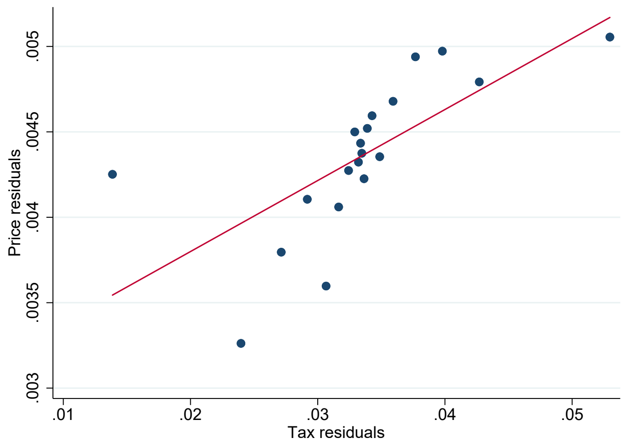
Figure OA.5: Heterogeneity Across Product Categories



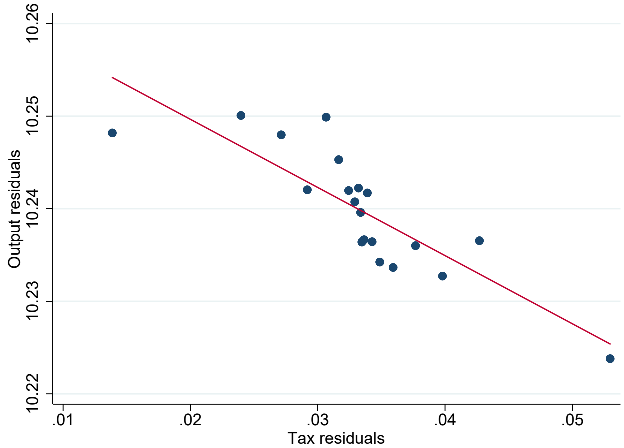
Notes: This figures shows estimates of the effects of sales taxes on price (panel A), quantity (panel B) and product variety (C) for different categories of products. Models for panels A, B and C are based on an augmented version of equation (17), in which tax rates are interacted with indicators for 5 different categories of goods. Panel D shows corresponding estimates of the demand elasticity, estimated using the methods described in Kroft et al. (2021). The blue dashed bars indicate 95% confidence intervals. The red vertical line indicates the average coefficient estimate across all 9 years, and the red area denotes the 95% confidence interval around that average.

Figure OA.6: Binscatter Plots of Regression Residuals

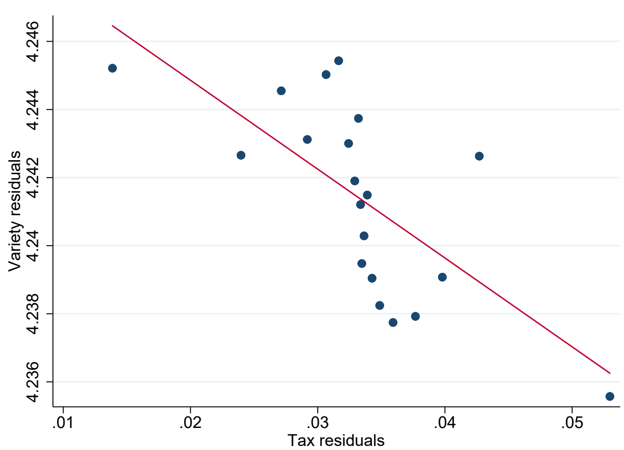
Panel A: log Prices



Panel B: log Quantity



Panel C: log Product Variety



Notes: This figures shows binscatter plots of regression residuals from models estimating the effects of sales taxes on price (panel A), quantity (panel B) and product variety (C). The number of bins is set to 20. All residuals are based on equation (17) and estimated by OLS. The red lines show the linear fit, the slope of which corresponds to our main estimates reported in Table 2.