

WORKING PAPER · NO. 2021-63

A Welfare Analysis of Tax Structures with Love-of-Variety Preferences

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

MAY 2021

A WELFARE ANALYSIS OF TAX STRUCTURES WITH LOVE-OF-VARIETY PREFERENCES

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

May 2021

This paper extends and replaces a previously-circulated 2017 working paper by the same authors titled “Quantifying the welfare gains of variety: A sufficient statistics approach.” We thank Simon Anderson, Raj Chetty, Julie Cullen, Amy Finkelstein, Nathan Hendren, Louis Kaplow, Henrik Kleven, Nicholas Li, Jesse Shapiro, Rob Porter, Aviv Nevo, Stephen Coate, and numerous seminar participants for helpful comments. We thank Eileen Driscoll, Robert French, Adam Miettinen, Boriana Miloucheva, Pinchuan Ong, Shahar Rotberg, Marc-Antoine Schmidt, Stephen Tino, Jessica Wagner, Ting Wang, and Haiyue Yu for extremely valuable research assistance. We gratefully acknowledge funding from the Social Sciences and Humanities Research Council (SSHRC). Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors(s) and do not necessarily reflect the views of the SSHRC. This research is based in part on data from The Nielsen Company (US), LLC and marketing databases provided through the Nielsen Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the Nielsen data are those of the researchers and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein. The review by the NBER Board of Directors that accompanies official NBER publications.

© 2021 by Kory Kroft, Jean-William P. Laliberté, René Leal Vizcaíno, and Matthew J. Notowidigdo. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

A Welfare Analysis of Tax Structures with Love-of-Variety Preferences
Kory Kroft, Jean-William P. Laliberté, René Leal Vizcaíno, and Matthew J. Notowidigdo
May 2021
JEL No. H20,H22,H71

ABSTRACT

This paper reassesses the general trade-off between ad valorem and specific taxation using an economic model that features love-of-variety preferences and encompasses a wide range of market conduct –including both quantity and price competition – while allowing for firm entry and exit. We derive formulas for efficiency and incidence of both types of taxes that depend on the responsiveness of product variety to taxes and the effect of a change in product variety on consumer surplus. We use our formulas to derive a desirability condition for when ad valorem taxes are more efficient than specific taxes and a condition for when ad valorem taxes lead to greater pass-through than specific taxes. We identify and estimate the model parameters using a quasi-experimental “county border pair” research design that uses state-level and county-level variation in sales taxes combined with detailed scanner data covering grocery stores in the U.S. Our empirical results indicate that sales taxes are slightly overshifted onto consumer prices, have a large effect on quantity demanded, and have a more modest effect on the variety of products available to consumers. Using the estimated parameters, we recover consumers' love-of-variety, infer whether or not product variety is socially optimal (at current tax rates), and implement our desirability condition. We find that specific taxes are more efficient at the margin than ad valorem taxes given the estimated love-of-variety. This suggests that policymakers should consider using specific taxes and tariffs in markets with substantial product differentiation.

Kory Kroft
Department of Economics
University of Toronto
150 St. George Street
Toronto, ON M5S 3G7
CANADA
and NBER
kory.kroft@utoronto.ca

Jean-William P. Laliberté
Department of Economics
University of Calgary
2500 University Drive NW
Calgary, Alberta, T2N 1N4
CANADA
jeanwilliam.lalibert@ucalgary.ca

René Leal Vizcaíno
Bank of Mexico
Avenida 5 de Mayo 2
Colonia Centro
Código postal 06000
Alcaldía Cuauhtémoc
Ciudad de México
México
rlealv@banxico.org.mx

Matthew J. Notowidigdo
University of Chicago
Booth School of Business
5807 S Woodlawn Ave
Chicago, IL 60637
and NBER
noto@chicagobooth.edu

1 Introduction

One of the longstanding questions in economics concerns the relative efficiency and incidence of ad valorem taxation versus specific taxation. Starting with Wicksell (1896, 1959), a voluminous literature has found that in models featuring monopoly and oligopoly markets with homogeneous products, ad valorem taxes are more efficient than specific taxes and are associated with lower consumer prices both in the short run with a fixed number of firms and in the long run with free entry (Suits and Musgrave 1953, Cournot 1960, Dierickx, Matutes and Neven 1988, Delipalla and Keen 1992, Myles 1996, Skeath and Trandel 1994, Keen 1998, Denicolò and Matteuzzi 2000, Auerbach and Hines 2001, Blackorby and Murty 2007). Similar results that compare ad valorem tariffs to specific tariffs exist in the international trade literature (see Helpman and Krugman 1989 and Kowalczyk 1990).

While several influential papers have shown that specific taxes can dominate ad valorem taxes in the long run when consumers exhibit love-of-variety preferences (see, for example, Anderson, De Palma and Kreider 2001*b* and Hamilton 2009), they consider specific settings; for example, they assume a particular type of market structure and/or class of preferences. They also do not attempt to derive general formulas that connect the efficiency cost or pass-through of a tax with economic fundamentals, as in Chetty (2009) (efficiency) and Weyl and Fabinger (2013) (pass-through). The remark in Keen (1998) thus still 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.”

Beyond this academic interest, understanding the efficiency and incidence of different tax structures is also practically relevant since governments rely on both ad valorem taxes and specific taxes as major sources of tax revenue. For example, in the U.S. in 2019, roughly 336 billion dollars in state tax revenue was collected through ad valorem taxes (retail sales taxes) and 77 billion dollars was collected via specific taxes.¹ Given the prevalence of both types of taxes, it is important to evaluate the robustness of the theoretical results derived in special settings and to derive formulas characterizing the efficiency cost and incidence of each tax that are tightly connected to estimable reduced-form elasticities (or “policy” elasticities, in the language of Hendren 2016). Such formulas facilitate empirical implementation using quasi-experimental tax variation, and thus can shed light on the welfare effects of alternative

¹See the U.S. Census Bureau’s 2019 State Government Tax Tables. About 100 billion dollars is collected in federal excise tax revenue, a fraction of which is levied through specific taxes (e.g. gasoline) and another through ad valorem taxes (e.g. airline tickets). See <https://www.taxpolicycenter.org/briefing-book/what-are-major-federal-excise-taxes-and-how-much-money-do-they-raise> for details.

tax structures that are available to policymakers.

In this paper, we reassess the general trade-off between ad valorem and specific taxation using an economic model that features love-of-variety preferences and encompasses a wide range of market conduct – including both quantity and price competition – while allowing for firm entry and exit. We derive formulas for efficiency and incidence of specific and ad valorem taxation that depend on the responsiveness of product variety to taxes and the effect of a change in product variety on consumer surplus – hereafter referred to as the “variety effect”. Our formulas allow us to derive a straightforward desirability condition for when ad valorem taxes are more efficient than specific taxes and a condition for when ad valorem taxes lead to greater pass-through than specific taxation.

We estimate the inputs into our formulas using quasi-experimental variation in ad valorem taxes (sales taxes), use the resulting parameters estimates to calibrate the welfare cost of an ad valorem tax, and consider two additional applications. First, we use the estimated parameters to counterfactually calibrate the efficiency cost and pass-through rate of specific taxes – even though we do not actually observe independent variation in specific taxes in our empirical setting. This is possible because our theory of taxation shows that the efficiency cost and pass-through rate of an ad valorem tax and a specific tax are functions of the same set of model parameters. This facilitates the implementation of our desirability condition in order to assess whether ad valorem or specific taxes dominate in our specific empirical setting with differentiated products, and our findings point to the welfare superiority of specific taxes, in contrast to the conventional wisdom. We also implement our pass-through comparison and find that specific taxes lead to greater pass-through than ad valorem taxes, which corroborates existing results on pass-through in the literature. Second, we use the estimated parameters to recover the “business-stealing” effect which arises in our model since the marginal entrants do not internalize the sales that they capture from other firms in the market (Mankiw and Whinston 1986). Comparing the estimated business-stealing effect to the estimated variety effect, our findings indicate that variety is below the socially optimal level.²

Section 2 characterizes the objectives of consumers and firms. Consumers exhibit love-of-variety preferences and we allow them to misoptimize with respect to taxes following Chetty, Looney and Kroft (2009), Taubinsky and Rees-Jones (2018), Farhi and Gabaix (2020), and Kroft et al. (2020a).³ Our model of firm behavior follows Weyl and Fabinger (2013) and spans

²For papers on socially optimal variety, see Spence (1976), Dixit and Stiglitz (1977), Vives (2001), Parenti, Ushchev and Thisse (2017) and Dhingra and Morrow (2019). Relative to this work, we show how to combine our model with reduced-form estimates of the effects of taxes in order to assess whether variety is socially optimal at current tax rates.

³Our results on how tax salience interact with the welfare cost and incidence of taxes in the long run mirror

a wide range of market conduct permitting both price-setting and quantity-setting models of competition using conjectural variations. The equilibrium number of firms is determined by the free-entry condition that profits are equal to zero in the long run. Finally, the government has a revenue requirement and tax revenues (and profits) are rebated to consumers in a lump-sum fashion.

In Sections 3 (specific taxes) and 4 (ad valorem taxes), we derive formulas for the marginal excess burden and pass-through of specific taxation (Proposition 1) and ad valorem taxation (Proposition 2). We show that the marginal excess burden depends on the following reduced-form effects: (1) the causal effect of taxes on output, (2) the causal effect of taxes on producer prices and (3) the causal effect of taxes product variety. The effect of taxes on output is scaled by the baseline tax rate as well as a measure of tax salience. The effect of taxes on prices is scaled by output, and the effect of taxes on product variety is scaled by the variety effect. While the first two terms are standard in models with homogeneous products (with the exception of the salience parameter) (see, for example Besley 1989 and Delipalla and Keen 1992), the last term is new and comes from product differentiation and love-of-variety preferences.⁴ Intuitively, if firms react to taxation by reducing product variety, there will be a negative externality since firms do not take into account consumer surplus at the margin when deciding whether to offer a new product; they only internalize the effects of the tax on profits.

Our pass-through formulas with free entry build on results in Sullivan (1985), Bulow and Pfleiderer (1983), Weyl and Fabinger (2013), Adachi and Fabinger (2017) and Kroft et al. (2020a), who link pass-through rates of taxes to economic fundamentals such as 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. We extend these analyses of pass-through by taking into account the endogeneity of product variety to taxes and present formulas in terms of the same set of economic fundamentals alongside the variety effect.⁵ In the same way that Weyl and Fabinger’s pass-through formula for imperfect competition nests the homogeneous

the results in Kroft et al. (2020a) and so we do not emphasize them as much in this paper.

⁴See de Meza (1982), Seade (1985), Stern (1987), Myles (1987) and Myles (1989) for papers studying specific taxation in the short run (i.e., without allowing for free entry and exit).

⁵Strictly speaking, we show that the pass-through rate for both specific and ad valorem taxation depends on the variety effect, which enters into the marginal excess burden formulas under a parallel demands assumption following Kroft et al. (2020b). This assumption requires that the inverse market demand curve shifts in parallel when new varieties are added or removed from the market. This assumption allows us to empirically implement our welfare formulas and assess whether variety is socially optimal; however, it is not necessary for other exercises such as using the full structure of the model to construct valid estimates of the counterfactual reduced-form effects of a specific tax (such as the pass-through rate).

product “short run” (fixed variety) conjectural analysis of Delipalla and Keen (1992) and the differentiated products Bertrand pricing analysis in Anderson, De Palma and Kreider (2001a), our formula with free entry nests the corresponding “long run” analyses in each of these papers.⁶

Using our new formulas, we characterize the conditions for when introducing a small tax in an untaxed market increases welfare and producer prices (Corollaries 1 and 2 for specific taxation and Corollaries 3 and 4 for ad valorem taxation). For both taxes, we find that under certain conditions taxes can increase welfare.⁷ The logic is that provided that the variety effect is not too large, there will be an excessive number of firms in the market due to business stealing, and taxes can correct this distortion and thus improve welfare – an intuitive application of the theory of second-best. Turning to pass-through, we show that the condition for the overshifting of specific taxes onto consumer prices depends on the curvature of the firm’s own demand, and the condition for the overshifting of ad valorem taxes depends on the curvature of the firm’s own demand, the price elasticity of demand, and the variety effect. We also compare the short-run and long-run effects of taxation on prices (Corollary 3 for specific taxation and Corollary 6 for ad valorem taxation).

Section 5 presents our desirability condition and a comparison of pass-through rates of specific taxes and ad valorem taxes (Proposition 3). For welfare, we find that the desirability condition depends on the variety effect, the elasticity of demand and the conduct parameter. In particular, when the love-of-variety is large relative to the firm’s markup, specific taxation dominates ad valorem taxation. Intuitively, when consumers have a sufficiently strong preference for variety and there is insufficient variety in the market (relative to the social optimum), a tax increase that reduces variety imposes additional efficiency costs. Since ad valorem taxes tend to reduce variety more than specific taxes through what Keen (1998) calls a “multiplier effect”, this implies that specific taxation has a more favorable effect on welfare when consumers have a strong love-of-variety.⁸

Turning to pass-through, the condition for whether specific taxes are shifted more (or

⁶Throughout this paper we use the terminology in Anderson, De Palma and Kreider (2001a) and refer to analyses for a fixed number of varieties as “short run”, and analyses allowing for endogenous number of varieties with respect to taxes as “long run”.

⁷Besley (1989) previously demonstrated this possibility for specific taxation in a model of Cournot quantity competition with homogeneous products.

⁸To see this logic, consider an increase in the fixed cost of production of ΔF and suppose that profits are initially 0. In order for the firm to break even with a specific tax (holding constant demand responses 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 the consumer price by $\Delta p = \Delta F/q(1 - \tau)$. Thus, ad valorem taxation is likely to lead to lower product variety than specific taxation.

less) to prices than ad valorem taxes depends on the variety effect and the price elasticity of demand. If the variety effect is large enough, ad valorem taxes correspond to higher consumer and producer prices than specific taxes, contrasting with the results in Delipalla and Keen (1992) as well as Anderson, De Palma and Kreider (2001*a*). It is noteworthy that our results are different than Anderson, De Palma and Kreider (2001*a*) since they consider a setting with differentiated products and Bertrand competition. We show that this discrepancy comes from the fact that Anderson, De Palma and Kreider (2001*a*) impose an important assumption of business stealing. We show that if one relaxes this restriction and allows entry to be “business augmenting”, then there exist parameter regions where the variety effect can be large enough to reverse the ranking of pass-through across the tax structures (Corollary 7). Put a slightly different way, our results are consistent with Anderson, De Palma and Kreider (2001*a*) when we impose the same assumption of business stealing, but our analysis shows that this restriction is not necessary for stability and – perhaps surprisingly – we show it is exactly this restriction that ends up ruling out parameter regions where ad valorem taxes have both larger efficiency costs and lead to higher consumer and producer prices relative to specific taxes.

In Section 6, we use our tax formulas to organize our empirical analysis, which focuses on estimation of the following reduced-form effects of ad valorem taxation: (1) the causal effect of taxes on pre-tax prices (β^p), (2) the causal effect of taxes on quantity demanded (β^Q), and (3) the causal effect of taxes on product variety (β^J). Using Nielsen Retail Scanner data covering grocery stores selling consumer goods in the U.S. combined with detailed product-level and county-level retail sales tax data, we estimate these causal effects by implementing a cross-sectional “county border pair” research design (Holmes 1998, Dube, Lester and Reich 2010). This design exploits tax rate differentials between taxed and tax-exempt products (including products taxed at reduced rates) across nearby stores located in contiguous counties across state borders and thus controls flexibly for store-level heterogeneity as well as product-level heterogeneity that is common across stores within a county border pair. Our estimates indicate that sales taxes are slightly overshifted onto consumer prices and have a large effect on quantity demanded. We also find that sales taxes meaningfully reduce the variety of products available to consumers, with the magnitude being about one-third of the effect of sales taxes on aggregate quantity demanded.

Lastly, in Section 7, we recover the model parameters, calibrate the welfare formula for ad valorem taxes, and consider several applications. We start with identification of the model parameters. Under the assumption of constant marginal cost, our theoretical formulas for the reduced-form effects of ad valorem taxes on equilibrium prices, market output, and product

variety depend on (1) the price elasticity of demand (ϵ_D), (2) the tax salience parameter (θ_τ), (3) the conduct parameter (v_q/J), (4) the elasticity of inverse marginal surplus (ϵ_{ms}), which governs the curvature of demand, and (5) the variety effect ($\tilde{\Lambda}$). We calibrate $\epsilon_D = 1.223$ and $\theta_\tau = 0.528$ using estimates from Kroft et al. (2020a).⁹ This leaves v_q/J , ϵ_{ms} , and $\tilde{\Lambda}$, which we identify using the model-based expressions for the reduced-form effects of taxes on prices, market output, and variety which are functions of these three parameters. As a result, we can solve for the parameters that make the model exactly match the three corresponding reduced-form estimates ($\beta^p, \beta^Q, \beta^J$).

A key advantage of our identification approach is the transparent connection between the reduced-form empirical estimates and the underlying model parameters. Specifically, we show that the long-run free-entry condition determines the markup in terms of the reduced-form estimates (i.e., the markup is given by $-\beta^p/(\beta^Q - \beta^J)$). We then use the estimated markup to pin down v_q/J , given ϵ_D , through the firm's first-order condition. Next, given v_q/J , we show that the pass-through rate pins down ϵ_{ms} independently of ϵ_D (as in Bulow and Pfleiderer 1983). Lastly, $\tilde{\Lambda}$ is pinned down by the relative magnitude of the reduced-form effects of taxes on variety and the total quantity demand (i.e., β^Q/β^J). Intuitively, when β^Q increases relative to β^J (holding ϵ_D fixed), demand falls a lot for a given change in variety, which in turn reveals that consumers have a high willingness-to-pay for variety.

Turning to our empirical results, we estimate a conduct parameter of $v_q/J = 0.096$ (where 0 is perfect competition and 1 is perfect collusion), which suggests a high degree of competition, as would be expected in our setting of retail grocery stores. We estimate the elasticity of inverse marginal surplus to be $\epsilon_{ms} = -0.935$. According to our theoretical analysis, this parameter must be negative whenever there is overshifting of ad valorem taxes, as we find empirically. Finally, we estimate the variety effect to be $\tilde{\Lambda} = 0.133$, which is modest given that we find a small effect of taxes on variety relative to the effect on total quantity. This can be given a willingness-to-pay interpretation – i.e., an exogenous 10 percent reduction in variety reduces average willingness-to-pay by 1.33 percent.¹⁰ Combining these model parameter estimates with our reduced-form estimates ($\beta^p, \beta^Q, \beta^J$), we calculate a marginal excess burden of ad valorem taxes of about 9 percent of total firm revenues, which is several times larger than a standard Harberger benchmark that is adjusted for tax salience (about 1.3 percent of total

⁹In Kroft et al. (2020a), we identify ϵ_D using a Two-Stage Least Squares estimator, and θ_τ is identified by taking the ratio of the tax elasticity to the price elasticity of demand, adjusted for pass-through of taxes onto consumer prices.

¹⁰One 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 price elasticity of demand. This would imply an estimate of the variety effect of around 0.818.

firm revenues). This estimate is almost twice as large as an alternative benchmark that accounts for imperfect competition but ignores consumers' love-of-variety preferences, which comes from our extension of Besley (1989) to ad valorem taxation.

We conclude with two additional counterfactual exercises that use the full structure of the model. First, we calibrate the business-stealing effect which requires an estimate of the markup and the responsiveness of firm demand to new varieties. Comparing the estimated business-stealing effect to the variety effect allows us to assess whether the privately optimal product variety is insufficient or excessive relative to the socially optimal variety. We find that the variety effect is larger than the business-stealing effect, which implies that variety is lower than the social optimum. The love-of-variety parameter we estimate is about twice the value that would lead us to conclude that variety was socially optimal (holding all other model parameters constant).

Second, we calibrate the marginal cost of public funds and pass-through rates for ad valorem taxes and counterfactual specific taxes, and we illustrate the sensitivity of the incidence and efficiency costs of both types of taxes to the variety effect. Without any love-of-variety, we find less overshifting and lower efficiency costs for ad valorem taxes (relative to a counterfactual specific tax that raises the same revenue). At the actual love-of-variety estimated in our data, we still find less overshifting for ad valorem taxes, but we find that ad valorem leads to higher efficiency costs. This illustrates the practical relevance of the novel theoretical results. Specifically, both our theoretical analysis and the model calibrations point towards the conclusion that in differentiated product markets with meaningful love-of-variety, we should not be surprised to find specific taxes are more efficient than ad valorem taxes.

Our theoretical analysis generalizes and unifies several existing results in the literature on taxation. Our main contribution is to build on the papers which consider the welfare and incidence effects of taxation (specific taxation and ad valorem taxation) in the long-run when entry and exit are allowed. Our key innovation relative to this literature is to consider a theory of taxation that nests various forms of market competition (quantity and price competition) and allows for love-of-variety preferences and tax salience. Our pass-through formulas generalize Bulow and Pfleiderer (1983), Anderson, De Palma and Kreider (2001*a*), Weyl and Fabinger (2013), Adachi and Fabinger (2017) and Kroft et al. (2020*a*). Our welfare analysis builds on Anderson, De Palma and Kreider (2001*b*) who 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.¹¹ To focus on

¹¹Kay and Keen (1983) consider a model of monopolistic competition with Dixit-Stiglitz preferences and

the distortion to product variety, they assume that demand is completely inelastic to taxes and thus, abstract from the “static” distortionary effect of taxation. Our desirability condition accounts for distortions to both output and product variety. Interestingly, we show that when demand is more elastic, specific taxation is more likely to be a preferable form of excise taxation than ad valorem taxation which contrasts with the speculation of Anderson, De Palma and Kreider (2001*b*). Hamilton (2009) considers a specific model of retailer competition and emphasizes the role of *multi-product* firms demonstrating that one can get overshifting of 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 one can get specific taxation to dominate ad valorem taxation even with single product firms depending on the strength of consumers’ preferences for variety.

Our empirical analysis builds on the literature evaluating the reduced-form effects of sales taxes on prices and output (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, Baker, Johnson and Kueng forthcoming, and Kroft et al. 2020*a*). Our main innovation to this literature is to empirically examine the effect of sales taxes on product variety. Our empirical results are consistent with Cawley et al. (2020), who find that a beverage tax in Philadelphia resulted in fewer taxed products available in stores. A second innovation of our paper is to show that one can use the reduced-form effects of ad valorem taxes along with the ad valorem tax formulas to counterfactually estimate the effects of specific taxes – even though we do not actually observe independent variation in specific tax rates in our empirical setting. This broadly relates to Zoutman, Gavrilova and Hopland (2018) who use an insight from taxation theory to show that when a market is perfectly competitive, one can identify both the elasticity of demand and elasticity of supply using a single tax rate. Intuitively, both the reduced-form effects of ad valorem taxes and specific taxes on economic outcomes depend on the same set of economic fundamentals. Provided that one can structurally recover the primitives using variation in one type of tax (ad valorem), one can then simulate the reduced-form effects associated with the other tax (specific). This shows that under the model assumptions, one does not require exogenous variation in both ad valorem taxes and specific taxes to compare their welfare and price effects. This “shortcut” is quite useful in empirical settings where the researcher does not observe a specific tax,

fixed aggregate demand and show that specific taxation can dominate ad valorem taxation as the taste for variety becomes infinitely strong.

but is interested in comparing the efficiency costs and incidence between the alternative tax structures.

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 (Pirttilä 1997). Finally, our analysis assumes that consumers are identical and thus does not take into account distributional considerations. 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.

2 The Model

We consider a differentiated product market (the “inside market”) which is subject to a specific tax t on consumers that applies to each product in the market. 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.

Consumers

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

Following Chetty, Looney and Kroft (2009), consumer demand for product variety j is given by $q_j = q_j(p_1, \dots, p_J, t)$ which is a function of *both* prices and the tax. In the neoclassical full-optimization model, demand depends only on the total tax-inclusive price: $q_j = q_j(p_1 + t, \dots, p_J + t, 0)$. In this case, equal changes in prices and taxes affect demand symmetrically. We allow for tax salience by considering the possibility that $q_j(p_1, \dots, p_J, t) \neq q_j(p_1 + t, \dots, p_J + t, 0)$.

In order to connect our tax formulas to empirically estimable objects, it is necessary to relate observed demand $q_j(p_1, \dots, p_J, t)$ to consumer willingness-to-pay. We begin with the following assumptions which mirror assumptions A1 and A2 in Chetty, Looney and Kroft (2009) and Kroft et al. (2020a).

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

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

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

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

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

Assumption 2 implies that when tax-inclusive prices are fully salient, agents maximize utility. We allow for salience effects by considering the possibility that $q_j(p_1, \dots, p_J, 0) > q_j(p_1, \dots, p_J, t) > q_j(p_1 + t, \dots, p_J + t, 0)$. In what follows, we assume that the “observed” demand function $q_j(\cdot)$ is symmetric and twice differentiable and denote by $q^J(p, t)$ demand corresponding to symmetric prices and J firms: $q^J(p, t) \equiv q_j(p, \dots, p, t)$. We define market demand as $Q(p, t, J) = Jq^J(p, t)$ and the inverse market demand $wtp(Q, J) \equiv P(Q, J) \equiv Q^{-1}(p, 0, J)$ which corresponds to willingness-to-pay when taxes are fully salient. For a fully salient tax, market demand is $Q(p + t, 0, J)$ and inverse market demand is $P(Q, J)$. For a non-salient tax, we define the degree of inattention to the tax, θ_t , as the ratio of the tax elasticity of market demand and the price elasticity of market demand $\theta_t \equiv \frac{\partial Q / \partial t}{\partial Q / \partial p} = \frac{\epsilon_{Dt}}{\epsilon_D}$, where $\epsilon_D \equiv -\frac{\partial Q}{\partial p} \frac{p+t}{Q}$ is the price elasticity of market demand and $\epsilon_{Dt} \equiv -\frac{\partial Q}{\partial t} \frac{p+t}{Q}$ is the tax elasticity of market demand. Under symmetry, Assumption 2 implies that $P(Q, J) = \frac{\partial u}{\partial Q}(Q, J)$ through the first-order condition for Q . Assuming θ_t is constant over (p, t, J) it can be shown that $Q(p, t, J)$

satisfies $wtp(Q(p, t, J), J) = p + \theta_t t$; moreover $\epsilon_D = -\frac{p+t}{Qmwt p(Q)}$ where $mwt p(Q) = \frac{\partial wtp}{\partial Q}(Q, J)$ where we drop the J argument for simplicity and we let $\epsilon_D^* \equiv \frac{p}{p+t}\epsilon_D$. Next, the elasticity of inverse marginal surplus is given by $\epsilon_{ms} \equiv \frac{ms(Q)}{ms'(Q)Q}$, where $ms(Q) \equiv -mwt p(Q)Q$ is marginal consumer surplus. Finally, we define the “variety effect” as the effect of a change in varieties on consumer surplus (holding prices and output constant):

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

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

For some of our results below, we assume that $P(Q, J)$ shifts in parallel when variety J changes.

Assumption 3. *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 3, $P(Q, J)$ 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. (2020b) for discrete choice models, and in the Appendix we present a microfoundation for parallel demands in our continuous choice model.

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. A given 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 :

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

The term ν_p is the conjectural price variation of the other firms' price as a function of p_j .

The first-order condition for p_j is given by:

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

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

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

We assume that $\frac{\partial \pi_j}{\partial p_j}(p_j, p)$ is strict single crossing (from above) in p_j and decreasing in p so that a unique symmetric equilibrium $p(t)$ exists.¹² By letting $\nu_q \equiv \frac{1}{mwp(Q)} \times \frac{1}{\frac{dq_j}{dp_j}} = \frac{1}{mwp(Q)} \times \frac{1}{\frac{\partial q_j}{\partial p_j} + \nu_p \sum_{k \neq j} \frac{\partial q_j}{\partial p_k}}$ we can rewrite the first-order condition as a generalized Lerner index:

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

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.¹³

We treat the number of firms as a continuous variable, a standard procedure in this literature following Seade (1980), Besley (1989) and Delipalla and Keen (1992). In the “long run”, the number of firms $J(t)$ in the symmetric equilibrium is determined by the free-entry condition $\pi_j(p(J, t), J, t) = 0$:

$$p(J(t), t)q^J(p(J(t), t), t) - c(q^J(p(J(t), t), t)) - F = 0 \quad (2)$$

Moreover, we impose assumptions that give uniqueness for the firm pricing decision and entry problem, so that there is a unique solution $J(t)$.

¹²The case of strategic complementarities, where $\frac{\partial \pi_j}{\partial p_j}(p_j, p)$ is increasing in p allows for the existence of multiple symmetric equilibria. However, in that case if we assume there is a continuous and symmetric equilibrium selection $p(t)$, the same results follow.

¹³The 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.

Assumption 4. We require that $\pi_j(p_j, p_{-j}, J, t)$ is concave in p_j and decreasing in J . We also assume that the conduct parameter ν_q is constant, which implies $\frac{d\nu_q}{dt} = 0$. Let¹⁴

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

The two stability conditions for the firms' problem 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 $\epsilon_D^* = \frac{p(1+\theta_\tau\tau)}{p(1+\tau)+t} \epsilon_D$.¹⁵

Lastly, we define the long-run demand as $Q_L(t) = Q(p(t), t, J(t))$.¹⁶ The effect of taxes on long-run demand, taking into account the endogeneity of variety to taxes, is given by $\frac{dQ_L}{dt} = \frac{\partial Q}{\partial t} + \frac{\partial Q}{\partial J} \frac{dJ}{dt} + \frac{\partial Q}{\partial p} \frac{dp}{dt}$.

3 Welfare, Pass-through and Entry Effects of a Specific Tax

In this section, we consider 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 extend previous theoretical results on taxation with imperfect competition in two ways: first, we allow for “behavioral” agents that need not optimize with respect to taxes following Chetty, Looney and Kroft (2009), Taubinsky and Rees-Jones (2018), Farhi and Gabaix (2020), and Kroft et al. (2020a); second, we allow taxes to affect product variety through the free-entry decision of firms following Besley (1989) and Delipalla and Keen (1992). In this section, we first present results for the general case and then consider several specialized cases, in order to connect our results with the literature on taxation.

We assume throughout that tax revenue $R = tQ$ and profits $J\pi$ are redistributed to the representative consumer as a lump-sum transfer. As is standard in this literature, the consumer treats profits and tax revenue as fixed when choosing consumption, failing to consider the external effects on the lump-sum transfer. Given the assumption of quasilinear utility,

¹⁴Under Assumption 3, $\Delta = \left[2 - \frac{\nu_q}{J} + \frac{\epsilon_D^* - \frac{\nu_q}{J}}{\epsilon_S \frac{\nu_q}{J}} + \frac{\frac{\nu_q}{J}}{\epsilon_{ms}} \right] - \frac{\Lambda \epsilon_D}{(p(1+\tau)+t)q} \left(1 + \frac{\epsilon_D^* - \frac{\nu_q}{J}}{\epsilon_S \frac{\nu_q}{J}} + \frac{1}{\epsilon_{ms}} \right)$.

¹⁵Here, Δ and ϵ_D^* are written in the general form that depends on the specific tax rate t and the ad valorem tax rate τ for convenience, anticipating our results in Sections 4 and 5. The ad valorem tax rate τ and corresponding salience parameter θ_τ are introduced below, but at this point can be shut down by assuming $\tau = 0$ and $\theta_\tau = 1$.

¹⁶Here we use the assumption that for each tax t there is a unique symmetric price equilibrium $p(t)$ where $J(t)$ firms enter the market. The notation Q_L serves to mathematically differentiate the functions $Q(\cdot, \cdot, \cdot) : \mathbb{R}^3 \rightarrow \mathbb{R}$ and $Q_L(t) : t \rightarrow Q(p(t), t, J(t))$.

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):

$$W(p, t, J) = \underbrace{u(Q, J) - (p + t)Q}_{CS} + \underbrace{pQ - Jc\left(\frac{Q}{J}\right)}_{J\pi} - JF + \underbrace{tQ}_R \quad (3)$$

Consider a small increase in the tax starting from t_0 . Under Assumptions 1 and 2, a first-order approximation to the marginal excess burden of taxation is:

$$\frac{dW(p(t), t, J(t))}{dt} = \underbrace{(\theta_t t_0 + p_0 - c'(q_0)) \frac{dQ_L}{dt}}_{\text{Quantity effect}} + \underbrace{(\Lambda_0 + \pi_0 - [p_0 - c'(q_0)]q_0) \frac{dJ}{dt}}_{\text{Diversity effect}} \quad (4)$$

where $p_0, q_0, Q_0, J_0, \pi_0, \Lambda_0$ are all variables evaluated at the equilibrium corresponding to t_0 .¹⁷

Equation (4) shows that the marginal excess burden of taxation is a combination of two main 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 $p_0 + \theta_t t_0$ and the social marginal cost $c'(q_0)$. With no pre-existing taxes ($t_0 = 0$) or when $\theta_t = 0$, the first term depends only on the markup which represents a 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 becomes $(\Lambda_0 - F) \frac{dJ}{dt}$. Thus, whether the change in variety induced by taxes lowers (increases) welfare along this margin depends on whether the love-of-variety exceeds (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 $t_0 = 0$, $p_0 = c'(q^*)$ and $\Lambda_0 = -\pi_0$.

Equation (4) nests canonical formulas in public economics for: *homogeneous or differentiated products, perfect or imperfect competition, fixed or endogenous variety, fully optimizing or behavioral agents*. First, in the case of perfect competition where $p = c'(q)$ and J is fixed, equation (4) reduces to $\frac{dW}{dt} = \theta_t t_0 \frac{dQ_L}{dt}$. The simple case with fully optimizing agents ($\theta_t = 1$) corresponds to the classic analysis by Harberger (1964), while the extension to behavioral agents corresponds to the analysis in Chetty, Looney and Kroft (2009). Next, Auerbach and Hines (2001) consider a model of homogeneous products ($\Lambda_0 = 0$) with imperfect competition

¹⁷See Appendix for proof.

and fixed variety ($\frac{dJ}{dt} = 0$) and fully optimizing consumers ($\theta_t = 1$). In this case, the marginal excess burden of taxation is given by $\frac{dW}{dt} = (p_0 + t_0 - c'(q_0))\frac{dQ_L}{dt}$. This formula also holds in the case of *second-best variety*: a central planner chooses J optimally considering that pricing decisions are left to firms.¹⁸

In the case where goods are homogeneous ($\Lambda_0 = 0$), consumers optimize ($\theta_t = 1$) and 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 $\frac{dW}{dt} = (p_0 + t_0 - c_0)\frac{dQ_L}{dt} - F\frac{dJ}{dt}$.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

Clearly, for taxes to increase welfare, it must be the case that there is a market failure in the baseline equilibrium. The market failure in this case arises from non-convexities due to fixed costs. We can also connect the market failure to the concept of externalities by re-expressing equation (4) 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 $\frac{dW}{dt} = (p_0 + \theta_t t_0 - c'(q_0))J_0\frac{dq_L}{dt} + (\Lambda_0 + \theta_t t_0 q_0 + \pi_0)\frac{dJ}{dt}$. It is noteworthy that 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. Interestingly, this raises the possibility that taxes can *increase* social welfare. To see the connection to business stealing and inefficient entry, we note that $\frac{dq_L}{dt} = \frac{\partial q}{\partial J}\frac{dJ}{dt} + \frac{dq}{dt}|_J$ where $\frac{dq}{dt}|_J \equiv \frac{\partial q}{\partial p}\frac{dp}{dt} + \frac{\partial q}{\partial t}$. Assuming for simplicity that $\pi_0 = t_0 = 0$, we may decompose the formula for marginal excess burden as:

$$\frac{dW}{dt} = \underbrace{\left(J_0(p_0 - c'(q_0))\frac{\partial q}{\partial J} + \Lambda_0 \right) \frac{dJ}{dt}}_{\leq 0} + \underbrace{(p_0 - c'(q_0))J\frac{dq}{dt}|_J}_{< 0} \quad (5)$$

¹⁸The 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:

$$\frac{dW}{dt} = \left(\frac{\partial W}{\partial Q}\frac{dQ}{dJ} + \frac{\partial W}{\partial J} \right) \frac{dJ}{dt} + \frac{\partial W}{\partial Q}\frac{dQ_L}{dt} = \frac{\partial W}{\partial Q}\frac{dQ_L}{dt} = (p + \theta_t t - c'(q))\frac{dQ_L}{dt}$$

¹⁹Besley (1989) assumes Cournot competition but the formula is valid for other types of competition through the conduct parameter.

²⁰One can also show that $\frac{dW}{dt} = \theta_t t_0 \frac{dQ_L}{dt} + (p_0 - c'(q_0))J\frac{dq_L}{dt}$.

The first term in parentheses in equation (5) is the marginal welfare gain of an 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 negative. Note that this is an application of the theory of second-best. There is a market distortion due to excessive firm entry in the economy without taxes and introducing a tax, which is a second type of distortion, counteracts the first distortion leading potentially to a more efficient outcome. If the business-stealing effect is sufficiently strong so that $\frac{dq_L}{dt} > -\frac{\Lambda_0}{J_0(p_0 - c'(q_0))} \frac{dJ}{dt} > 0$, then taxes can actually increase welfare. A necessary condition is for firm-level output to rise in response to an increase in the tax so that the “competition effect” due to fewer firms dominates the standard distortionary effect of taxation.

In practice, equations (4) or (5) 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 now show how one may exploit the long-run free-entry condition to provide a remarkably simpler representation for the marginal excess burden that maps more easily to empirically estimable objects. We begin with the following lemma.

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

$$\frac{\varepsilon_{p,t}}{\varepsilon_{q,t}} = -\frac{p - c'(q)}{p} \quad (6)$$

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

Proof. See Appendix. □

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

We now restate the tax formula in equation (4) using Lemma 1. It is first useful to define the pass-through rate as $\rho_t \equiv 1 + \frac{dp}{dt}$.

Proposition 1. *Assume $\nu_q \in (0, J]$. Under Assumptions 1, 2, 3 and Lemma 1, the marginal excess burden, pass-through and the entry effects of a small change in the specific tax t are*

given respectively by²¹

$$\frac{dW}{dt} = \Lambda_0 \frac{dJ}{dt} - Q_0 \frac{dp}{dt} + \theta_t t_0 \frac{dQ_L}{dt} \quad (7)$$

$$\rho_t = 1 - (1 - \omega_t) \theta_t \quad (8)$$

$$\frac{dJ}{dt} = -\frac{\theta_t J \epsilon_D}{p_0 + t_0} \left[\frac{1 + \frac{\epsilon_D^* - \frac{\nu_q}{J_0}}{\frac{\nu_q}{J_0} \epsilon_S} + \frac{1}{\epsilon_{ms}}}{\Delta} \right] \quad (9)$$

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

Proof. See Appendix. □

Equation (7) 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. (2020a) to the case of love-of-variety preferences and endogenous entry with imperfect competition. In order to develop intuition for this formula, consider a tax decrease as represented in Figure 1. Let (p, Q) be the starting equilibrium producer price and aggregate quantity for a specific tax $t > 0$ and note that if the tax is not fully salient, then $\theta_t \in (0, 1)$. In this case, the willingness-to-pay for the marginal aggregate unit is $wtp(Q) = p + \theta_t t$. For the purpose of this example, assume $\frac{\partial \pi}{\partial t} < 0$, so that when the tax disappears increasing profits drives the entry of new firms $J' > J$, which in turn shifts the aggregate demand curve out and a new equilibrium is reached at (p', Q') . Note that before and after the tax change the long-run free-entry condition is satisfied and so the average total cost is equal to the producer price in each case. Additionally, note that welfare is given by the area between below aggregate demand and above equilibrium price up to the equilibrium quantity. Finally, the difference in welfare between the two scenarios can be represented by the sum of three terms: the area between the inverse aggregate demand curves (Λ_0), an additional trapezoid with base $Q' - Q$ and height $\theta_t t$ ($\theta_t t_0 \frac{dQ_L}{dt}$), minus the price effect $(p' - p)Q'$ ($Q_0 \frac{dp}{dt}$). Thus, this graphical representation of the marginal welfare gain connects with equation (7).

To develop intuition for our other expressions, we consider several special cases. We begin by considering the case where $\theta_t = 1$, $\Lambda_0 = 0$ and $t_0 = 0$ in order to connect our results to the literature.

²¹While the welfare formula $\frac{dW}{dt}$ does not depend on Assumption 3, without Assumption 3 the term Λ_0 that appears in the formulas for ρ_t and $\frac{dJ}{dt}$ would have to be substituted by $Q * \left(\frac{\partial P}{\partial J} + \frac{\partial^2 P}{\partial J \partial Q} q \nu_q \right)$. See Appendix.

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

$$\frac{dW}{dt} = -Q_0 \frac{dp}{dt} \quad (10)$$

$$\rho_t = \frac{2 + \frac{\epsilon_D^* - \frac{\nu_q}{J_0}}{\epsilon_S \frac{\nu_q}{J_0}}}{2 - \frac{\nu_q}{J_0} + \frac{\epsilon_D^* - \frac{\nu_q}{J_0}}{\epsilon_S \frac{\nu_q}{J_0}} + \frac{\frac{\nu_q}{J_0}}{\epsilon_{ms}}} \quad (11)$$

$$\frac{dJ}{dt} = -\frac{J\epsilon_D}{p_0} \left[\frac{1 + \frac{\epsilon_D^* - \frac{\nu_q}{J_0}}{\epsilon_S} + \frac{1}{\epsilon_{ms}}}{2 - \frac{\nu_q}{J_0} + \frac{\epsilon_D^* - \frac{\nu_q}{J_0}}{\epsilon_S \frac{\nu_q}{J_0}} + \frac{\frac{\nu_q}{J_0}}{\epsilon_{ms}}} \right] \quad (12)$$

Proof. See Appendix. □

Corollary 1 shows that the welfare cost of taxation depends only on the price effect, as in Besley (1989) and Delipalla and Keen (1992).²² 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. Thus 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.

Our formula in Proposition 1 generalizes the marginal excess burden formula in these papers to allow for a love-of-variety and pre-existing taxes. When consumers have a preference for variety, there is an additional effect on consumer surplus since varieties are affected by the tax change. Intuitively, when firms decide to enter (or exit) the market, they do not internalize the positive effects of product creation (or product destruction) on consumer surplus. Thus, there is a new term that depends on the variety effect and the responsiveness of varieties to a change in taxes. Finally, when there are pre-existing taxes, one must account for the fiscal externality on government revenue.

Turning to pass-through, while Besley (1989) and Delipalla and Keen (1992) considered

²²Note 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.

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. (2020a).²³ Our contribution relative to these papers is to better highlight the economic forces by connecting the expressions for pass-through and entry effects to intuitive concepts such as the conduct parameter, the elasticity of demand and supply, and the elasticity of the inverse marginal surplus (Corollary 1), as well as generalize by allowing for differentiated products and a love-of-variety (Proposition 1). Our general formula in Proposition 1 indicates that a lower θ_t leads to a lower incidence on consumers so long as demand is log concave as in Kroft et al. (2020a). When $\theta_t = 0$, we see that $\rho_t = 1$ so that incidence is full on consumers.

Finally, we see that the effects of taxes on entry/exit depend on the price elasticity of market demand. 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. This is the reason that the variety effect Λ_0 enters the pass-through formula in (8) scaled by ϵ_D in Proposition 1. 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 $\frac{\nu_q}{J_0} + \frac{\epsilon_D^* - \frac{\nu_q}{J_0}}{\epsilon_S} + \frac{\frac{\nu_q}{J_0}}{\epsilon_{ms}} > 0$. We show below 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.

Next, we derive conditions for when taxes increase welfare and lead to overshifting or undershifting in a model with a love-of-variety.

Corollary 2. *Assume $\nu_q \in (0, J]$, $\theta_t \in [0, 1]$, and $\pi_0 = 0$.*

1. *Overshifting: a small specific tax increases producer prices if and only if:*

$$\begin{aligned} \frac{dp}{dt} &\geq 0 \\ \Leftrightarrow 1 - \frac{1}{\epsilon_{ms}} &\geq 0 \end{aligned}$$

2. *Starting from no tax $t_0 = 0$, introducing a small specific tax, increases welfare if and*

²³In the Appendix, we formally establish the connection between the pass-through formulas in our paper and Delipalla and Keen (1992).

only if:

$$\begin{aligned}\frac{dW}{dt} \geq 0 &\Leftrightarrow \frac{\Lambda_0}{Q_0} \frac{dJ}{dt} \geq \rho_t - 1 \\ &\Leftrightarrow \frac{1}{\epsilon_{ms}} - 1 \geq \frac{\Lambda_0 \epsilon_D}{p_0 q_0 \frac{\nu_q}{J_0}} \left[1 + \frac{1}{\epsilon_{ms}} + \frac{\epsilon_D^* - \frac{\nu_q}{J_0}}{\epsilon_S \frac{\nu_q}{J_0}} \right]\end{aligned}$$

3. Therefore, if $\Lambda_0 = 0$, starting from no tax $t_0 = 0$, introducing a small specific tax, increases welfare if and only if there is no overshifting:

$$\frac{dW}{dt} \geq 0 \Leftrightarrow \frac{dp}{dt} \leq 0 \Leftrightarrow \frac{1}{\epsilon_{ms}} - 1 \geq 0$$

Proof. See Appendix. □

In Besley (1989) and Delipalla and Keen (1992), the condition for whether $\frac{dW}{dt} \geq 0$ is $\frac{1}{\epsilon_{ms}} - 1 \geq 0$. Thus, we see that the taste for variety Λ_0 now enters into the generalized condition. Intuitively, when Λ_0 is large, there is likely to be insufficient entry in the market and a tax that causes less variety lowers welfare further. On the other hand, the condition for overshifting is unchanged when there are differentiated products and a taste for variety. For more intuition on the overshifting condition, note that the Lerner condition from the firms' optimization can be used to interpret the elasticity of marginal surplus as follows: $\nu_q mwp(Q) = \frac{1}{\frac{\partial q}{\partial p_i}}$ implies $1 - \frac{1}{\epsilon_{ms}} = \nu_q Q * mwp^2(Q) \frac{\partial^2 q}{\partial p \partial p_i}$ where $\frac{\partial^2 q}{\partial p \partial p_i}$ is the cross-partial derivative of individual firm demand with respect to individual price and market price. Therefore, $1 - \frac{1}{\epsilon_{ms}}$ is a measure of the degree of "strategic complementarity" in the oligopoly pricing game. As such, we can expect that the stronger the strategic complementarity, the stronger the multiplier effect of an increase in taxes due to each firm increasing its own price as a function of taxes and as a result of each of the other firms increasing their prices.

The pass-through formula contained in Proposition 1 extends results in Kroft et al. (2020a) and Weyl and Fabinger (2013) to the case of free entry. Our final corollary compares the short-run pass-through formula in Kroft et al. (2020a) to the long-run pass-through formula in Proposition 1 and highlights the key economic forces differentiating them. Note in this case:

$$\rho_t - \rho_t^{SR} = \frac{\partial p}{\partial J} \frac{dJ}{dt}$$

where we define $\rho_t^{SR} \equiv 1 - (1 - \omega_{SR}) \theta_t$ as the pass-through of a specific tax in the short run

and $\omega_{SR} \equiv \frac{1}{1 + \frac{\epsilon_D^* - \frac{\nu_q}{J}}{\epsilon_S} + \frac{\nu_q}{J\epsilon_{ms}}}$ as the pass-through of a specific tax in the short run when consumers fully optimize (Kroft et al. 2020a). The following Lemma is useful for comparing the short-run and long-run effects of taxation on prices.

Lemma 2. *The effect of competition on prices and output is given respectively by:*

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

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 ($\frac{\partial q}{\partial J} < 0$) whenever $\frac{\Lambda\epsilon_D}{(p+t)q} + \frac{\nu_q}{J} \left(1 - \frac{1}{\epsilon_{ms}} \right) < 1$.

One can see in Lemma 2 that there is a connection between business stealing and pass-through of a specific tax. This relates to results in Weyl and Fabinger (2013) who demonstrate that when there is a per-firm exogenous quantity that enters the market, the increase is equivalent to a reduction in cost (or increase in the tax).²⁴ We also see from Lemma 2 that, conditional on the business-stealing effect, a stronger variety effect implies that adding new varieties to the market leads to a greater increase in prices. Intuitively, new varieties raise the marginal utility of existing products and thus increase the willingness-to-pay for these products. On the other hand, there is an indirect effect which works in the opposite direction as a stronger variety effect lowers business stealing with the strength of this effect depending on pass-through of specific taxes. Thus, a larger Λ can increase or decrease $\frac{\partial p}{\partial J}$ and consequently increase or decrease long-run pass-through.

Corollary 3. *Entry and short-run versus long-run pass-through:*

Assume $\epsilon_D > 0$. The following three conditions are equivalent:

$$\frac{dJ}{dt} < 0 \iff \frac{\partial \pi}{\partial t} < 0 \iff \frac{\nu_q}{J} + \frac{\epsilon_D^* - \frac{\nu_q}{J}}{\epsilon_S} + \frac{\frac{\nu_q}{J}}{\epsilon_{ms}} > 0$$

Moreover, assume constant marginal cost ($\epsilon_S = \infty$) and assume that the variety effect Λ is bounded by $\frac{1}{\epsilon_{ms}} \frac{\Lambda\epsilon_D}{(p+t)q} < 1$ so that $\frac{\partial p}{\partial J} < 0$ holds by Lemma 2. Then

$$\frac{\partial \pi}{\partial t} < 0 \iff \rho_t > \rho_t^{SR}$$

²⁴In Weyl and Fabinger (2013), the thought experiment is to increase the quantity of each of the goods produced by the firms in the market; thus, there is no new variety. In our case, there is a new variety that is introduced which is why the variety effect is present. When Λ is larger, there is less business stealing; intuitively, a stronger love-of-variety implies that goods are less substitutable and dampens the business-stealing effect.

Proof. See Appendix. □

Corollary 3 shows that when taxes reduce profits, there are fewer varieties in the market and this acts to increase prices further due to a competition effect.²⁵

4 Welfare, Pass-through and Entry Effects of an Ad Valorem Tax

In this section, we consider the marginal welfare gain, pass-through and entry effects associated with a small increase in the ad valorem tax τ . The pre-tax or producer price for product j is still p_j and the post-tax or consumer price is now given by $p_j(1 + \tau)$ for all $j = 1, \dots, J$. We continue to assume that tax revenue and profits are redistributed to the representative consumer as a lump-sum transfer.

Analogous to the specific tax case, consumer demand for product variety j is given by $q_j = q_j(p_1, \dots, p_J, \tau)$ which is a function of *both* prices and the commodity tax and we allow for tax salience by considering the possibility that $q_j(p_1, \dots, p_J, \tau) \neq q_j(p_1(1 + \tau), \dots, p_J(1 + \tau), 0)$. We assume that the “observed” demand function $q_j(\cdot)$ is symmetric and twice differentiable and denote by $q^J(p, \tau)$ demand corresponding to symmetric prices and J firms: $q^J(p, \tau) \equiv q_j(p, \dots, p, \tau)$. Market demand is given by $Q(p, \tau, J) = Jq^J(p, \tau)$ and the willingness-to-pay for unit Q when variety is J is the inverse market demand $wtp(Q, J) \equiv P(Q, J) \equiv Q^{-1}(p, 0, J)$. We define the degree of inattention to the ad valorem tax, $\theta_\tau \equiv \frac{\frac{\partial Q(p, 0, J)}{\partial \tau}}{\frac{\partial Q(p, 0, J)}{\partial p}} \times \frac{1}{p}$, as the ratio of the demand responses to the tax and price starting from $\tau = 0$, normalized by p . Assuming θ_τ is constant over (p, τ, J) it can be shown that $Q(p, \tau, J)$ satisfies $wtp(Q(p, \tau, J), J) = p(1 + \theta_\tau \tau)$; moreover, we define $\epsilon_D \equiv -\frac{p(1+\tau)}{Qmwtp(Q)}$ where $mwtp(Q) = \frac{\partial wtp}{\partial Q}(Q, J)$ where we drop the J variable for simplicity and we let $\epsilon_D^* \equiv \frac{p(1+\theta_\tau \tau)}{p(1+\tau)}\epsilon_D$. Finally, as before, $\epsilon_{ms} \equiv \frac{ms(Q)}{ms'(Q)Q}$, where $ms(Q) \equiv -mwtp(Q)Q$ is marginal consumer surplus. In the interest of space, we do not characterize Assumptions 1 and 2 for the case of an ad-valorem tax but the exact same logic

²⁵It is also useful to compare the long-run overshifting condition $1 - \frac{1}{\epsilon_{ms}} \geq 0$ with the short-run overshifting condition $0 \geq \frac{1}{\epsilon_{ms}}$ that applies in the simple case with constant marginal cost $\epsilon_S = \infty$ and full salience $\theta_t = 1$. A corollary of this, is that overshifting in the short run implies overshifting in the long run. In a similar model, Anderson, De Palma and Kreider (2001a) provide formulas for overshifting in the short run and long run. To obtain their Proposition 2 (short run) for instance, one can substitute $\epsilon_{DD} = -\epsilon_D$, $1 - \tilde{E} = \frac{1}{\epsilon_{ms}}$ and $\frac{\nu_q}{\epsilon_D} = -\frac{1}{\epsilon_{dd}}$. However a translation of their overshifting long-run condition is not as straightforward; moreover our overshifting conditions in Corollaries 2 and 5 are more easily comparable to the analogous short-run overshifting conditions.

holds replacing $q_j(p_1, \dots, p_J, t)$ with $q_j(p_1, \dots, p_J, \tau)$ and $q_j(p_1 + t, \dots, p_J + t, 0)$ with $q_j(p_1(1 + \tau), \dots, p_J(1 + \tau), 0)$.

On the supply side, using the short hand $\nu_q = \frac{dQ}{dq}$ and $p = \frac{wtp(Q)}{1 + \theta_\tau \tau}$, we can differentiate $\pi = pq - c(q) - F$ to get the analogous first-order condition as a generalized Lerner index:

$$\frac{p - mc(q)}{p(1 + \tau)} = \frac{\frac{\nu_q}{J}}{(1 + \theta_\tau \tau)\epsilon_D} \quad (13)$$

Similar to the previous section, define $\epsilon_S \equiv \frac{c'(q)}{c''(q)q}$. Total welfare, W , with an ad valorem tax is given by:

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

Consider a small increase in the ad valorem tax rate starting from τ_0 . Under the analogous Assumptions 1 and 2 for ad valorem taxation, a first-order approximation to the marginal excess burden of ad valorem taxation is:

$$\frac{dW}{d\tau} = \underbrace{(p_0(1 + \theta_\tau \tau_0) - c'(q_0)) \frac{dQ_L}{d\tau}}_{\text{Quantity effect}} + \underbrace{(\Lambda_0 + \pi_0 - [p_0 - c'(q_0)]q_0) \frac{dJ}{d\tau}}_{\text{Diversity effect}} \quad (14)$$

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

Lemma 3. *In the long run (when the free-entry condition (2) is satisfied), for any tax rate τ , the following envelope condition holds:*

$$\frac{\varepsilon_{p,\tau}}{\varepsilon_{q,\tau}} = -\frac{p - c'(q)}{p} \quad (15)$$

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

Proof. See Appendix. □

Next, we characterize welfare and pass-through by again exploiting the long-run free-entry condition in Lemma 2. In order to do so, we first define the pass-through for an ad valorem tax as $\rho_\tau \equiv \frac{1}{p} \frac{d(p(1 + \tau))}{d\tau} = 1 + \frac{1 + \tau}{p} \frac{dp}{d\tau}$.

Proposition 2. *Assume $\nu_q \in (0, J]$. Under the analogous Assumptions 1 and 2 for ad valorem taxation, and Assumption 3, the marginal excess burden, pass-through and the entry effects of*

a small change in the ad valorem tax τ are given respectively by:²⁶

$$\frac{dW}{d\tau} = \Lambda_0 \frac{dJ}{d\tau} - Q_0 \frac{dp}{d\tau} + \theta_\tau \tau_0 p_0 \frac{dQ_L}{d\tau} \quad (16)$$

$$\rho_\tau = 1 - (1 - \omega_\tau) \frac{(1 + \tau_0)\theta_\tau}{1 + \theta_\tau \tau_0} \quad (17)$$

$$\frac{1}{p_0} \frac{dJ}{d\tau} = - \frac{\theta_\tau J \epsilon_D}{(1 + \tau_0)p_0} \left[\frac{1 + \frac{\epsilon_D^* - \frac{\nu_q}{J_0}}{\epsilon_S} + \frac{1}{\epsilon_{ms}} + \frac{1 - \frac{\nu_q}{J_0}}{\epsilon_D^*}}{\Delta} \right] \quad (18)$$

where $\omega_\tau = \frac{\Delta + \frac{\nu_q}{J_0} \left[1 - \frac{1}{\epsilon_{ms}} + \frac{\Lambda_0}{(1 + \theta_\tau \tau_0)p_0 q_0} - \frac{1}{\epsilon_D^*} \right]}{\Delta}$ is the pass-through formula when there is full optimization ($\theta_\tau = 1$).

Proof. See Appendix.

The formulas derived in Proposition 2 are straightforward extensions of the case of specific taxes. We only briefly interpret the formulas here and discuss differences between specific taxation and ad valorem taxation in the next section. Before we do that, we present the main corollaries that follow the special cases considered in Section 3. $\square$

Corollary 4. Consider the case of full-optimization ($\theta_\tau = 1$), homogeneous products ($\Lambda_0 = 0$) and no pre-existing taxes ($\tau_0 = 0$). The marginal excess burden and pass-through formulas are given respectively by:

$$\frac{dW}{d\tau} = -Q_0 \frac{dp}{d\tau} \quad (19)$$

$$\rho_\tau = \frac{2 + \frac{\epsilon_D^* - \frac{\nu_q}{J_0}}{\epsilon_S} - \frac{\frac{\nu_q}{J_0}}{\epsilon_D^*}}{2 - \frac{\nu_q}{J_0} + \frac{\epsilon_D^* - \frac{\nu_q}{J_0}}{\epsilon_S} + \frac{\frac{\nu_q}{J_0}}{\epsilon_{ms}}} \quad (20)$$

$$\frac{1}{p_0} \frac{dJ}{d\tau} = - \frac{J \epsilon_D}{p_0} \left[\frac{1 + \frac{\epsilon_D^* - \frac{\nu_q}{J_0}}{\epsilon_S} + \frac{1}{\epsilon_{ms}} + \frac{1 - \frac{\nu_q}{J_0}}{\epsilon_D^*}}{2 - \frac{\nu_q}{J_0} + \frac{\epsilon_D^* - \frac{\nu_q}{J_0}}{\epsilon_S} + \frac{\frac{\nu_q}{J_0}}{\epsilon_{ms}}} \right] \quad (21)$$

Proof. See Appendix. $\square$

We see that the pass-through expression in Corollary 3 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

²⁶While the welfare formula $\frac{dW}{d\tau}$ does not depend on Assumption 3, without Assumption 3 the term Λ that appears in the formulas for ρ_τ and $\frac{dJ}{d\tau}$ would have to be substituted by $Q * \left(\frac{\partial P}{\partial J} + \frac{\partial^2 P}{\partial J \partial Q} q \nu_q \right)$. See Appendix.

$\rho_\tau < \rho_t$ whenever $\Lambda_0 = 0$. This is a standard result in the 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 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 does not go to the government. Thus, firms face less of an incentive to increase prices at the margin under an ad valorem tax. Note that unlike the specific tax case, even if demand is inelastic so that $\epsilon_D = 0$, we can still have $\frac{dJ}{d\tau} < 0$. This is the reason that there is an additional Λ_0 term in the numerator of ω_τ compared to ω_t where it only enters through Δ_t and is always scaled by ϵ_D . Thus, ad valorem taxes have stronger effects on entry and exit and we will see below can lead to greater long-run effects on prices than specific taxes.

Now we characterize the conditions under which a small change in the ad valorem tax increases welfare and producer prices.

Corollary 5. *Assume $\nu_q \in (0, J]$, $\theta_\tau \in [0, 1]$, and $\pi_0 = 0$.*

1. *Overshifting: a small tax increases producer prices if and only if:*

$$\begin{aligned} \frac{dp}{d\tau} &\geq 0 \\ \Leftrightarrow 1 - \frac{1}{\epsilon_{ms}} &\geq \frac{1}{\epsilon_D^*} - \frac{\Lambda_0}{p_0 q_0 (1 + \theta_\tau \tau_0)} \end{aligned}$$

2. *Starting from no tax $\tau_0 = 0$, introducing a small specific tax, increases welfare if and only if:*

$$\begin{aligned} \frac{dW}{d\tau} \geq 0 &\Leftrightarrow \frac{\Lambda_0}{p_0 Q_0} \frac{dJ}{d\tau} \geq \rho_\tau - 1 \\ &\Leftrightarrow \frac{1}{\epsilon_D^*} + \frac{1}{\epsilon_{ms}} - 1 \geq \frac{\Lambda_0 \epsilon_D}{p_0 q_0 \frac{\nu_q}{J_0}} \left[1 + \frac{1}{\epsilon_D^*} + \frac{1}{\epsilon_{ms}} + \frac{\epsilon_D^* - \frac{\nu_q}{J_0}}{\epsilon_S \frac{\nu_q}{J_0}} \right] \end{aligned}$$

3. *Therefore, if $\Lambda_0 = 0$, starting from no tax $\tau_0 = 0$, introducing a small tax, increases welfare if and only if there is no overshifting:*

$$\frac{dW}{d\tau} \geq 0 \Leftrightarrow \frac{dp}{d\tau} \leq 0 \Leftrightarrow \frac{1}{\epsilon_D^*} + \frac{1}{\epsilon_{ms}} - 1 \geq 0$$

Proof. See Appendix. □

The overshifting condition for ad valorem taxes shows that when $\Lambda_0 = 0$, producer prices increase and welfare is reduced whenever $1 - \frac{1}{\epsilon_{ms}} \geq \frac{1}{\epsilon_D^*}$ which is the same condition as in

Delipalla and Keen (1992) and Anderson, De Palma and Kreider (2001a). Note also that in this case, if $\frac{dp}{d\tau} \geq 0$, then $\frac{dp}{dt} \geq 0$ since $\epsilon_D^* > 0$. While it might appear at first glance from our condition that overshifting is possible when demand log concave (for sufficiently large values of Λ_0), infact this is not true. The reason is that the parameter region featuring overshifting $\left(1 - \frac{1}{\epsilon_{ms}} \geq \frac{1}{\epsilon_D^*} - \frac{\Lambda_0}{p_0 q_0 (1 + \theta_\tau \tau_0)}\right)$ and log-concave demand $\left(\frac{1}{\epsilon_{ms}} - 1 \geq 0\right)$ violates the stability condition $\Delta > 0$. Interestingly, in the general case, our results show that there are scenarios where $\frac{dp}{d\tau} \geq 0$, but $\frac{dp}{dt} \leq 0$ for sufficiently large Λ_0 . This is in contrast to the results in Anderson, De Palma and Kreider (2001a) which we elaborate on below.

The pass-through formula contained in Proposition 2 extends results in Kroft et al. (2020a) and Weyl and Fabinger (2013) to the case of free-entry. Our final corollary compares the short-run pass-through formula for an ad valorem tax in Kroft et al. (2020a) to the long-run pass-through formula in Proposition 2 and highlights the key economic forces differentiating them.²⁷ For concreteness, we define $\rho_\tau^{SR} \equiv 1 - \left(1 - \omega_{SR} \frac{mc}{p}\right) \theta_\tau$ as the pass-through of an ad valorem tax in the short run.

Lemma 4. *The effect of competition on prices and output is given respectively by:*

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

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

Proof. See Appendix. □

Corollary 6. *Entry and short-run versus long-run pass-through:*

The following three conditions are equivalent:

$$\frac{dJ}{d\tau} < 0 \iff \frac{\partial \pi}{\partial \tau} < 0 \iff 1 + \frac{\epsilon_D^* - \frac{\nu_q}{J}}{\epsilon_S \frac{\nu_q}{J}} + \frac{1}{\epsilon_{ms}} + \frac{1 - \frac{\nu_q}{J}}{\epsilon_D^*} > 0$$

Moreover, assume constant marginal cost ($\epsilon_S = \infty$) and assume that the the variety effect

²⁷It is also useful to compare the long-run overshifting condition $1 + \frac{\Lambda_0}{p_0 q_0 (1 + \theta_\tau \tau_0)} \geq \frac{1}{\epsilon_D^*} + \frac{1}{\epsilon_{ms}}$ with the short-run overshifting condition $0 \geq \frac{1}{\epsilon_D^*} + \frac{1}{\epsilon_{ms}}$ that applies in the simple case with constant marginal cost $\epsilon_S = \infty$ and full salience $\theta_\tau = 1$. Note that even if $\epsilon_D = 0$ we are able to recover $\frac{dJ}{d\tau} < 0$ and $\frac{\partial \pi}{\partial \tau} < 0$, which is not the case for specific taxation.

Λ is bounded by $\left(\frac{1}{\epsilon_{ms}}\right) \frac{\Lambda \epsilon_D}{p(1+\tau)q} < 1$ so that $\frac{\partial p}{\partial J} < 0$ holds by Lemma 4. Then

$$\frac{\partial \pi}{\partial \tau} < 0 \iff \rho_\tau > \rho_\tau^{SR}$$

Proof. See Appendix. □

5 Comparing the Welfare, Pass-through and Entry Effects of Specific and Ad Valorem Taxes

In this section we discuss the relationship between ad-valorem and specific taxation. We generalize the basic setup by letting $q_j = q_j(p_1, \dots, p_J, \tau, t)$ be a function of prices and both (τ, t) and allow for tax salience by considering the possibility that $q_j(p_1, \dots, p_J, \tau, t) \neq q_j(p_1(1+\tau) + t, \dots, p_J(1+\tau) + t, 0)$. As before, 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: $q^J(p, \tau, t) \equiv q_j(p, \dots, p, \tau, t)$. We define market demand as $Q(p, \tau, t, J) = Jq^J(p, \tau, t)$ and the willingness-to-pay for unit Q when variety is J as the inverse market demand $wtp(Q, J) \equiv P(Q, J) \equiv Q^{-1}(p, 0, 0, J)$. 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 $\theta_t \equiv \frac{\partial Q / \partial t}{\partial Q / \partial p}$, the ratio of the demand responses to the tax and price starting from $\tau = t = 0$.

Assuming that θ_τ and θ_t are constant over (p, τ, t, J) , we can show that $Q(p, \tau, t, J)$ satisfies $wtp(Q(p, \tau, t, J), J) = p(1+\theta_\tau\tau) + \theta_t t$. As above, we have the following definitions: $mwtp(Q) = \frac{\partial wtp}{\partial Q}(Q, J)$, $\epsilon_{ms} \equiv \frac{ms(Q)}{ms'(Q)Q}$, where $ms(Q) \equiv -mwtp(Q)Q$ is marginal consumer surplus, $\epsilon_D \equiv -\frac{p(1+\tau)+t}{Qmwtp(Q)}$, $\epsilon_D^* \equiv \frac{p(1+\theta_\tau\tau)}{p(1+\tau)+t}\epsilon_D$ and $\epsilon_S \equiv \frac{c'(q)}{c''(q)q}$. The generalized Lerner condition in this setup is the following:

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

To be able to compare the welfare effects of taxation, we first 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:*

$$\rho_\tau > \rho_t \Leftrightarrow \frac{\Lambda_0 \epsilon_D}{p_0 q_0} > 1$$

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. *Starting from $\tau_0 = t_0 = 0$, the desirability condition ranking $MCPF_\tau$ and $MCPF_t$ is given by:²⁸*

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

Proof. See Appendix. □

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 $\Lambda_0 = 0$, 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 2001a;b). Second, we see that if the variety effect is sufficiently strong, both of these results can be overturned. In particular, if consumers have a strong preference for variety, then consumer prices might be higher under ad valorem taxation and welfare might be lower. Additionally, we see that if consumer prices

²⁸It 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 (4) 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).

are higher under ad valorem taxation, then welfare is higher under specific taxation, but the converse need not be true.

Anderson, De Palma and Kreider (2001*b*) contains what appears to be the first counterexample that ad valorem is not always to be preferred to specific taxation in a setting with love-of-variety and endogenous entry. They consider a model of Bertrand competition without an outside good so that demand is completely inelastic with respect to taxation (see footnote (28)). This eliminates the static distortionary effect of taxation which tends to favor ad valorem taxation. They show that in a discrete choice model with exponentially distributed match values (Proposition 5), that in the absence of taxation, variety is efficient and that specific taxation is more efficient than ad valorem taxation. This result comes from the fact that with completely inelastic demand ($\epsilon_D = 0$), variety does not change with a specific tax (as can be seen from equation (12) by setting $\epsilon_D = 0$) and so continues to be efficient even with taxation; on the other hand, variety is reduced below the efficient level with an ad valorem tax even with inelastic demand. Importantly, this logic implicitly assumes that the tax increase is large; for a small tax change, our results indicate that if variety is efficient, 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.

Our desirability condition 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),²⁹ it follows that when the variety effect is large, specific taxation has more favorable welfare effects than ad valorem taxation. Our results imply that in markets with stronger love-of-variety and more elastic demand, specific taxation is more likely to be a preferable form of excise taxation than ad valorem taxation. It is noteworthy that this contrasts with the speculation of Anderson, De Palma and Kreider (2001*b*) who argue in their conclusion that specific taxation is likely to be more desirable on inelastic commodities and ad valorem may be preferable on elastic products where losses associated with the static demand distortion may outweigh the distortion to product variety.

On the other hand, it is less well known that consumer prices can be higher under ad

²⁹In 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 $J(\tau) < J(t)$.

valorem when the variety effect is large. In fact, Anderson, De Palma and Kreider (2001a) demonstrate in a model of Bertrand pricing with differentiated products, that in the long run consumer prices are always lower with ad valorem prices, similar to results in models featuring homogeneous products and Cournot quantity competition. So what explains the discrepancy in results? The next corollary sheds light on this.

Corollary 7. *Consider the ranking of pass-through for the two tax structures (ρ_τ and ρ_t). Then $\rho_\tau > \rho_t$ implies $\frac{\partial q}{\partial J} > 0$ and $1 - \frac{1}{\epsilon_{ms}} > 0$ so that $\rho_t > 1$.*

Proof. See Appendix. □

Thus, while Proposition 3 shows that we can obtain a rank reversal in the pass-through of specific versus ad valorem taxation if the variety effect is sufficiently strong, Corollary 7 establishes the scenario where this is satisfied. In particular, in order for ad valorem taxes to have a larger effect on prices than specific taxes, there must be a business-augmenting effect so that $\frac{\partial q}{\partial J} > 0$ which is a violation of the restriction imposed in Anderson, De Palma and Kreider (2001a). On the other hand, when $\frac{\partial q}{\partial J} < 0$ is satisfied, $\rho_\tau < \rho_t$ which is the common result reported in the literature but only a particular case of our model.

6 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 we then describe the empirical strategy. Lastly, we discuss the reduced-form estimates that we will then use as inputs into our model-based estimation and calibrations in Section 7.

6.1 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 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-modules* (e.g. fresh eggs, milk, aluminum foil,

batteries, frozen desserts). For computational reasons, we restrict our analytical sample to the top 20 percent of modules in terms of total U.S. sales. To address the concern that sales taxes are spatially correlated across regions of the U.S. in ways that may endogenously reflect the geographic distribution of consumer preferences, we also 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). Grocery stores typically sell both food and nonfood products, and effective tax rates on food and nonfood products differ in most states. This provides us with cross-sectional variation in tax rates across produce-modules *within* stores, and we primarily rely on this source of variation in taxes in our reduced-form analysis to estimate the causal effects of sales taxes.³⁰

We define all of our variables at the level of module (m), store (r), and year (n), which requires aggregating the data over time and across products. In the RMS data, we first aggregate weekly revenue and quantities sold to the yearly level separately for each UPC. The average yearly price for product j in store r is calculated by dividing the total yearly revenue (from the sales 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 UPCs with positive sales within a module-by-store-by-year cell.

Finally, we 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. 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

³⁰We use cross-sectional variation to approximate 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.

counties in our sample are part of 497 different county border pairs. 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.

The sample average tax rate in our sample is 3.4 percent (calculated as a weighted average weighting each observation by the inverse of the number of pairs that a store is part of), which 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. Because prices are calculated from fixed effects regressions, the sample mean of this variable is normalized to 0 in the underlying UPC-level data.³¹ In our final border-county sample, the mean (log) price is 0.005 and the standard deviation is 0.114. For quantity and product variety, the mean and standard deviation (in parentheses) are equal to 10.241 (1.372) and 4.241 (0.969), respectively. Table 1 also reports summary statistics split by above-median and below-median tax rates (based on the median tax rate within each module-by-year cell). Average pre-tax prices are roughly the same for above-median and below-median tax rate cells, while quantity and product variety is slightly lower in above-median tax rate cells.

6.2 Estimation Strategy

We estimate the effects of sales taxes using the following “county border pair” regression model:

$$\log y_{mrnd} = \sum_{z=2006}^{2014} \beta^{y,z} [\log(1+\tau_{mcn}) \times 1\{z = n\}] + \delta_{mdn} + \delta_{rn} + \varepsilon_{mrnd} \quad (22)$$

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 (22), 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., $\beta^y = \frac{1}{9} \sum \beta^{y,z}$ for $y \in \{p, Q, J\}$). This econometric approach is numerically equivalent to

³¹The sample used in these regressions also includes grocery stores in non-border counties. Statistics reported in Table 1 are calculated for our border-county subsample only.

estimating the model separately for each year and then averaging the coefficients. 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 (on prices, quantity and product variety) are therefore identified under the assumption that tax rates are uncorrelated with unobservable determinants of demand conditional on these fixed effects. The identifying assumption is that governments do not set local tax rates on products based on their relative market shares. For example, our estimate of the effect of taxes on quantity demanded would be biased upwards for β^Q 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 subjecting these goods to higher sales taxes. 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. Our approach therefore mimics a cross-sectional difference-in-differences strategy, where the first difference is across products and the second difference is across stores within county border pairs.

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). In supplementary analyses, 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.

6.3 Reduced-form Estimates

The main results from estimating equation (22) 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 (22) in Panel A of Table 1. We find a small amount of overshifting of taxes on consumer 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 $\hat{\beta}^Q = -0.676$ (s.e. 0.155). This estimate is very similar to previous work studying tax salience using the same scanner data; for example, Kroft et al. (2020a) estimate an output elasticity of -0.650 (s.e. 0.084) using a different source of tax variation.³²

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. Another way to interpret the magnitude is using the formula for the variety elasticity in terms of structural parameters, and then inspecting the parameters that are consistent with this reduced-form estimate; we do this in Section 7 below and find that it leads to reasonable estimates of the underlying structural parameters governing the love-of-variety, the shape of demand curve, and the market conduct parameter. 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 our setting.

We also assess the reliability and robustness of these main results in two ways. 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 Figures OA.1 and OA.2. These figures show that the county border pair estimates are fairly stable across years and clustered around the across-year simple average. We gain precision by pooling the OLS and 2SLS estimates across years, and these figures show that our model-based estimates are not

³²The estimates in this paper are based on a pooled cross-sectional research design comparing across counties that share a state border, while Kroft et al. (2020a) used within-county variation in tax rates over time. One reason we focus on a cross-sectional research design is that we see the long-run free-entry condition in the theory as potentially taking a long time to emerge. Future work could try to combine the sources of variation to trace out the dynamic adjustments of product variety to sales taxes over time.

sensitive to the specific choice of years in the sample.

Second, 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 how prices, quantity, and product variety vary with the tax rate they are subject to (on the basis of whether they are food or nonfood products) versus the tax rate that applies to other products in the same jurisdiction. To implement these tests, we residualize the outcome and tax variables on module-by-border pair-by-year fixed effects and calculate the difference in mean residuals between the two counties in each county-pair d in each year n , separately for food and nonfood products. We obtain a dataset with two observations per county-pair-year cell, one for food products and one for nonfood products. All variables (outcomes and tax rates) are measured as within-cell (between-county) differences in mean residuals. These results are reported in Appendix Table OA.1, and they represent a kind of “horse race” model that exploits residual variation across food and nonfood products. Formally, we regress $\log(p)$, $\log(Q)$ and $\log(J)$ on their own tax rate (i.e., average food tax rates for food products and average nonfood tax rates for nonfood products) as well as on the other category’s tax rate. In column (1) only the “own” tax rate is included, in column (2) only the “other” tax rate is included, and in column (3) both average tax rates are included simultaneously. For quantity and variety, the placebo (other) tax rate has no residual explanatory power: the between-county differences in quantity and variety are unrelated to differences in tax rates that apply to *other* products. By contrast, prices appear to increase both as a function of own and other tax rates, but the coefficient on own tax rate is larger in magnitude. This suggests stores raise *all* prices when taxes increase on *some* products, and raise prices even more for products for which the tax increase directly applies.³³ Finally, in column (4), we include cell fixed effects to mimic our main results that include store fixed effects. Here, the coefficient on own taxes is identified by differencing between food and nonfood products within counties. Overall, the reported coefficients are similar to the preferred estimates of the effects of taxes reported in Table 2.

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 product variety. These reduced-form effects are robust to alternative specifications and stable over time, and they are useful for calibrating the marginal excess burden of ad valorem

³³Note that our main results that do not take into account such cross-product spillovers may therefore understate the true extent of overshifting.

taxation, as well as identifying the main model parameters.

7 Model calibrations

This section considers identification and estimation of the deeper model parameters and calibrates the main welfare formula. We also show how estimating the deeper parameters allows us to learn about whether or not variety is socially optimal at current tax rates and compare the incidence and efficiency costs of the existing ad valorem taxes to a counterfactual specific tax.

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

This subsection describes how we identify and estimate the deeper model parameters using the reduced-form estimates. We rely on the formulas in Section 4 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 (see Table 1). We calibrate the price elasticity of demand to be $\epsilon_D = 1.223$ and the tax salience parameter to be $\theta_\tau = 0.528$, based on the estimates in Kroft et al. (2020a). Both of these parameters are estimated using similar retail scanner data. This leaves three remaining unknown parameters: the love-of-variety parameter ($\tilde{\Lambda}_0 \equiv \Lambda_0 \frac{J}{pQ}$), the elasticity of inverse marginal surplus (ε_{ms}), and the conduct parameter (v_q/J). We can then solve for the values of the three unknown parameters that make the model-implied effects of taxes exactly match the reduced-form estimates in Table 2.

Before solving for the parameter estimates, we can build on the theoretical analysis in Section 4 to provide an intuitive discussion of how each parameter is identified. The conduct parameter is identified by the long-run free-entry condition (Lemma 3), 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 3 reports the markup estimate of $(p - c'(q))/p = 0.080$. Given this estimate of the markup, we can then use the generalized Lerner index defined in equation (13) 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 $v_q/J = 0.096$ (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. (2020a) for more detailed discussion).³⁴

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 Proposition 2). Corollary 4 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 4 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:

$$\begin{aligned}\tilde{\Lambda}_0 &= k_1 + k_2 * (1/\epsilon_{ms}) \\ k_1 &\equiv (1 + \theta_\tau \tau) \left(\frac{\hat{\beta}^Q}{\hat{\beta}^J} \left(1 + \frac{1 - \frac{\nu_q}{J}}{\epsilon_D^*} \right) + \frac{\frac{\nu_q}{J}}{\epsilon_D^*} - 2 \right) \\ k_2 &\equiv (1 + \theta_\tau \tau) \frac{\hat{\beta}^Q}{\hat{\beta}^J}\end{aligned}\tag{23}$$

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

³⁴Kroft et al. (2020a) 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 is 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 the shorter run (quarterly) analysis in Kroft et al. (2020a).

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 3 reports $\epsilon_{ms} = -0.936$.³⁵ 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 (23) to identify and estimate $\tilde{\Lambda}$. 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 for variety. Using equation (23), we estimate the variety effect to be $\tilde{\Lambda}_0 = 0.133$. This parameter can be given a willingness-to-pay interpretation: an exogenous 10 percent reduction in variety reduces average willingness-to-pay by 1.33 percent. This magnitude is smaller than the reciprocal of the demand elasticity (0.818), 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 the true love-of-variety).³⁶

7.2 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 6. Since the estimates

³⁵Formally, 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., $\tilde{\Delta} > 0$). As discussed above in Section 4, the stability condition will be violated whenever the estimate of $\tilde{\Lambda}$ is very large, which will generally be the case whenever $(\hat{\beta}^Q/\hat{\beta}^J)$ is very large.

³⁶Since we calibrate the tax salience parameter and the price elasticity of demand, Appendix Table OA.2 shows the sensitivity of our results to alternative values of these parameters. Interestingly, varying both parameters but holding constant the product of the two terms leads to very similar values of the variety effect. Increasing the demand elasticity while holding constant the tax salience parameter leads to a larger change in the variety effect than the curvature parameter, while decreasing the demand elasticity actually leads to a negative love-of-variety parameter estimate. If such “hate-of-variety” preferences are ruled out a priori, then this implies that our model puts a lower bound on the demand elasticity given our reduced-form empirical estimates. The calibrated value of the price elasticity of demand comes from relatively “short run” variation in local prices, and our analysis in the present paper focuses on “long run” effects of taxes allowing for entry and exit of products, it is probably more likely that the calibrated demand elasticity is an under-estimate rather than over-estimate since demand curves are generally more elastic over time (Samuleson 1947).

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 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 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 $d\widetilde{W}/d\tau = -0.085$ (see Panel D of Table 3). This is larger in magnitude than a standard Harberger benchmark adjusted for salience effects ($d\widetilde{W}/d\tau = -0.013$), 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 $d\widetilde{W}/d\tau = -0.052$. This is the estimate one would calibrate based on our extension of the theoretical results in Besley (1989) to cover ad valorem taxes. The reason we find a larger negative effect of taxes on welfare (relative to a Besley (1989)-style benchmark) is that we find a positive love-of-variety estimate and a negative effect of taxes on product variety – and according to the full marginal excess burden (MEB) formula, the MEB is increasing in the product of these two terms.

7.3 Welfare analysis of product variety

Using the full structure of the model, we can infer whether or not variety is socially optimal. To see how this is achieved, we first re-write the marginal excess burden formula that connects to equation (5) as follows:

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

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 costs of additional variety from the love-of-variety benefits. In order to calibrate the expression, we need an estimate of the love-of-variety parameter and the mark-up (see subsection 7.1 above), along with an estimate of $\frac{\partial \log(q)}{\partial \log(J)}$. Lemma 4 above shows that

this partial elasticity can be defined in terms of the conduct parameter, demand elasticity, and inverse elasticity of marginal surplus. Using all of the estimated parameters described in previous subsection and listed in Table 3, we estimate $\frac{\partial \log(q)}{\partial \log(J)} = -0.728$ (see Panel B of Table 4). Multiplying this partial elasticity by the average mark-up and adding the love-of-variety parameter leads to an estimate of $\partial \widetilde{W} / \partial \log(J) = 0.075$, 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 4, 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 $\tilde{\Lambda}_0 = 0.065$, 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 (see, e.g., 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 4 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).

7.4 Comparing ad valorem taxes to specific taxes

In this subsection, we 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 order to compare the efficiency costs of ad valorem and specific taxes, we use the marginal cost of public funds ($MCFP_\tau$) concept defined above; i.e., the ratio of the effect of taxes on welfare to the effect of taxes on revenue ($-\frac{dW/dt}{dR/dt}$). Table 5 reports the counterfactual

results. Column (1) reports the estimated pass-through (ρ_τ) and $MCFP_\tau$ at current average (ad valorem) tax rate of 0.034. The pass-through rate that is reported in column (1) closely matches the reduced-form empirical estimate in Table 2 and indicates a small amount of overshifting. The $MCFP_\tau$ of 0.083 is very similar in magnitude to the marginal excess burden reported in the bottom of Table 3 (of -0.085).³⁷

We next estimate the pass-through (ρ_t) and marginal cost of public funds ($MCFP_t$) of a counterfactual specific tax, which we estimate by using the theoretical formulas for ρ_t and $MCFP_\tau$ of a specific tax, holding constant all of the model parameters estimated above. The results in column (2) show that this counterfactual specific tax would lead to even more overshifting (i.e., greater magnitude of pass-through), but a smaller efficiency cost (0.067, compared to 0.083 in column (1)).

For comparison, in columns (3) and (4) we redo the same calculations at the same values of all of the model parameters except for the love-of-variety parameter, which we set to zero in these two columns. In this case, we continue to find more overshifting in the case of a specific tax, but a larger efficiency cost. In other words, we are able to replicate the longstanding result that ad valorem taxes dominate specific taxes in terms of efficiency cost when there is no love-of-variety. By contrast, we find that the actual love-of-variety we estimate (which we view as fairly modest in size and well below the love-of-variety parameter one would estimate using a standard logit choice model) “flips” this result and leads us to conclude that ad valorem taxes actually lead to greater – not lower – efficiency costs.

In the final two columns (columns (5) and (6)), we increase the love-of-variety parameter from the baseline value in the first two columns (and continue to hold all of the other model parameters constant). As in the first two columns, we find larger efficiency costs of ad valorem taxes, but we now find greater overshifting for specific taxes. These results reinforce the conclusions from the theoretical analysis: while the conventional view is that ad valorem taxes “dominate” specific taxes on efficiency grounds and have lower pass-through, we find that both of these can be overturned when love-of-variety is strong enough.

The remaining panels of Table 5 give additional statistics that provide further economic intuition behind these results. Panel C shows that the short-run (fixed-variety) pass-through rate is smaller in magnitude for all values of the variety effect for ad valorem taxes, but is slightly larger in magnitude for specific taxes. Panel D shows that the difference between the short-run and long-run pass-through can be understood as a consequence of the (long-run)

³⁷This is not surprising since the pre-existing tax level is small and the amount of overshifting is also small, which together imply that $d \log(R)/d \log(1 + \tau)$ is very close to 1.

effect of taxes on variety and the (short-run) effects of taxes on firm profits. Both of these effects are negative for ad valorem taxes and positive for specific taxes. Lastly, Panel E shows that there are larger business-stealing effects when there is no love-of-variety (comparing $\frac{\partial \log(q)}{\partial \log(J)}$ in columns (3) and (4) to columns (1) and (2)), and in the final two columns with a large love-of-variety the business-stealing effect switches sign and becomes positive (i.e., increased entry actually raises demand for existing products). This is also the scenario where pass-through of ad valorem taxes is larger than pass-through of specific taxes in the long run, in contrast to previous theoretical analyses that restrict the sign of the business-stealing effect to rule out this possibility (e.g., Anderson, De Palma and Kreider 2001a). The final row shows that all of these scenarios continue to satisfy the stability condition that $\Delta > 0$ (see Proposition 2), which implies that the restrictions in previous work are not necessary for stability, and our more general formulas show that these restrictions would lead to incorrect conclusions regarding the ranking of pass-through when love-of-variety is large.

In the Appendix we report additional calibration results covering alternative counterfactual scenarios. Appendix Table OA.3 shows sensitivity to alternative values of the inverse elasticity of marginal surplus and reports results for even greater love-of-variety. By decreasing ϵ_{ms} , we are able to find combinations where increases in counterfactual specific taxes also lead to reductions in variety, similar to the baseline scenario studying ad valorem taxes. In this case, we find lower pass-through in the short run, as well as short-run reductions in profits from larger taxes. The final two columns show that the stability condition can be eventually violated when ϵ_{ms} is decreased and the love-of-variety is very large. Lastly, Appendix Table OA.4 shows how pass-through varies with love-of-variety when ϵ_{ms} is increased (rather than decreased, as in OA.3). We find that at baseline parameter estimates greater love-of-variety leads to larger pass-through, but we find this comparative static is reversed when $\epsilon_{ms} = -0.468$ (i.e., half the magnitude of the baseline estimate). In this case, both types of taxes now encourage entry and raise profits in the short run, and in this case greater love-of-variety leads to lower – rather than higher – pass-through in the long run.

Taken together, we conclude from our empirical calibrations that even a relatively modest love-of-variety is enough to overturn the standard efficiency cost ranking, implying that ad valorem taxes may be suboptimal tax instruments when applied to product markets with substantial product differentiation.

8 Conclusion

In this paper, we develop new formulas to study the efficiency cost and incidence of ad valorem taxes relative to 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 and tax salience. 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 output, producer prices and product variety in terms of the full set of model parameters. We then combine retail scanner data from the U.S. with quasi-experimental variation in (ad valorem) sales taxes to estimate these three reduced-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 disconnecting the love-of-variety 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. The larger the reduced-form variety response to taxes (holding constant the other parameters and the other reduced-form estimates), the larger the love-of-variety. While there are many natural extensions to our model (e.g., allowing for multi-product firms, as in Hamilton 2009), we believe the identification approach in this paper is likely to extend naturally to many richer settings.

Lastly, we believe our framework can be used to study particular commodity markets subjected to “sin taxes” (e.g., 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 output to the tax (alongside the markup), then it is separable from the variety effect. As a result, extended versions of our formulas that account for these kind of externalities (i.e., 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, alongside the market demand elasticity and market conduct. Overall, we conclude from both our theoretical results and calibrations that 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

- 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.
- 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.
- 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*.
- Baker, Scott R, Stephanie Johnson, and Lorenz Kueng.** forthcoming. “Shopping for Lower Sales Tax Rates.” *American Economic Journal: Macroeconomics*.

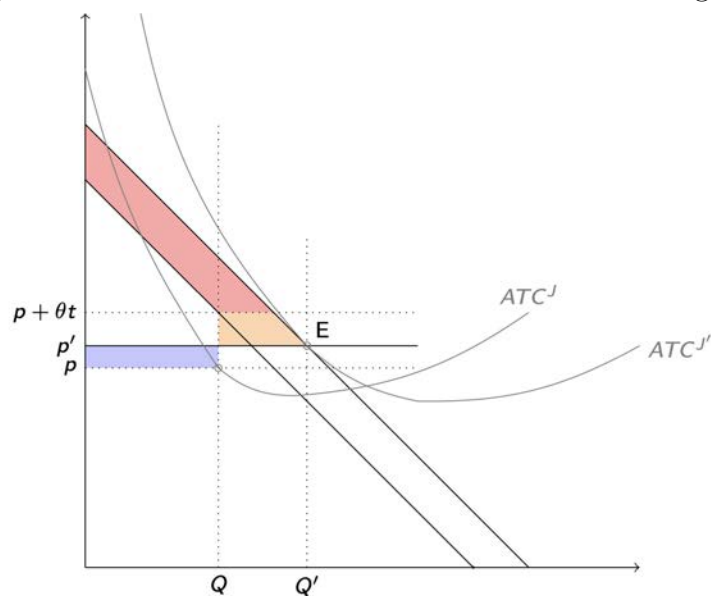
- Barzel, Yoram.** 1976. "An Alternative Approach to the Analysis of Taxation." *Journal of Political Economy*, 84(6): 1177–97.
- Besley, Timothy.** 1989. "Commodity Taxation and Imperfect Competition: A Note on the Effects of Entry." *Journal of Public Economics*, 40: 359–367.
- Besley, Timothy J ;, and Harvey S Rosen.** 1999. "Sales taxes and prices: An empirical analysis." *National Tax Journal*, 52(2): 157.
- Blackorby, Charles, and Sushama Murty.** 2007. "Unit versus ad valorem taxes: Monopoly in general equilibrium." *Journal of Public Economics*, 91(3-4): 817–822.
- Boone, Christopher, Arindrajit Dube, Lucas Goodman, and Ethan Kaplan.** 2016. "Unemployment Insurance Generosity and Aggregate Employment."
- 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.
- Cournot, A.** 1960. *Researches into the Mathematical Principles of the Theory of Wealth*. London:Frank Cass & Co.
- 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.

- de Meza, David.** 1982. “Generalized Oligopoly Derived Demand with an Application to Tax Induced Entry.” *Bulletin of Economic Research*, 34(1): 1–16.
- Denicolò, Vincenzo, and Massimo Matteuzzi.** 2000. “Specific and Ad Valorem Taxation in Asymmetric Cournot Oligopolies.” *International Tax and Public Finance*, 7: 335–242.
- Dhingra, Swati, and John Morrow.** 2019. “Monopolistic competition and optimum product diversity under firm heterogeneity.” *Journal of Political Economy*, 127(1): 196–232.
- Dierickx, I., C. Matutes, and D. Neven.** 1988. “Indirect taxation and cournot equilibrium.” *International Journal of Industrial Organization*.
- 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. “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.
- Handbury, Jessie, and David E Weinstein.** 2015. “Goods prices and availability in cities.” *The Review of Economic Studies*, 82(1): 258–296.
- Harberger, Arnold C.** 1964. “The Measurement of Waste.” *American Economic Review*, 54(3): 58–76.
- Hart, Oliver.** 1982. “Reasonable Conjectures.” *Working Paper*.
- Helpman, Elhanan, and Paul Robin Krugman.** 1989. *Trade Policy and Market Structure*. Cambridge, MA: The MIT Press.
- Hendren, Nathaniel.** 2016. “The policy elasticity.” *Tax Policy and the Economy*, 30(1): 51–89.
- 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.
- Johnson, Terry R.** 1978. “Additional Evidence on the Effects of Alternative Taxes on Cigarette Prices.” *Journal of Political Economy*, 86(2): 325–8.

- 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.
- 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.
- Kowalczyk, Carsten.** 1990. “Monopoly and trade policy.” National Bureau of Economic Research.
- Kroft, Kory, Jean-William Laliberté, René Leal Vizcaíno, and Matthew Notowidigdo.** 2020a. “Salience and Taxation with Imperfect Competition.” *NBER Working Paper*.
- Kroft, Kory, René Leal-Vizcaíno, Matthew J. Notowidigdo, and Ting Wang.** 2020b. “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.
- Myles, Gareth D.** 1989. “Product Variety and Tax Policy.” *Working Paper*.
- Myles, Gareth D.** 1996. “Imperfect Competition and the Optimal Combination of ad valorem and Specific Taxation.” *International Tax and Public Finance*, 3: 29–44.
- Myles, G. D.** 1987. “Optimal Commodity Taxation with Imperfect Competition.” *Warwick Economic Research Papers*.
- Parenti, Mathieu, Philip Ushchev, and Jacques-François Thisse.** 2017. “Toward a theory of monopolistic competition.” *Journal of Economic Theory*, 167: 86–115.
- 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.
- Samuleson, Paul.** 1947. “Foundations of economic analysis.” *Harvard University Press, Cambridge, Mass.*
- Seade, J.** 1985. “Profitable Cost Increases and the Shifting of Taxation: Equilibrium Responses of Markets in Oligopoly.” *Working Paper*.

- Seade, Jesus.** 1980. “On the Effects of Entry.” *Econometrica*, 48(2): 479–489.
- Skeath, Susan E., and Gregory A. Trandel.** 1994. “A Pareto comparison of ad valorem and unit taxes in noncompetitive environments.” *Journal of Public Economics*, 53(1): 53–71.
- Spence, Michael.** 1976. “Product Differentiation and Welfare.” *Source: The American Economic Review*, 66(2): 407–414.
- Stern, Nicholas.** 1987. “The Effects of Taxation, Price Control and Government Contracts in Oligopoly and Monopolistic Competition.” *Journal of Public Economics*, 32: 133–158.
- Suits, D. B., and R. A. Musgrave.** 1953. “Ad valorem and unit taxes compared.” *Quarterly Journal of Economics*, 67(4): 598–604.
- Sullivan, Daniel.** 1985. “Testing Hypotheses about Firm Behavior in the Cigarette Industry.” *Journal of Political Economy*, 93(3): 586–598.
- 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.
- 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.
- Wicksell, K.** 1896, 1959. “Taxation in the Monopoly Case (Translation from the original).” In *Readings in the Economics of Taxation.*, ed. R. A. Musgrave and C. S. Shoup. London: Allen & Unwin.
- Zoutman, Floris T., Evelina Gavrilova, and Arnt O. Hopland.** 2018. “Estimating Both Supply and Demand Elasticities Using Variation in a Single Tax Rate.” *Econometrica*, 86(2): 763–771.

Figure 1: Welfare Effect of a Tax Decrease in the Long Run



Notes: In this figure (p, Q) is the starting equilibrium producer price and quantity for a specific tax $t > 0$. When the tax disappears, entry of new firms $J' > J$ drives the aggregate demand out and a new equilibrium is reached at (p', Q') . The welfare change is represented in the picture by the sum of the red and orange areas minus the blue area which correspond to equation (7), up to a first-order approximation.

Table 1: Summary statistics

	Full sample		Above-median tax rate		Below-median tax rate	
	Average	Standard deviation	Average	Standard deviation	Average	Standard deviation
Tax rate, τ_{mcn}	0.034	0.032	0.055	0.027	0.018	0.026
<i>Key outcomes in reduced-form analysis:</i>						
log Prices (pre-tax)	0.005	0.114	0.002	0.111	0.006	0.115
log Quantity	10.241	1.372	10.024	1.370	10.414	1.349
log Product Variety	4.241	0.969	4.229	0.957	4.251	0.978
<i>Sample size statistics:</i>						
N (module-store-year observations)	11,624,918		4,764,127		6,860,791	
N (stores)	3,822		3,554		3,662	
N (modules)	198		198		198	
N (years)	9		9		9	
N (counties)	543		509		513	
N (border pairs)	497		490		495	

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 effective on September 1. Prices, quantity, and variety are measured yearly. Median tax rates are calculated separately for each module-year cell. For many modules that are generally tax exempt, the median rate is zero, which results in fewer observations above the median (strictly above zero) than below (equal to zero). 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. See main text and Data Appendix for more details.

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

Dependent Variable:	Prices (1)	Quantity (2)	Variety (3)
Panel A: County Border Pair OLS Estimates			
$\log(1 + \tau_{mcn})$	0.038 (0.016)	-0.676 (0.155)	-0.236 (0.074)
Panel B: 2SLS Estimates Using State-Level Tax Rate as Instrument			
$\log(1 + \tau_{mcn})$	0.039 (0.016)	-0.731 (0.154)	-0.243 (0.075)
<i>Specification:</i>			
Store fixed effects	y	y	y
Module $\times$ County Border Pair fixed effects	y	y	y

Notes: Sales tax rates 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 coefficients are linear combinations of nine 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: Estimating Variety Effect and Calibrating Welfare Formulas

<u>Panel A: Calibrated parameters</u>	
Average tax rate, τ_0	0.034
Tax salience parameter, θ_τ	0.528
Demand elasticity, ϵ_D	1.223
<u>Panel B: Reduced-form estimates</u>	
Pass-through of taxes into pre-tax prices, $d\log(p)/d\log(1+\tau)$	0.039
Quantity response, $d\log(Q)/d\log(1+\tau)$	-0.731
Variety response, $d\log(J)/d\log(1+\tau)$	-0.243
<u>Panel C: Model parameters estimated by matching reduced-form estimates</u>	
Markup, $(p - c'(q))/p$	0.080
Implied conduct parameter, v_q/J	0.096
Inverse elasticity of marginal surplus, ϵ_{ms}	-0.936
Variety effect parameter, $\tilde{\Lambda}_0$	0.133
<u>Panel D: Calibrated welfare formulas</u>	
Full marginal excess burden (MEB) formula, $d\tilde{W}/d\tau$	-0.085
Alternative MEB formula benchmarks:	
Harberger / Chetty-Looney-Kroft benchmark, $\theta_\tau * \tau_0 * d\log(Q)/d\log(1+\tau)$	-0.013
Besley(1989)-style benchmark; i.e., full MEB formula with $\tilde{\Lambda}_0 = 0$	-0.052

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 variety effect parameter, inverse elasticity of marginal surplus, and 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 main formula with benchmarks from Harberger/Chetty-Looney-Kroft and Besley (1989). See text for details.

Table 4: Socially Optimal Variety Calibrations

	Baseline	Counterfactual scenarios	
	(1)	(2)	(3)
<u>Panel A: Variety effect parameter, $\tilde{\Lambda}_0$</u>			
Estimated variety effect parameter	0.133	0.065	0.000
<u>Panel B: Socially optimal variety calculations</u>			
$\partial \log(q)/\partial \log(J)$	-0.728	-0.819	-0.907
Business-stealing effect, $\partial \log(q)/\partial \log(J) * (p - c'(q))/p$	-0.058	-0.065	-0.072
$\partial \tilde{W}/\partial \log(J) = \text{Variety effect } (\tilde{\Lambda}_0) + \text{business-stealing effect}$	0.075	0.000	-0.072
(>0 implies variety is below the social optimum)			

Notes: This table reports results using the parameter estimates from Table 3 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 3, 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 rate), and column (3) sets variety effect to 0.

Table 5: Counterfactual Scenarios Comparing Ad Valorem and Unit Tax Taxes

Variety effect parameter, $\tilde{\Lambda}_0$	Baseline variety effect estimate, $\tilde{\Lambda}_0 = 0.157$		No variety effect counterfactual, $\tilde{\Lambda}_0 = 0.000$		Large variety effect counterfactual, $\tilde{\Lambda}_0 = 1.000$	
	Ad		Ad		Ad	
	valorem tax ($d\tau$)	Specific tax (dt)	valorem tax ($d\tau$)	Specific tax (dt)	valorem tax ($d\tau$)	Specific tax (dt)
	(1)	(2)	(3)	(4)	(5)	(6)
<u>Panel A: Pass-through of taxes into pre-tax prices</u>						
$d\log(p)/d\log(1+\tau)$ or $d\log(p)/dt$	0.039	0.058	0.035	0.058	0.061	0.056
Difference b/w ad valorem and specific tax	-0.019		-0.023		0.005	
<u>Panel B: Marginal cost of public funds (MCPF)</u>						
$MCFP_\tau$ or $MCPF_t$	0.083	0.067	0.047	0.070	0.311	0.045
Difference between ad valorem and specific tax	0.017		-0.023		0.267	
<u>Panel C: Short-run pass-through</u>						
$d\log(p)/d\log(1+\tau) J$ or $d\log(p)/dt J$	0.013	0.061	0.013	0.061	0.013	0.061
Difference between SR and LR pass-through	0.026	-0.003	0.022	-0.002	0.047	-0.005
<u>Panel D: The effects of taxes on variety and profits</u>						
$d\log(J)/d\log(1+\tau)$ or $d\log(J)/dt$	-0.243	0.024	-0.244	0.024	-0.234	0.023
$\partial\log(\pi)/\partial\log(1+\tau)$ or $\partial\log(\pi)/\partial t$	-0.041	0.004	-0.041	0.004	-0.041	0.004
<u>Panel E: Competitive effects of entry</u>						
$\partial\log(p)/\partial\log(J)$	-0.108	-0.106	-0.092	-0.091	-0.209	-0.205
$\partial\log(q)/\partial\log(J)$	-0.728	-0.717	-0.907	-0.893	0.432	0.426
Stability condition (must be >0)	1.812	1.812	1.801	1.801	1.882	1.882

Notes: This table reports counterfactual estimates of reduced-form effects of specific taxes under different assumptions on variety effect based on using the model parameter estimates of Table 3. 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 variety effect is large (columns (5) and (6)).