Quantifying the Welfare Gains of Variety: A Sufficient Statistics Approach

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

First version: April 2017. This version: July 2017

Abstract

This paper develops a new revealed-preference approach for valuing changes in product variety. We show that the "variety effect" – the change in consumer surplus resulting from a change in the number of available products, holding prices constant – can be represented graphically as the area between the inverse market demand curves before and after the change in product variety. Our key contribution is to derive a sufficient statistics formula for the variety effect under the assumption of parallel inverse market demand curves. This formula depends on the price elasticity of demand when variety is fixed and the price elasticity of demand when variety is permitted to vary. We demonstrate that a wide class of continuous and discrete choice models give rise to parallel inverse demand curves, showing that our formula is robust. We illustrate the value of our approach by considering an empirical application to taxes. In particular, we show how one can implement our sufficient statistics formula using reduced-form estimates of the effect of taxes on variety and the effect of taxes on prices and quantities in two cases: where variety is held constant and where variety responds to a change in taxes through firm entry or exit. Combining retail scanner data from grocery stores in the U.S. with detailed local sales tax data and using within-store and between-store variation in rates and exemptions, we estimate a large effect of sales taxes on product variety. Finally, we discuss several additional applications in Industrial Organization and Public Economics.

JEL codes: H21, H71, F12, L13.

Keywords: Market demand, consumer surplus, product variety, new goods.

*Kroft: University of Toronto and NBER, kory.kroft@utoronto.ca; Laliberté: University of Toronto, jw.plaliberte@mail.utoronto.ca; Leal-Vizcaíno: Northwestern University, renelealv@u.northwestern.edu; Notowidigdo: Northwestern University and NBER; noto@northwestern.edu. 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, Pinchuan Ong, Adam Miettinen, Boriana Miloucheva, Shahar Rot-berg, Marc-Antoine Schmidt, Jessica Wagner 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. Calculated (or Derived) based on data from The Nielsen Company (US), LLC and marketing databases provided by the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the Nielsen data are those of the researchers and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

1 Introduction

Quantifying the benefits to consumers from new products is important for a broad range of economic issues. In Industrial Organization (IO), it is central to whether markets provide an efficient level of product variety (Spence 1976ab; Dixit and Stiglitz 1977; Mankiw and Whinston 1986). In International Trade, it is crucial to a full accounting of the gains from trade (Feenstra 1994; Broda and Weinstein 2006).

The typical approach to measuring the gains from greater product variety in both the IO and Trade literatures has been to specify and estimate structural models of demand. In IO, the standard approach involves estimating Random Utility models that allow for rich observed and unobserved heterogeneity across both product characteristics and consumer tastes and examining the impact of greater variety on consumer surplus (see, e.g., Berry, Levinsohn, and Pakes 1995; Petrin 2002; Nevo 2003; Sheu 2014; Berry, Eizenberg and Waldfogel 2016). This approach allows for a welfare analysis of counterfactual scenarios, but requires modeling assumptions that may be hard to assess and may have meaningful effects on the ultimate welfare estimates. Moreover, fully-specified structural models are naturally specific to the particular setting, which makes it hard to extend and compare results from these exercises to other settings.

By contrast, the standard approach in Trade to measuring the welfare gains from product variety involves estimating parametric models of demand that ignore much of the heterogeneity that is the focus in the IO literature (see, e.g., Feenstra 1994; Broda and Weinstein 2006; Arkolakis, Costinot and Rodriguez-Clare 2012; Melitz and Redding 2015; Atkin, Faber and Gonzales 2016). This allows researchers to analyze the consequences of changes in product variety in a large number of markets simultaneously, but requires researchers to make strong functional assumptions on preferences such as Constant Elasticity of Substitution (CES) demand. One concern with CES preferences is that its restrictive since the elasticity of substitution pins down both the price elasticity of demand and the variety effect (see Benassy 1996).

In this paper, we propose a new revealed preference approach to studying the welfare consequences of changes in product variety, what we label the "variety effect".[1] Rather than specify a consumer's utility function as a function of the number of offered varieties in the

[1] To be clear, there is also a "price effect" associated with changes in variety which affects consumer surplus. The variety effect is defined as the change in consumer surplus due to the change in variety *holding market prices constant*.

market, the spirit of our approach is to use the market demand curve (for a given number of products) directly to evaluate changes in consumer surplus due to changes in variety. This approach follows the recent "sufficient statistics" literature (e.g., Heckman and Vytlacil 2007; Chetty 2009) which is based on "Marschak's maxim" that "one should solve well-posed economic problems with minimal assumptions". Marschak (1953) noted that for many questions of economic and policy analysis, it is not necessary to identify the full set of policy-invariant structural parameters. All that may be required are combinations of subsets of the structural parameters, corresponding to the parameters required to forecast particular policy reforms, which are often much easier to identify since they require fewer and weaker assumptions. Additionally, as Nevo (2011) has noted, aggregation has the advantage in that it allows for flexible, non-parametric functional forms. Although there may be many applications where we care about substitution between particular products (e.g., merger analysis), for some applications, it may be sufficient to work with overall demand.

For ease of exposition and to help with intuition, we begin with a simple model of symmetric products and derive an expression for the change in consumer surplus due to a small increase in variety. This expression is the sum of the "price effect" and the "variety effect". While the price effect is typically estimated directly using reduced-form methods, estimating the variety effect has required placing more structure on consumer demand (Hausman and Leonard 2002). We show that the variety effect corresponds to the area between the inverse market demand curves before and after a change in variety.[2] Identifying the variety effect therefore requires non-parametric identification of the demand curves before and after a change in variety and integration between these curves.

In practice, researchers often do not have enough information to trace out demand curves non-parametrically; for example, it is necessary to observe the price for which aggregate demand is close to 0. However, the "choke price" is not observable in most data sets. Thus, a key objective of this paper is to pursue a more easily implementable method to identify the variety effect. We show that under the assumption that the market inverse demand curve shifts *in parallel* in response to (exogenous) changes in variety, the variety effect depends on three key "sufficient statistics": the price elasticity of market demand when variety is held constant, the price elasticity of market demand when variety can vary, and the change in market output in response to the change in variety. Together, these statistics jointly reveal

[2]This is similar to the approach to value a new good using the area under the demand curve for the new good (Trajtenberg 1989; Hausman 1996; Bhattacharya 2015).

information about the vertical distance between the demand curves.

The intuition for our sufficient statistics formula is the following. Under parallel demands, the area between the demand curves corresponds to the area of a rectangle. The base of the rectangle is market output which is observable. The height of the rectangle is the difference between the final price that prevails after variety adjusts and the price that would have prevailed at the new level of output but on the original market demand curve. While the former is observable, the latter is a counterfactual that must be estimated. The key insight is that the difference between the fixed-variety price elasticity of demand and the variable-variety price elasticity of demand exactly pins down this difference in prices. We use these insights to provide a graphical representation of the variety effect using market demand curves. As in Einav, Finkelstein, and Cullen (2014), we view these graphs as providing useful intuition and therefore as an important contribution on their own.

We next demonstrate that our sufficient statistics formula is valid for a wide class of continuous and discrete choice models with symmetric preferences.[3] We first consider the Random Utility model and show that the symmetric Nested Logit model gives rise to parallel inverse demand curves. Next, we show that for a wide class of distributions of the random utility shock, the inverse aggregate demands are asymptotically parallel as the variety level increases. This result comes from Extreme Value Theory: when the random utility shocks are independent and identically distributed, the distribution of the maximum order statistic converges to a Gumbel distribution (the same as the Logit shocks) for a wide range of distributions. We show numerically that this convergence happens very quickly for many standard shock distributions (Normal, Gamma, and Exponential).[4] Lastly, we come back to continuous choice models and we characterize the utility functions that give rise to parallel demands. Here we consider commonly used functional forms, such as CES preferences and preferences that give rise to constant pass-through (Bulow and Pfleiderer 1983; Fabinger and Weyl 2016).[5] We also consider a behavioral application where consumers face attention costs when evaluating their most preferred option over a choice set, consistent with the idea that

[3]In the symmetric model, all of our welfare formulas are derived under the assumption that the number of varieties is a continuous variable. To handle the discrete case, we consider an extension to probabilistic entry and obtain similar results.

[4]Extreme Value Theory has been used in economics in a random utility context by Gabaix et. al. (2016) to show that there might exist robustly high equilibrium markups in large markets that are insensitive to the degree of competition as the number of firms increases.

[5]We also make the connection with Benassy (1996), who recognizes that in the usual CES parametrization, the taste for variety and the elasticity of substitution are intertwined and provides a CES parametrization where the variety effect is identified separately from the elasticity of substitution.

consumers may suffer from "choice overload" when choosing among a large set of options.

Moving beyond the symmetric model, we next consider our general model with asymmetric products and prices and we show that our sufficient statistics formula holds under a "symmetric pass-through" assumption. Specifically, we begin by showing that we can still obtain the variety effect as the area between the market demand curves before and after the change in variety. Next, under the assumption that all prices are increased simultaneously and in the same proportion in response to a change in the number of varieties, we show that we can express the area between the market demand curves in terms of our sufficient statistics. Importantly, we show that we can state the sufficient statistics in terms of the elasticity of market demand with respect to the *uniform price response* and therefore we do not need to estimate how total output changes with each individual price separately.

As a final step, we evaluate the robustness of our approach by considering the Random Coefficients Logit model, which has become the workhorse model for demand estimation in the IO literature following the seminal contribution of Berry, Levinsohn, and Pakes (1995) ("BLP"). We simulate a BLP model of demand with heterogeneous prices and product characteristics and compute the welfare gain from adding new varieties to the market using both exact consumer surplus and using our general sufficient statistics formula. Our results indicate that the welfare estimates are quite similar. This is true when we allow for unobserved product heterogeneity and endogenous selection of variety, where the new varieties have product characteristics that are very different from the existing varieties. The intuition is that the demand curves before and after the change in variety capture how consumers value the new varieties; if the new products are valued a lot, then the demand curve will shift further out to reflect this.

Our sufficient statistics formula can be used to evaluate changes in consumer surplus resulting from *any* shock that exogenously shifts variety in the market. We illustrate this for the case of sales taxes. The key insight is that sales taxes may shift the equilibrium level of variety in the market through firm entry and exit and thus affect consumers by altering their choice set. To test this idea empirically, we combine rich retail scanner data from grocery stores in the U.S. with detailed state and county sales tax data. We operationalize our framework by assuming that the *long-run* effects of taxes correspond to situations where variety may adjust to taxes and the *short-run* effects of taxes correspond to situations where variety is fixed, an assumption we test below. We also discuss how short-run and long-run taxes can be thought of as instrumental variables for price and variety in order to identify the

partial effect of varieties on consumer surplus.

Following the literature, we estimate the long-run effects of taxes using *cross-sectional* variation in sales tax rates and exemptions between and within counties in a difference-in-differences research design. We validate these cross-sectional results in a "border pair" subsample that compares similar counties in different states that share a state border, following the empirical strategy laid out in Holmes (1998), Dube, Lester and Reich (2010) and Hagedorn, Manovskii and Mitman (2016). The short-run effects of taxes are estimated by exploiting *time-series* variation in tax rates, and using fixed effects panel models to focus on within-store changes over time in prices and expenditures in response to changes in sales taxes. Our preferred estimates suggest a large effect of sales taxes on product variety in the long run. We also provide suggestive evidence that marginal varieties added or withdrawn in response to sales tax have relatively low market shares and we test the symmetric pass-through assumption and find that it holds in our setting. The long-run effects of taxes on output are larger (in magnitude) than the short-run effects and we find similar effects on after-tax prices in the short run and long run. Using our preferred long-run and short-run estimates, we estimate that average willingness-to-pay declines by a little more than one precent in response to an exogenous one percent decline in variety. This reflects consumers substituting away from their first choice to their next preferred choice (or not purchasing at all) when their first choice is eliminated. While this is a large valuation of variety, we show that it is an order of magnitude smaller than what one would conclude by using a Logit/CES that implicitly restricts the short-run price elasticity of demand and the variety effect to be determined by same parameter.

One concern with our approach is that the short-run and long-run effects of taxation may differ for reasons other than changes in product variety. For example, adjustment costs may give rise to long-run elasticities that exceed short-run elasticities (Chetty 2012). We provide an analysis of the LeChatelier Principle in the context of product differentiation. Our key result is to show that the formula for the variety effect remains valid even when the short-run and long-run vary for reasons other than product variety. The intuition for this result is an application of the envelope theorem: the long-run decisions of agents are set optimally and as a result, behavioral responses do not have a first-order welfare effect.

Our approach can be applied in many other settings where the welfare effect of increasing product variety is of interest. To support this claim we describe what our sufficient statistics approach has to offer to two longstanding questions in Industrial Organization and Public

Economics. First, we revisit the classic IO question of whether there is too little or too much product variety in the free-entry equilibrium. We consider a general model of imperfect competition with symmetric firms (nesting Cournot, Bertrand and Perfect Collusion, all of which are captured in a reduced-form by a conjectural variations parameter) and derive the social marginal welfare gain or loss from a small change in the number of varieties. We show that whether variety is insufficient or excessive depends on the relative magnitude of the variety effect and the "business-stealing effect", which is a negative externality arising because the marginal entrant does not account for the harm it imposes on its competitors (see Mankiw and Whinston 1986).[6]

Second, we consider is the welfare effect of taxation, which is of central interest in Public Economics. Standard formulas for measuring the welfare effects of taxes typically assume markets are competitive (Harberger 1964) or markets are characterized by imperfect competition but the number of firms and products are fixed (Auerbach and Hines 2001; Weyl and Fabinger 2013).[7] In differentiated product markets, taxes can distort product variety by reducing firm profitability, and thereby leading to exit for those firms at the margin. In this context, pass-through alone (as in Weyl and Fabinger 2013) is no longer a sufficient statistic for the welfare effects of taxation, instead we also need a measure of the variety effect. We derive a new formula for the marginal welfare gain from increasing commodity taxes in a general model of imperfect competition covering a wide range of market conduct when variety is endogenous. While the standard formula emphasizes the response of total output to the tax due to a fiscal externality, our formula shows additionally that one needs to account for the variety effect which arises since taxes distort product variety. Interestingly, the impact on welfare coming through a change in variety is ambiguous and depends on whether variety is excessive or insufficient at the prevailing equilibrium. Finally, we show how our new formula connects to existing formulas for the welfare effects of taxes under imperfect competition.

The rest of the paper proceeds as follows. Section 2 considers the symmetric model

[6]There is a vast literature on optimal product variety. This literature dates back to Spence (1976a), Spence (1976b), Dixit and Stiglitz (1977) with more recent contributions by Anderson, de Palma and Nesterov (1995), Berry and Waldfogel (1999), Gentzkow, Shapiro and Sinkinson (2014), and Berry, Eizenberg and Waldfogel (2015).

[7]Other papers in this literature include Seade (1987), Stern (1987), Myles (1989), Besley (1989), Delipalla and Keen (1992), Anderson, de Palma and Kreider (2001a, 2001b), Auerbach and Hines (2001), Weyl and Fabinger (2013) and Gillitzer, Kleven and Slemrod (2015). These papers typically assume a specific form of firm competition and impose specific structure on consumer preferences, and some of these papers focus purely on the short-run equilibrium holding the number of varieties fixed. Below we allow for both differentiated products and free entry in deriving our welfare expressions, without having to specify the form of imperfect competition (e.g. Bertrand vs Cournot).

of demand. Section 3 considers the general model with asymmetric products. Section 4 considers identification of the variety effect using short-run and long-run variation in sales taxes. Section 5 considers our empirical application. Section 6 considers several applications. Section 7 concludes.

2 Symmetric Model of Demand

For ease of exposition and to develop intuition, we begin with two classes of models with symmetric preferences, discrete choice (Random Utility) models (section 2.1.1) and continuous choice models (section 2.1.2). We show that for both classes of models, consumer surplus is equal to the integral of aggregate demand (see proposition 1). Next, we proceed to the welfare analysis of product variety, where we define and provide a graphical representation of the "price effect" and the "variety effect". We then derive a sufficient statistics formula for the variety effect under a "parallel inverse demands" assumption and consider several microfoundations for this assumption. In section 3 we introduce the general model that does not rely on symmetry and we extend our sufficient statistics formula for the variety effect.

2.1 Preferences, Demand, and Consumer Surplus

2.1.1 Symmetric Discrete Choice Model

We consider a population of statistically identical and independent consumers indexed by i of mass unity who choose to purchase a single product $j \in \{1, ..., J\}$ or the outside option $j = 0$. The number of products in the market, J, is our measure of product variety.

Preferences. The indirect utility of individual i who purchases product j is given by:

$$u_{ij}(y_i, p_j) = \alpha(y_i - p_j) + \delta_j + (1 - \sigma)\nu_i + \sigma\varepsilon_{ij} \tag{1}$$

where y_i is the consumer's income, p_j is the price of good j, δ_j is a firm-level characteristic and $(1-\sigma)\nu_i + \sigma\varepsilon_{ij}$ is an idiosyncratic match value between consumer i and product j, which captures heterogeneity in tastes across consumers and products. The utility of individual i who chooses the outside option is given by $u_{i0} = \alpha y_i + \varepsilon_{i0}$. Although linearity in price in equation ((1)) seems like a special case, Nevo (2011) shows that it may rationalizable by a quasi-linear utility function (no income effects). Specifying a distribution for the random utility shocks $(\nu_i, \varepsilon_{ij})$ gives rise to different models of discrete choice including probit, Logit,

Nested Logit or any from the Generalized Extreme Value (GEV) family.[8]

In general, we assume that for $j \neq 0$, the random utility shocks (ε_{ij}) are identical and independently (continuously) distributed (i.i.d.) and independent of (ν_i), but we allow ε_{i0} to be correlated with ν_i. In section 3, we relax the i.i.d. assumption and allow for general distributions of the random utility shocks.

Demand. Given the indirect utility function in equation (1), assuming there is 0 probability of ties, we may define the demand for product j as

$$q_j(p_1, \ldots, p_J, J) = \mathbb{P}\left(u_{ij}(y_i, p_j) = \max_{j' \in \{0,\ldots,J\}} u_{ij'}(y_i, p_{j'})\right). \tag{2}$$

Imposing symmetry, $\delta_j = \delta$ and $p_j = p$ for all $j = 1 \ldots J$, we may express aggregate demand (for all products excluding the outside good) when J varieties are available as:

$$Q(p, J) = \sum_{j=1}^{J} q_j(p, J). \tag{3}$$

Similarly, define $P(Q, J)$ to be the inverse aggregate demand corresponding to $Q(p, J)$.[9]

2.1.2 Symmetric Continuous Choice Model

We now introduce a class of continuous choice models.

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

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

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

Demand. The consumer's problem is

$$\max u_J(q_1, \ldots, q_J, m) = h_J(q_1, \ldots, q_J) + m \tag{4}$$
$$\text{subject to } m + \sum_{j=1}^{J} p_j q_j = y.$$

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

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

[8] See Train (2003).

[9] Given symmetry in preferences (the quality parameter δ does not depend on i or j) the symmetry in prices $p_j = p$ is motivated by assuming symmetric firms and a symmetric price equilibrium.

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

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

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

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

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

2.1.3 Consumer Surplus

In the discrete choice model we define consumer surplus as the expected maximum utility normalized by the marginal utility of income:

$$CS(p_1, \ldots, p_J, J) = \frac{1}{\alpha}\mathbb{E}\left[\max_{j\in\{0,\ldots,J\}} u_{ij}(y_i, p_j)\right]$$

while for the continuous choice model consumer surplus is given by the indirect utility function $u^*(p, J, y) = H_J(Q(p, J)) + y - p * Q(p, J)$.

Proposition 1. *For both models of preferences as defined in equations* (1) *and* (4)*, under the assumption of no income effects and symmetric prices $p_j = p$, we can represent consumer surplus (up to a constant) as*

$$CS(p, J) = \int_p^{\infty} Q(s, J)ds. \tag{5}$$

Thus, consumer surplus is equal to the integral of aggregate demand.[10]

[10] If we allow for income effects, the correct measure of welfare is either compensating variation or equivalent variation. In that case we can get the analogous result that compensating variation is the integral of the aggregate compensated demand. See example 4.

Proof. See Appendix. □

Finally, observe when there is a continuum of varieties then $CS(p, J) = \int_p^\infty Q(s, J)ds$ is differentiable in both p and J.

2.2 Welfare Effects of Variety

In this section, we consider the impact of a small change in the number of varieties J on consumer surplus. The starting point of this section is the expression for consumer surplus as the integral of aggregate demand $CS(p, J) = \int_p^\infty Q(s, J)ds$ which we have proved for both the discrete choice and continuous choice models. In what follows we we will assume that J is a continuous variable and $Q(p, J)$ is defined for any $J \in \mathbb{R}$ and is continuously differentiable.[11] We first introduce some definitions.

Definition 1. The "price effect" is defined as

$$-Q\frac{dp}{dJ}. \tag{6}$$

It arises since market prices may change when firms enter or exit the market. Due to the envelope theorem, there is no first-order effect on welfare due to consumer re-optimization when prices change, so only the mechanical effect of a price change affects consumer welfare.

Definition 2. The "variety effect" is defined as

$$\Lambda(Q, J) \equiv \int_{P(Q,J)}^\infty \frac{\partial Q}{\partial J}(s, J)ds. \tag{7}$$

Holding the effect on prices constant, a new variety increases welfare since consumers exhibit a "love of variety". The variety effect depends on how aggregate demand responds to a change in variety.

The total effect on consumer welfare is the sum of the price effect and the variety effect:

$$\frac{dCS}{dJ} = -Q\frac{dp}{dJ} + \Lambda. \tag{8}$$

Each term in equation (8) is illustrated in Figure 1 where we consider a reduction in product variety ($\Delta J < 0$). The price effect is the rectangular area where the base is the pre-existing output level and the height is the change in prices. The variety effect is given

[11] For the continuous choice model we have shown how this is a natural assumption if there is a continuum J of varieties. In the Appendix we show how in the case of discrete varieties the differentiability assumption can be reinterpreted if we allow for probabilistic entry.

by the area between the demand curves. To see the intuition for the variety effect, consider a market where consumers choose among a set of products. The aggregate demand curve for these products is given by equation (3). The inverse of this curve gives the maximum willingness to pay for products in the market across individuals. Now consider a reduction in the number of products available. In this case, some consumers will no longer be able to purchase their most preferred option. Thus, the maximum possible utility attainable will be lower for these consumers. We can represent this a shift down in the inverse aggregate demand curve. The area between the inverse aggregate demand curves before and after the change in variety (above the market price) corresponds exactly to the variety effect.

Another way to see this is to define the average change in willingness to pay for infra-marginal units as variety J changes as

$$\overline{\frac{\partial P}{\partial J}}(Q,J) = \frac{1}{Q}\int_0^Q \frac{\partial P}{\partial J}(s,J)ds.$$

We can then show the variety effect is the total change in willingness to pay for the units that are actually exchanged:

$$\Lambda = \int_p^\infty \frac{\partial Q}{\partial J}(s,J)ds = \int_0^Q \frac{\partial P}{\partial J}(s,J)ds = Q\overline{\frac{\partial P}{\partial J}}. \tag{9}$$

Clearly, if we could non-parametrically identify the demand curves before and after the change in variety, we can exactly compute the variety effect. However this might not be possible in practice. A key objective of this paper is to establish a method to identify the variety effect using reduced-form methods based on local information. The next theorem states that one may recover the variety effect in the symmetric model using a "sufficient statistics" approach (Chetty 2009).

Theorem 1. *The variety effect Λ is equal to the average change in willingness to pay times the quantity demanded:*

$$\Lambda(Q,J) = Q\overline{\frac{\partial P}{\partial J}}(Q,J) \tag{10}$$

Furthermore, if the inverse aggregate demands are parallel, $\frac{\partial P}{\partial Q}(Q,J) = \frac{\partial P}{\partial Q}(Q,J')$ for all J, J' and Q, then the average change in willingness to pay is equal to the marginal change in willingness to pay, $\overline{\frac{\partial P}{\partial J}}(Q,J) = \frac{\partial P}{\partial J}(Q,J)$. Therefore we obtain:

$$\overline{\frac{\partial P}{\partial J}} = \left(\frac{dP}{dQ} - \left.\frac{dP}{dQ}\right|_J\right)\frac{dQ}{dJ} \tag{11}$$

where $\left.\frac{dP}{dQ}\right|_J = \frac{\partial P}{\partial Q}$ denotes the slope of inverse demand when variety J is held fixed and $\frac{dP}{dQ} =$

$\frac{dP(Q(J),J)}{dJ} / \frac{dQ}{dJ}$ *denotes the slope of inverse demand when J is variable.*

Proof: See Appendix.

The expression for the variety effect in equation (10) and the average change in willingness-to-pay in equation (11) can be most easily understood geometrically. Figure 2 considers a reduction in variety and shows that the variety effect is the area of the rectangle with base Q_0 and height $d \equiv P_1 - P^*$. The base Q_0 is simply the aggregate output prior to the change in variety and is observable. The height d captures the change in willingness to pay as J changes, therefore $d = \frac{\partial P}{\partial J}$. However, d is not directly observable, it depends on the price induced by the change in variety, P_1, which is observable and the market price that would prevail at the same level of output but on the original demand curve, P^*, which is unobservable. To see how to recover an expression for d, consider a change in variety from J_0 to J. Note from Figure 2 that d must satisfy the following relationship $Q(P(J), J) = Q(P(J) - d, J_0)$. To solve for d, note that for a small change in J:

$$\begin{aligned} dQ &\approx Q(P(J), J) - Q(P_0, J_0) \\ &= Q(P(J) - d, J_0) - Q(P_0, J_0) \\ &\approx \frac{dQ}{dP}|_{J_0}(-d + P(J) - P_0) \\ &\approx \frac{dQ}{dP}|_{J_0}(-d + \frac{dP}{dQ}dQ). \end{aligned}$$

Thus, rearranging and solving for d yields:

$$d \approx \left(\frac{dP}{dQ} - \left.\frac{dP}{dQ}\right|_J \right) dQ. \tag{12}$$

In economic terms, d gives the reduction in the willingness to pay for the marginal unit. The upshot is that it's possible to identify d using the price elasticities of demand when J is fixed and when J is variable. In our empirical application, we interpret these as the "short-run" and the "long-run" price elasticities of demand, respectively. As long as one can estimate the short-run and the long-run price elasticity of demand, then one can recover the variety effect using purely reduced-form methods. In Section 4, we will show how to recover the variety effect associated with a commodity tax change.

2.3 Microfoundations for Parallel Demands

In this section, we consider the conditions under which aggregate demands satisfy the parallel demands assumption. We start with discrete choice models and then move to continuous

choice models.

2.3.1 Discrete Choice Models

We first describe the Nested Logit model which features parallel aggregate demands. Next we prove an asymptotic result that a wide class of random utility models feature parallel demands.

Example 1. In equation ((1)), if the random utility shocks (ε_{ij}) are drawn from the Gumbel distribution, and $(1-\sigma)\nu_i$ has the distribution derived in Cardell (1997), then this model corresponds to the Nested Logit model in which there are only two nests: one which includes $j = 1, ..., J$ and the other which includes only the outside option $j = 0$. Then

$$q_i(p_1, \ldots, p_J, J) = \frac{\left(\sum_{j=1}^{J} e^{\frac{\delta_j - \alpha p_j}{\sigma}}\right)^{\sigma}}{1 + \left(\sum_{j=1}^{J} e^{\frac{\delta_j - \alpha p_j}{\sigma}}\right)^{\sigma}} \frac{e^{\frac{\delta_i - \alpha p_i}{\sigma}}}{\sum_{j=1}^{J} e^{\frac{\delta_j - \alpha p_j}{\sigma}}}.$$

As the parameter σ goes to 0, the only random term in (1) is ν_i which is constant across all $j \neq 0$. When $\sigma = 1$, we retrieve the Logit model. The parameter σ/α characterizes a consumer's "love of variety". When σ/α is large, consumers value greater variety (higher J).[12] If we focus on symmetric varieties $\delta_j = \delta$ and a symmetric price equilibrium $p_j = p$, aggregate demand is equal to:

$$Q(p, J) = \frac{J^{\sigma} e^{\delta - \alpha p}}{1 + J^{\sigma} e^{\delta - \alpha p}}.$$

Observe then the inverse aggregate demand curve is given by

$$P(Q, J) = \frac{\delta}{\alpha} + \frac{\sigma}{\alpha} \log J - \frac{1}{\alpha} \log\left(\frac{Q}{1-Q}\right)$$

which is separable in J and Q. This implies that exogenous shifts in variety move the inverse aggregate demand curve in parallel, therefore the assumptions of theorem 1 are satisfied. Additionally, the coefficient on log variety is different than the coefficient on output. Thus, in this model, the value of additional variety is not pinned down fully by the price elasticity of demand, which is in contrast to the standard models considered in the trade literature (see for example, Feenstra 1994; Broda and Weinstein 2006).[13] In the Nested Logit model, consumer

[12] When $\sigma = 1$, the love of variety parameter $\frac{1}{\alpha}$ is inversely related to the elasticity of substitution considered in trade models with a representative consumer with a CES utility function. In fact, the Logit model aggregates to the CES model if we substitute $log(p)$ instead of prices p (Anderson, de Palma and Thisse 1987). The formal connection requires one to introduce a second stage where individuals choose a continuous quantity of the good.

[13] Benassy (1996) recognizes that in the usual CES parametrization taste for variety and the elasticity of substitution are intertwined and provides a CES parametrization where the variety effect is identified separately from the elasticity of substitution. See example 4 for further discussion.

surplus is given by the familiar "log sum" expression $CS = \frac{1}{\alpha}\log\left(1 + J^{\sigma}e^{\delta-\alpha p}\right)$, also referred to as the "inclusive value".[14] Thus, we get a closed-form solution for the variety effect. In particular, one can show that $\Lambda = \frac{\sigma}{\alpha}q$. The variety effect in this parametric model is equal to the average willingness to pay for a unit of the additional variety ($\frac{\sigma}{\alpha}$) multiplied by the market share (q) of the new variety.[15] Intuitively, if product variety increases by one, then evaluating the effect on consumer surplus requires the post-entry market share of the new good. If this is zero, then consumers don't value the new good and hence the welfare gains are nil.

Next, we show that there is a large class of models that admit a Nested Logit approximation. The random utility models in this class have in common that the distribution of the maxima of the shocks is asymptotically Gumbel, which implies that the aggregate inverse demands are asymptotically parallel. Therefore, for any model in this class, the structural estimates of the price effect and the variety effect will asymptotically converge to the structural estimates of the Nested Logit, which in turn can be estimated with reduced-form parameters, namely the price elasticities of demand when J is fixed and when J is variable.

Summarizing, a sufficient condition to get inverse parallel demands is that the random utility shocks (ε_{ij}) are i.i.d. Gumbel, independent of the size of σ, the distribution of ν_i and the distribution of ε_{i0}.[16] Furthermore, if the shocks (ε_{ij}) are assumed to be i.i.d., then they have to be Gumbel in order to satisfy the inverse parallel demands condition, as we show in the Appendix.[17] Nonetheless, using results from Extreme Value Theory we can show that for a large class of iid shocks, the inverse demands are asymptotically parallel. We now define the class of models that admit the asymptotic approximation, and provide a useful sufficient condition to show that a given model is in the class.

Definition 3. Let (ε_j) be i.i.d. distributed according to a continuous cdf F. We say that F is in the domain of attraction of the Gumbel distribution if

$$\max_{j\in\{1,\ldots,J\}} \varepsilon_j \overset{a}{\sim} Gumbel(\mu(J), \sigma(J)),$$

as $J \to \infty$ for some location and dispersion parameters $(\mu(J), \sigma(J))$.

[14] See Train (2003) for derivation.

[15] In the Nested Logit model, the assumptions of theorem 1 are satisfied, therefore average change in WTP is equal to the marginal change in WTP and given by $\frac{\sigma}{\alpha J}$ which is decreasing in J.

[16] Remember the starting point is the random utility specification $u_{ij} = \alpha(y_i - p_j) + \delta_j + (1-\sigma)\nu_i + \sigma\varepsilon_{ij}$, and $u_{i0} = \alpha y_i + \varepsilon_{i0}$.

[17] Without the independence assumption, there are other distributions of shocks that also give rise to parallel demands. If we do not assume independence of the shocks then the Gumbel distribution is not necessary. In this case, for any family of (downward sloping, increasing in J) parallel inverse demands there is a joint density of shocks that rationalizes it. In general, the shocks are not Gumbel and are correlated in this construction. See Appendix for proof.

Lemma 1. *Let x_0 be the supremum of the support of a cdf F that is twice continuously differentiable. If F satisfies that $\lim_{x\to x_0} \frac{F''(x)(1-F(x))}{F'^2} = -1$ then F is in the domain of attraction of the Gumbel distribution.*

See Resnick (1987) for a proof of the lemma and a full characterization of the domain of attraction of the Gumbel distribution. The characterization is outside the scope of the paper and the lemma is enough for our purposes. For example, if (ε_j) are iid $N(0, \sigma^2)$ or exponential the above lemma applies.

The next theorem provides the microfoundation for our key assumption of parallel inverse demands and states that inverse demands become parallel as variety increases for any random utility model with shocks in the Gumbel domain of attraction.

Theorem 2. *Let the random utility shocks (ε_j) be i.i.d. and distributed according to F in the domain of attraction of the Gumbel distribution. Then, for any large enough J and K there exists d such that for all $p \in \mathbb{R}$ we have $Q(p, J) \approx Q(p + d, K)$. Specifically, for all p we have*

$$Q(p, J) = \mathbb{P}\left(\max_{j\in\{1,\ldots,J\}} u_{ij}(p) > u_{i0}\right) \overset{a}{\sim} \mathbb{P}\left(\max_{j\in\{1,\ldots,K\}} u_{ij}(p + d) > u_{i0}\right) = Q(p + d, K)$$

Therefore the inverse demands are approximately parallel $P(Q, K) \approx P(Q, J) + d$ for all Q, for large enough J and K.

Proof: See Appendix.

In Figure 3, we assess this approximation theorem by numerically simulating different random utility models and calculating the bias that arises from assuming demands are parallel. Specifically, we simulate a model of a large number of consumers choosing with utility over products given by equation (1). We choose $\alpha = 1$ and $y = 1$ in the simulation and we consider a range of different shock distributions (Gumbel, Normal, Gamma, and Pareto). We then repeat this procedure for a range of different values of J to assess how the bias from assuming parallel demand varies with J when consider a hypothetical 20 percent increase in the number of products (from initial value of J). We compute the welfare gains exactly using numerical methods and compare the exact welfare gain to the approximate gains implied by assuming parallel demands based on the formula in equation (10). The results in Figure 3 show that the bias that arises from assuming parallel demands is a function of the number of varieties in the market, where bias is measured as the difference between the estimated (approximate) variety effect and the exact variety effect. The benchmark distribution is Gumbel where we know from theory that the demand curves are exactly parallel and therefore the bias is zero. For

both the Normal and Exponential distributions, we find that the bias is small in magnitude and converges to 0 fairly quickly as the number of varieties increase. On the other hand, with a Pareto distribution, there is a bias of roughly 20 percent, which does not vanish as varieties increase. In this case, the variety effect computed using our sufficient statistics formula is a lower bound on the true variety effect.

Discrete choice models of demand are used in many economic applications. In Labor Economics, classic examples are labor supply (for recent examples, see Card, Cardoso, Heining and Kline 2017, Lamadon, Mogstad, and Setzler 2017). They are also used in Urban Economics; in particular, in spatial location models (see Busso, Gregory and Kline 2013, Diamond 2016, Suarez Serrato and Zidar 2016). Typically in these models, the random error term is assumed to be distributed as Extreme Value Type I which delivers closed forms for demand and welfare. Theorem 2 shows that if the set of alternatives is sufficiently large, then there are a broader set of distributions for the error term that can rationalize these. The key is that these distributions are in the domain of attraction of the Gumbel (or Extreme Value Type I) distribution.

2.3.2 Continuous Choice Models

Inverse aggregate demands are parallel, $\frac{\partial P}{\partial Q}(Q, J) = \frac{\partial P}{\partial Q}(Q, J')$ for all J, J' and Q, if and only if they are linearly separable in Q and J. Therefore any family of parallel inverse demands can be written as $P(Q, J) = a(J) - f(Q)$ for increasing and differentiable functions $a(J)$ and $f(Q)$. Letting F be any antiderivative of f. Then

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

for some $\rho \in (0, 1)$ is a direct utility function that rationalizes $P(Q, J)$ given that firms are playing a symmetric price equilibrium. In general, when J is discrete:

$$u_J(q, m) = a(J) h_J(q) - F\left(h_J(q)\right) + m$$

rationalizes $P(Q, J)$ for any $h_J : \{1, \ldots, J\} \to \mathbb{R}$ which is symmetric in all its arguments, continuously differentiable, strictly quasi-concave, with $h_J(0, \ldots, 0) = 0$, and $h_J(q, \ldots, q) = Q$.

We now describe examples of demand functions generated from continuous choice models that feature parallel demands.

Example 2. Bulow and Pfleiderer (1981) obtain the following three categories of inverse

demands as the unique curves with the property of constant pass-through:

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

An important case is when $\beta_J = \beta$ for all J, then the inverse aggregate demands are linearly separable in J and Q and they shift in parallel as J moves.[18] The fact that these are the only class of curves for which marginal costs are passed-on in a constant fraction makes them a tractable benchmark and therefore they have been popular in applied work. Fabinger and Weyl (2016) generalize Bulow and Pfleiderer (1983) and characterize a broader class of "tractable equilibrium forms" of the form $P(Q, J) = \alpha_J + \beta Q^t + \gamma Q^u$ which allow for greater modeling flexibility. Again, as long as β and γ are independent of J, then we say that the inverse demands shift in parallel.

Example 3. This example shows that our revealed-preference approach allows for rational preferences that display *hate-of-variety* ($a'(J) < 0$). Imagine there is a marginal cost of consumption cJ for each unit of some good that is consumed; that is, for each unit consumed, the agent faces a constant cost of evaluating each of J varieties before he chooses. Preferences are given by

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

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

Example 4. This example illustrates two limitations of our approach regarding the no-income effects and parallel demands assumptions. When there are income effects we cannot use the

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

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

Marshallian demands to estimate welfare, however we extend our analysis by turning to Hicksian compensated demands. This might be of special importance for international trade models (e.g. Broda and Weinstein 2006; Arkolakis, Costinot and Rodriguez-Clare 2012) in which both gains from variety and general equilibrium effects are important. We also show the similarity of our flexible approach to estimate welfare gains from variety to the approach in Benassy (1996) and Behrens, Kanemoto and Murata (2017) in the case of CES sub-utility functions.

Let

$$U_J(q_0, (q_j)_{j=1}^J) = b(J) q_0^{1-\alpha} \left(\frac{1}{J} \int_0^J q_j^\rho dj \right)^{\frac{\alpha}{\rho}}.$$

Fix $p = p_j$ to be the symmetric price of the inside market, let income be I, and the price of q_0 be normalized to 1. Then the Marshallian aggregate demand for the inside good is independent of J and given by $Q(p, I, J) = \frac{\alpha I}{p}$, but the Hicksian compensated demands crucially depend on $b(J)$:

$$Q^C(p, U, J) = \frac{U}{b(J)} \left(\frac{\alpha}{p(1-\alpha)} \right)^{1-\alpha}.$$

In this case, using either compensating variation or equivalent variation[19] as a measure of consumer welfare we can obtain the variety effect:

$$\Lambda(p, J) = \frac{\partial CV}{\partial J}(p, J) = \frac{\partial EV}{\partial J}(p, J) = I \frac{b'(J)}{b(J)}.$$

As a comparison, under the classical version of CES used by Dixit and Stiglitz (1977) we have $b(J) = J^{\frac{\alpha}{\rho}}$, but under Benassy's parametrization $b(J) = J^\nu$. The implied variety effect in each of these cases is given by:

$$\Lambda_{DS} = I \frac{\alpha}{\rho J}$$

$$\Lambda_{Benassy} = I \frac{\nu}{J}$$

Note that in Dixit and Stiglitz's parametrization the variety effect is pinned down by the

[19] The compensating variation starting from period 0 varieties J_0 and inside market price p_0 as a function of period 1 varieties J_1 and inside market price p_1 is:

$$CV(p_1, J_1; p_0, J_0) = I_0 - I_0 \frac{b(J_0)}{b(J_1)} \left(\frac{p_1}{p_0} \right)^\alpha$$

Similarly, the equivalent variation is given by:

$$EV(p_1, J_1; p_0, J_0) = I_0 \frac{b(J_1)}{b(J_0)} \left(\frac{p_0}{p_1} \right)^\alpha - I_0$$

substitution parameter ρ and the share of income devoted to the inside market α, while in Benassy's parametrization the variety effect is identified separately from the elasticity of substitution. An alternative approach to obtain the variety effect is to integrate the difference of the appropriate aggregate compensated demands, although in this particular example the inverse compensated demands are not parallel.[20]

3 General Model: Asymmetric Demands and Prices

The symmetric model of section 2 helped us to build intuition and microfound our key assumptions, however it is not realistic enough to be taken to the data. Here we introduce the general model with asymmetric product characteristics and prices. We start with the description of the preferences for both discrete choice and continuous choice and then proceed in a unified treatment using aggregate demands.

Preferences (discrete choice). We consider a population of statistically identical and independent consumers of mass unity. The utility of individual i who purchases product $j = 0, \ldots, J$ is given by:

$$u_{ij}(y_i, p_j) = \alpha(y_i - p_j) + \eta_{ij}$$

where y_i is the consumer's income, p_j is the price of good j (where we normalize $p_0 = 0$) and η_{ij} is the random utility term, which captures heterogeneity in tastes across consumers and products. We allow (η_{ij}) to have a general joint cumulative distribution function F over $\mathbb{R}^J$ that admits a density function f.[21] This corresponds to the general Additive Income Random Utility Model (AIRUM) in McFadden (1981) which satisfies sufficient conditions for aggregation as first noted by Gorman (1953).[22]

Given there is 0 probability of ties, then we define the demand for product j as

$$q_j(p_1, \ldots, p_J, J) = \mathbb{P}\left(u_{ij}(y_i, p_j) = \max_{j' \in \{0,\ldots,J\}} u_{ij'}(y_i, p_{j'})\right)$$

[20]This is a result of both p and $b(J)$ entering multiplicatively into $Q^C(p, U, J)$. The bias of the formula based on parallel compensated demands is given by $\frac{\Lambda_{parallel}}{\Lambda} = \frac{\alpha}{1-\alpha}$ for any $b(J)$.

[21]The condition for F to be characterized by a density function f is that it be absolutely continuous with respect to the Lebesgue measure in $\mathbb{R}^J$. When $\eta_{ij} = \delta + (1-\sigma)\nu_i + \sigma\varepsilon_{ij}$ with (ε_{ij}) i.i.d. and independent of (ν_i) we obtain the symmetric model of section 2. The extent into which this formulation captures the Random Coefficients Logit Model is discussed below in example 6.

[22]Alternatively, our primitive could be a translation-invariant probabilistic choice system (TPCS). A probabilistic choice system is translation invariant if the choice probabilities are the same when the decision maker faces prices $(p_0, \ldots, p_J)$ and $(p_0 + s, \ldots, p_J + s)$. An AIRUM induces a TCPS and the other way around, for details see McFadden (1981).

and aggregate demand (for all products excluding the outside good) is defined as

$$Q(p_1, \ldots, p_J, J) = \sum_{j=1}^{J} q_j(p_1, \ldots, p_J, J).$$

Preferences (continuous choice). Let the representative consumer's utility function when J varieties are available, be given by

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

for any $h_J : \{1, \ldots, J\} \to \mathbb{R}$ which is continuously differentiable, strictly quasi-concave and $h(0, \ldots, 0) = 0$ and where the linear good m is interpreted as money. Importantly, h_J is no longer assumed to be symmetric.

Let $q_j(p_1, \ldots, p_J, J)$ be the Marshallian demand for variety j and $Q(p_1, \ldots, p_J, J) = \sum_{j=1}^{J} q_j(p_1, \ldots, p_J, J)$ the aggregate demand.

The following proposition contains an expression that generalizes the integral form of the variety effect in definition 2 to the case of asymmetric demands and prices.[23]

Proposition 2. *Consider the case where the number of products goes from J to M (with $M > J$). In this case, the welfare gain from variety or variety effect is equal to the following line integral:*

$$\Lambda = \int_0^{\infty} \sum_{j=J+1}^{M} q_j(\mathbf{p_J}, p_{J+1} + s, p_{J+2} + s, \ldots, p_M + s) ds$$

where $\mathbf{p_J} = (p_1, p_2, ..., p_J)$ are the prices for the existing J products, and $(p_{J+1}, \ldots, p_M)$ are the prices at which the new $M - J$ products are introduced.

Proof. See Appendix. □

The intuition is that the welfare gain from the new varieties can be calculated by integrating the demand for new varieties.[24] This result implies that if we non-parametrically identify the demand curves before and after the change in variety, we can exactly compute the variety effect. More often, in practice the econometrician does not have enough information to identify the demands non-parametrically (e.g. this would require to observe prices for which the demand of some variety is arbitrarily close to 0 but these prices might never show up in the data), however the econometrician always has information that is local to the

[23] This holds for both continuous and discrete choice as long as preferences are quasi-linear in money. In the case where there are income effects the analogous expression holds by substituting the Marshallian demands with Hicksian compensated demands. For a derivation in the discrete choice case when only one variety is introduced and income effects are allowed see Bhattacharya (2016).

[24] Notice that the prices of the old varieties are kept fixed, while the integral is taken over uniform increases in the price of the new varieties.

market equilibrium. The main objective in what follows is to give plausible and parsimonious sufficient conditions to identify the variety effect using reduced-form methods based on local information.

Similar to proposition 1, the variety effect can also be stated in terms of aggregate demands as:

$$\Lambda = \int_0^\infty Q_M(\mathbf{p_M} + s\mathbf{1_M})ds - \int_0^\infty Q_J(\mathbf{p_J} + s\mathbf{1_J})ds$$

where $\mathbf{1_K}$ is a $K-$dimensional vector of ones, and $\mathbf{p_M} = (\mathbf{p_J}, p_{J+1}, \ldots, p_M)$.

Next, we assume that there exists some price index d such that $Q_M(\mathbf{p_M} + (s+d)\mathbf{1_M}) = Q_J(\mathbf{p_J} + s\mathbf{1_J})$ for all $s \in \mathbb{R}$.[25] In other words, increase prices starting from $\mathbf{p_M}$ by some constant amount d until total quantity demanded equals quantity demanded when there are J products in the market. Under this assumption, it follows that:

$$\Lambda = \int_0^d Q_J(\mathbf{p_J} - s\mathbf{1_J})ds.$$

By the mean value theorem for integrals, there exists $d' \in [0, d]$ such that

$$\Lambda = \int_0^d Q_J(\mathbf{p_J} - s\mathbf{1_J})ds = d * Q_J(\mathbf{p_J} - d'\mathbf{1_J}).$$

Intuitively, as $M \to J$ we have $d' \to d$ and we obtain a first order approximation for the variety effect:

$$\Lambda \approx d * Q_J(\mathbf{p_J}).$$

We summarize these observations in the following theorem.

Theorem 3. *Assume that for all M and J there exists some d such that $Q_M(\mathbf{p_M} + (s + d)\mathbf{1_M}) = Q_J(\mathbf{p_J} + s\mathbf{1_J})$ for all $s \in \mathbb{R}$. Then there exists $d' \in [0, d]$ such that*

$$\Lambda = d * Q_J(\mathbf{p_J} - d'\mathbf{1_J}). \tag{13}$$

Furthermore, let $\mathbf{p}_M^1 = \mathbf{p}_M + s\mathbf{1_M}$, $\Delta P = s$, and $\Delta Q = Q_M(\mathbf{p_M^1}) - Q_J(\mathbf{p_J})$ then

$$d = \left(\frac{\Delta P}{\Delta Q} - \left.\frac{dP}{dQ_J}\right|_J\right)\Delta Q + o\left((s-d)^2\right) \tag{14}$$

where $\left.\frac{dP}{dQ_J}\right|_J = \left.\left(\frac{dQ_J(\mathbf{p_J}+t\mathbf{1_J})}{dt}\right)^{-1}\right|_{t=0}$.

Proof. Observe by assumption $Q_M(\mathbf{p_M^1}) = Q_J(\mathbf{p_J} + (s-d)\mathbf{1_J})$, then the second part of the theorem follows directly from the first-order Taylor approximation:

$$Q_M(\mathbf{p_M^1}) = Q_J(\mathbf{p_J}) + (s-d)\frac{dQ_J(\mathbf{p_J} + t\mathbf{1_J})}{dt} + o\left((s-d)^2\right)$$

[25] This is related to the price index in Feenstra (1994). However, in Feenstra, the price index is defined as the (common) price reduction that would have to occur when there are J goods in the market in order to give the same *utility* as when there are M goods.

where $\frac{dQ_J(\mathbf{p_J}+t\mathbf{1_J})}{dt}$ is the directional derivative in the direction $\mathbf{1_J}$. $\square$

We now have an expression for the variety effect in (13) that is similar to equation (10), and an expression for the average change in the willingness-to-pay (14) similar to (11), therefore theorem 3 is a general version of theorem 1.

Several features of theorem 3 are worth highlighting. First, observe we defined $\mathbf{p}_M^1 = \mathbf{p}_M + s\mathbf{1_M}$, this is equivalent to assuming that all prices adjust uniformly after the introduction of the new varieties. Although it seems restrictive, having prices adjust in the same direction $\mathbf{1_J}$ as the vertical shift d allows us to identify d by a simple application of the Taylor approximation.[26] In the Appendix we discuss how this assumption can be relaxed, however uniform price adjustments is a testable assumption and in our empirical application below we show that the data is consistent with it.

Second, we interpret the directional derivative $\left.\frac{dQ_J}{dP}\right|_J = \frac{dQ(\mathbf{p_J}+t\mathbf{1_J})}{dt} = \sum_{j=1}^{J} \frac{\partial Q_J}{\partial p_j}$ as the short-run slope of aggregate demand in the direction of uniform price changes, that connects the interpretation of (14) with equation (11) in the symmetric model, namely that d can be captured using the difference in the slopes of inverse demands when variety changes and when variety is fixed:

$$d = \left(\frac{dP}{dQ} - \left.\frac{dP}{dQ_J}\right|_J\right) dQ.$$

Furthermore, if we observe the change in aggregate demand Q_J when all prices are increased simultaneously, we do not need to estimate each partial derivative separately and we can directly estimate $\left.\frac{dQ_J}{dP}\right|_J$. When we take the model to the data, we assume a tax change is passed on to the consumers symmetrically across products, therefore the tax inducing the uniform price change that allows to estimate the total derivative $\left.\frac{dQ_M}{d\mathbf{p}}\right|_M$ without having to estimate the partial derivatives.

Finally, the most important assumption in theorem 3 was the existence of the price index d such that $Q_M(\mathbf{p_M} + (s+d)\mathbf{1_M}) = Q_J(\mathbf{p_J} + s\mathbf{1_J})$ for all s. To understand the result in terms of inverse demands, observe that for any $Q \in \mathbb{R}^+$ there is a unique s such that $Q_J(\mathbf{p^J}+s\mathbf{1_J}) = Q$, this implicitly defines the inverse aggregate demand $P(Q, J)$. Furthermore, $Q_M(\mathbf{p_M}+(s+d)\mathbf{1_M}) = Q_J(\mathbf{p_J}+s\mathbf{1_J})$ is equivalent to $P(Q, J)+d = P(Q, M)$ for all $Q \in \mathbb{R}^+$, which shows that the inverse aggregate demands are parallel. Moreover, by theorem 3, the

[26] A class of models for which we obtain symmetric pass-through is that of linear-quadratic revenues with constant marginal costs and Bertrand pricing. However, it might be unreasonable to assume uniform price adjustments when, e.g., there is one large incumbent with several products and several small entrants with one product each.

price index d is equal to the average change in willingness-to-pay. As the next example shows, in the Nested Logit model d has a closed form.

Example 5. We revisit the Nested Logit model from example 1 in the general case where attributes δ_j and prices p_j are not assumed to be symmetric. We describe the aggregate demands as a function of the price level and show that the inverse are parallel.

In the Nested Logit model aggregate demand is given by

$$Q_J(p_1, \ldots, p_J) = \frac{\left(\sum_{j=1}^{J} e^{\frac{\delta_j - \alpha p_j}{\sigma}}\right)^{\sigma}}{1 + \left(\sum_{j=1}^{J} e^{\frac{\delta_j - \alpha p_j}{\sigma}}\right)^{\sigma}}.$$

Consider, as in proposition 2, the case where the number of products goes from J to M (with $M > J$), where $\mathbf{p_J} = (p_1, p_2, ..., p_J)$ are the prices for the existing J products, and $(p_{J+1}, \ldots, p_M)$ are the prices at which the new $M - J$ products are introduced. The inverse demands are parallel if there exists d such that for all s we have

$$Q_J(\mathbf{p^J} + (s-d)\mathbf{1_J}) = \frac{\left(\sum_{j=1}^{J} e^{\frac{\delta_j - \alpha(p_j + s - d)}{\sigma}}\right)^{\sigma}}{1 + \left(\sum_{j=1}^{J} e^{\frac{\delta_j - \alpha(p_j + s - d)}{\sigma}}\right)^{\sigma}} = \frac{\left(\sum_{j=1}^{M} e^{\frac{\delta_j - \alpha(p_j + s)}{\sigma}}\right)^{\sigma}}{1 + \left(\sum_{j=1}^{M} e^{\frac{\delta_j - \alpha(p_j + s)}{\sigma}}\right)^{\sigma}} = Q_M(\mathbf{p^M} + s\mathbf{1_M}).$$

From where we obtain:

$$d = \frac{\sigma}{\alpha} log \left(\frac{\sum_{j=1}^{M} e^{\frac{\delta_j - \alpha p_j}{\sigma}}}{\sum_{j=1}^{J} e^{\frac{\delta_j - \alpha p_j}{\sigma}}} \right)$$

the average change in willingness to pay.

Example 6. One of the most important tools of the Industrial Organization economist is the Random Coefficients Logit model popularized by Berry, Levinsohn, and Pakes (1995).[27] In this model, the indirect utility for consumer i of choosing alternative j is

$$v_{ij} = V_{ij} + \varepsilon_{ij} = \underbrace{x_j \beta_i - \alpha_i p_j + \xi_j}_{V_{ij}} + \varepsilon_{ij}.$$

If the ε_{ij} are distributed Gumbel then

$$q_{ij} = \frac{e^{V_{ij}}}{1 + \sum_{k=1}^{J} e^{V_{ik}}}$$

Consider, as in Example 5, the case where the number of products goes from J to M (with $M > J$). The welfare effect of variety in money metric is given by the population average

[27] See also Nevo (2003) for a discussion of welfare estimates in this model under different assumptions.

equivalent variation:

$$EV = \int \underbrace{\frac{ln(\sum_{j=0}^{M} e^{V_{ij}}) - ln(\sum_{j=0}^{J} e^{V_{ij}})}{\alpha_i}}_{EV_i} dF_D(i)dF_\eta(\eta) \tag{15}$$

where F_D represents the demographics probability distribution over (α_i, β_i) and the distribution F_η captures the random part of β_i in case there is.

Conditional on the random coefficients we can see from example 5 and theorem 3 that a first order approximation to the individual i's equivalent variation is $\widehat{EV_i} = Q_i * d_i$ for:

$$d_i = \frac{1}{\alpha_i} log \left(\frac{\sum_{j=1}^{M} e^{V_{ij}}}{\sum_{j=1}^{J} e^{V_{ij}}} \right) \tag{16}$$

and $Q_i = \sum_{j=1}^{J} q_j$. Note that d_i has a a similar expression to EV_i in expression (15), however d_i does not depend on the value of the outside option by not including the term for V_{i0}. Next, averaging over the random coefficients we obtain:

$$\widehat{EV} = \int [Q_i * d_i] \, dF_D(i)dF_\eta(\eta).$$

Finally, note that if the marginal utility of money α_i is constant across the population i, and either the taste parameter β_i is constant across i or the products are symmetric ($x_j\beta_i + \xi_j$ is constant across j) then $d_i = d$ for all i. In that case, the equivalent variation takes the simple form:

$$\widehat{EV} = d \int Q_i dF_D(i)dF_\eta(\eta) = d * Q.$$

In Table 1, we present results from simulations designed to assess bias from the parallel inverse demands assumption when the α_i and β_i coefficients are allowed to be random. The model simulation follows a similar structure to the model used in Berry (1994), with the additional element that the marginal utility of income is also allowed to be a random variable in some of the simulations. We simulate $50,000$ individuals and we choose specific functional forms for the distributions of the following parameters: marginal utility of income (α_i), unobserved product attribute (ξ_j), observed product characteristic (x_j), product price (p_j), and the random coefficient (β_i). Each column reports results from a different set of assumptions about each of these distributions. In all simulations, initial variety is $J = 50$ and increases to $M = 55$. The first row in the table reports the exact change in consumer surplus, the next two rows report the approximate change in consumer surplus along with the bias from using the Logit formula but ignoring individual-specific heterogeneity and ignoring heterogeneity in the unobserved product attributes ξ_j, and the final rows report the approximate change in consumer surplus along with the bias from using the parallel demands approximation from

theorem 3.

In column (1), we simulate the model without any observed or unobserved individual-level and product-level heterogeneity other than allowing for heterogeneity through the shock ε_{ij} and product prices. In this case, we expect (and find) very little bias from assuming parallel inverse demands.[28] In fact, since without random coefficients the model collapses to the Logit model, we find the same population average equivalent variation. In columns (2) and (3), we add random coefficients (α_i and β_i). This introduces a small amount of bias in the Logit welfare formula and in the the parallel demand welfare formula, but the approximation is still fairly accurate.[29] We find the same pattern of results in columns (4) through (6) which add unobserved product-level heterogeneity (through the ξ_j terms). The last three columns explore the case where the distribution of the ξ_j terms is different for the $M - J$ products that are added, as compared to the distribution for the original J varieties. In the Logit formula ignoring this unobserved product-level heterogeneity, there is a significant upward bias in the population average equivalent. However, the parallel demands formula continues to be a fairly accurate approximation, even though the set of products added is not a random sample of the full set of M products. Overall, the simulation results suggest that even though the parallel demands assumption does not hold exactly in richer demand models with both observed and unobserved individual-level and product-level heterogeneity, the approximation may nevertheless be fairly accurate, especially as compared to counterfactual approaches that ignore selection of marginal varieties as exemplified by the Logit model used here.[30]

Example 7. For completeness, we now present a continuous choice example. Let

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

with H differentiable and concave. Then the first-order conditions imply:

$$\frac{p_i - a(J)}{p_j - a(J)} = \frac{a_i}{a_j}\left(\frac{x_i}{x_j}\right)^{\rho-1}$$

Therefore if the number of varieties is increased from J to M the average change in willingness-to-pay is approximated by $d = a(M) - a(J)$ as both J and M become large, and inverse

[28] The bias that is observed is numerical and arises since the formula is only exact for infinitesimal changes and the actual change in varieties in the simulations is from 50-55. We verified that if we substitute in the exact derivatives then we get the exact formula.

[29] This second source of bias, from ignoring the random coefficients, is always negative by Jensen's inequality $\mathbb{E}(Q_i * d_i) \geq \mathbb{E}(Q_i) * \mathbb{E}(d_i)$.

[30] The Logit model can perform poorly at generating counterfactuals (as in Debreu's classic "red-bus blue-bus example") but in general it is good at summarizing observed prices and quantities into a welfare measure. See the discussion in Nevo (2011) regarding Petrin's (2002) criticism of the Logit model.

demands are approximately parallel $Q_J(\mathbf{p^J} + s\mathbf{1_J}) \approx Q_M(\mathbf{p^M} + (s+d)\mathbf{1_M})$ for all s.

Example 8. Let preferences be given by

$$u_J(q_1, \ldots, q_J, m) = \hat{u}_J(q_1, \ldots, q_J) + H\left(\sum_{j=1}^{J} q_j\right) + m$$

where $\hat{u}_J$ is quasi-concave and homogeneous of degree 1 and H is concave. Then:

$$\begin{aligned}\max_{(q_j)} \hat{u}_J(q_1, \ldots, q_J) + H\left(\sum_{j=1}^{J} q_j\right) + y - \sum_{j=1}^{J} p_j q_j \\ = \max_{Q} \left[\max_{(\hat{q}_j)\in\Delta^J} \hat{u}_J\left(\hat{q}_1, \ldots, \hat{q}_J\right) - \sum_{j=1}^{J} p_j \hat{q}_j\right] Q + H\left(Q\right) + y\end{aligned}$$

where Δ^J is the simplex of dimension J. Let $Q(\mathbf{p_J}, J)$ be the aggregate demand as before and define the indirect subutility:

$$K(\mathbf{p_J}, J) = \max_{(\hat{q}_j)\in\Delta^J} \hat{u}_J\left(\hat{q}_1, \ldots, \hat{q}_J\right) - \sum_{j=1}^{J} p_j \hat{q}_j.$$

Notice that K has the property that $K(\mathbf{p_J} + s\mathbf{1_J}, J) = K(\mathbf{p_J}, J) - s$, so it has some similarity to the negative of a price index. Moreover, the aggregate demand satisfies:

$$Q(\mathbf{p_J}, J) = \arg\max_{Q} K(\mathbf{p_J}, J)Q + H(Q).$$

Finally, letting $d = K(\mathbf{p_M}, M) - K(\mathbf{p_J}, J)$, we can observe $Q(\mathbf{p_J} + (d+s)\mathbf{1_J}, J) = Q(\mathbf{p_M} + s\mathbf{1_M}, M)$ for any $s \in \mathbb{R}$, therefore inverse aggregate demands are parallel (the first assumption of theorem 3 is satisfied).

4 Estimation of the Variety Effect

To illustrate how one may implement our sufficient statistics formula for the variety effect, we next turn to an empirical application of sales taxes. In principle, any policy that exogenously shifts variety in the market – e.g., taxes, tariffs, price controls, regulations, etc. – is suitable to identify the variety effect for that policy. As we now show, all that is required are reduced-form effects of the policy on prices and output where variety is held fixed and where variety can directly respond to the policy. To fix ideas, consider a commodity tax, τ, in a market. As with sales taxes, we assume that the tax applies to some products in the market, but not others and therefore drives a wedge in relative prices. Additionally, taxes may affect the number of varieties in the market since they reduce demand for taxable products. With taxes,

the variety effect can be pinned down as follows:

$$\Lambda = -Q \left[\left. \frac{dp}{d\tau} \right|_J \frac{\frac{dQ}{d\tau}}{\left. \frac{dQ}{d\tau} \right|_J} - \frac{dp}{d\tau} \right] \frac{1}{\frac{dJ}{d\tau}}. \tag{17}$$

Clearly, to take formula (17) to the data, we need to estimate the effect of taxes when variety is held fixed and variety is variable. To operationalize this, we will assume that the fixed varieties case corresponds to the "short-run" and the variable varieties corresponds to the "long-run". We view this as a plausible assumption, since varieties are unlikely to respond to taxes in the short-run. Nevertheless, we can and do directly test this assumption by empirically examining whether varieties are endogenous with respect to taxes in the short-run.

The intuition for the variety effect is illustrated in Figure 4. Here we consider a small increase in taxes. As discussed above, the base of the rectangular area is given by pre-existing output before the tax change. The height of the rectangle is given by the difference between "long-run" change in price as a result of the tax change and the "short-run" change in price re-scaled by the ratio of the long-run output effect to the short-run output effect of the tax. The re-scaling serves to extend the price effect up the demand curve so that it's measured at the long-run output level. The identification of the variety effect thus comes from a policy instrument that shifts varieties in the market (such sales taxes), and is observed in a setting where variety is held constant and in a setting where variety can respond endogenously to the policy change. In the case of a standard CES demand model, both of the short-run and long-run effects are linked together by a single elasticity parameter. What our framework highlights is that in order to separately identify the demand elasticity (holding variety constant) from the variety effect, one requires two separate sources of variation. Conceptually, one needs an instrument to trace out demand holding variety constant and an instrument for variety. Practically, finding plausibly exogenous shocks to variety is likely to be challenging. Therefore, instead, we trace out the "long run" demand curve that allows both prices and variety to respond to cost shifter and show that this can be combined with the "short run" demand curve to identify the variety effect.[31]

[31] In the Appendix, we consider a more formal econometric framework for the identification of the variety effect in an Instrumental Variables (IV) framework.

4.1 The Principle of Le Chatelier and Parallel Inverse Demands

A concern with using short-run and long-run price elasticities of demand to identify the variety effect is that differences between the two may conflate changes in the level of variety with changes that might occur for other reasons, such as adjustment costs. In this section, we extend the model by incorporating an outside market represented by the variable y and we assume the consumer can only adjust y in the long run. We show the sufficient statistics formula is robust to the existence of this outside market by an application of the envelope theorem.

We start from a continuous choice model where all firms in the inside market are symmetric, we denote p the symmetric equilibrium price of the inside market, and Q the aggregate quantity. Let $u(Q, y, J) - pQ$ be the utility function of the consumer and assume u is supermodular and quasiconcave. Let

$$Q^*(y, p, J) = \underset{Q}{\operatorname{argmax}}\, u(Q, y, J) - pQ$$

be the aggregate demand of the inside good conditional on (p, y, J), and let

$$y^*(p, J) = \underset{y}{\operatorname{argmax}}\, u(Q^*(y, p, J), y, J) - pQ^*(y, p, J)$$

be the optimal choice of y given (p, J). Finally, define the long-run aggregate demand $Q(J) = Q^*(y^*(p(J), J), p(J), J)$.

Observe the long-run change in aggregate demand for the inside market given an exogenous change in variety J has three components:

$$\frac{dQ(J)}{dJ} = \frac{\partial Q^*}{\partial p}\frac{dp(J)}{dJ} + \frac{\partial Q^*}{\partial y}\frac{dy(p(J), J)}{dJ} + \frac{\partial Q^*}{\partial J} \tag{18}$$

the indirect effect of variety through equilibrium price p, the indirect effect of variety through the outside variable y, and the direct effect of variety J.

Assume the following parallel inverse demands condition:

Assumption. *(Parallel Inverse demands) For all J and all y there exists d such that for all p then $Q(y, p, J) = Q(y_0, p + d, J_0)$.*

In particular, for all J there exists $d(J)$ such that $Q(J) = Q(y_0, p(J) + d(J), J_0)$. Then

$$\begin{aligned} dQ &= Q(J_1) - Q(J_0) = Q(y_0, p_1 + d, J_0) - Q(y_0, p_0, J_0) \\ &\approx \frac{\partial Q^*}{\partial p} * (dp + d)\,. \end{aligned}$$

And so we can calculate the vertical shift

$$d \approx \left(\frac{dp}{dQ^*} - \frac{dp}{dQ} \right) * dQ. \tag{19}$$

Define the indirect utility function

$$w(y, p, J) = u(Q^*(y, p, J), y, J) - pQ^*(y, p, J)$$

and note that in a long-run equilibrium, welfare is $v(J) = w(y^*(p(J), J), p(J), J)$ from the consumer perspective. Taking the first-order conditions yields the following:

$$\frac{dv(J)}{dJ} = \frac{\partial w}{\partial p}\frac{dp}{dJ} + \frac{\partial w}{\partial J} + \frac{\partial w}{\partial y}\frac{dy^*}{dJ} = -Q\frac{dp}{dJ} + \Lambda$$

where $\frac{\partial w}{\partial y} = 0$ by the envelope theorem. Furthermore, the parallel inverse demands condition implies $-Q * d \approx \Lambda dJ$ and so

$$dv(J) \approx -Q * (dp + d)\,. \tag{20}$$

In other words, we can estimate the welfare effect in (20) by estimating pass-through (dp) and the vertical shift parameter (d) through equation (19). To estimate the latter, we need the short-run slope of demand (keeping both variety J and the outside market demand y fixed) and the long-run slope of demand when both y and J are adjusted. Therefore we have shown that our formula for the variety effect is robust to the existence of outside markets which only adjust in the long run. However, this is not true if there exists some externality from the outside market as we explain in the Appendix.

5 Empirical Results

The previous section showed that the sufficient statistics for the variety effect are $\frac{dJ}{d\tau}$, $\frac{dp}{d\tau}$, $\left.\frac{dp}{d\tau}\right|_J$, $\frac{dQ}{d\tau}$, and $\left.\frac{dQ}{d\tau}\right|_J$. To estimate these objects, we rely on several data sets. For measures of p, Q, and J, we use Nielsen's Retail Scanner Data. This is micro data which records weekly prices and output by product (Universal Product Code, UPC) for stores across the U.S. from 2006-2014. Each UPC in the data belongs to a "product module".[32] We aggregate the micro data to the store-module-time level to create two samples, one for the long-run analysis (time = yearly) and another for the short-run analysis (time = quarterly). Price is constructed as a time-varying module-store level average. For our empirical analysis below, we use aggregate expenditures per module and later obtain output elasticities from expenditure

[32]See Table A1 in the Appendix for examples of UPCs and the organizational hierarchy of the Nielsen data.

and price elasticities. Variety is defined as the count of UPCs with positive sales within a module and store (over the relevant time period).[33] To measure τ, we rely on hand-collected local sales tax rates and exemptions. These rates and exemptions vary by county, quarter and module.[34] Note that even though the finest level of sales tax variation is at the county level, we collapse the Nielsen data to the store level and limit our sample to grocery stores only since the distribution of store types, and therefore products, varies substantially across locations. Details on data construction and descriptive statistics for our sample (Table A3) are provided in the Appendix.

Grocery stores sell products that are subject to sales taxes (e.g., toothpaste) and other products that are exempt (e.g., food) generating within-store differences in after-tax prices between products. Additionally, local sales taxes vary substantially across the U.S., ranging from zero in Montana, Oregon, New Hampshire and Delaware to a maximum rate of 9.75 percent in Tennessee, as can be seen in Figure 5. This gives us two layers of identification, one that operates *across* stores and another that operates across products *within* stores. We combine these sources of variation to identify the long-run effects of taxation in a difference-in-differences (DD) research design.[35] Since we are exploiting *cross-sectional* variation which corresponds to the steady-state, we interpret the resulting estimates as "long-run" elasticities. In particular, our estimates will incorporate the response of product variety to sales taxes.[36]

To estimate the short-run effects of taxation, we rely on the panel structure of our data spanning the 2006-2014 period. We estimate regression models that control for module fixed effects interacted with store fixed effects, thus exploiting only within-store, time-series variation in sales tax rates. In practice, exemptions rarely change over our sample period so the identification is coming from changes in rates both at the state and county level. We interpret these regression results as corresponding to the "short-run" effects of taxation because firm entry and exit, and hence variety, is unlikely to adjust instantaneously to high-frequency variation in sales taxes, an assumption we empirically test below. In the Appendix, we formally describe our empirical models for the short-run and long-run effects of taxation.

[33]The definitions of price and variety are similar to the corresponding definitions in Handbury and Weinstein (2015).

[34]See Table A2 in the Appendix for examples of sales tax exemptions (Panel A is food modules and Panel B is non-food modules).

[35]Strictly speaking, it's not a pure DD design, since taxability status for a given module may vary across states and over time. In the cross-sectional models we estimate, we control for module and store fixed effects.

[36]Similarly, Atkin, Faber and Gonzales (2016) use cross-sectional variation in store-level prices to estimate long-run elasticities of substitution across stores.

5.1 Empirical Estimates of $\frac{dJ}{d\tau}$, $\frac{dp}{d\tau}$, and $\frac{d(pQ)}{d\tau}$

Our main estimates for the long-run effects of taxation are contained in Panel A of Table 2. We estimate cross-sectional regression models separately for each year between 2006 and 2014, and then take a simple linear combination of all the coefficient estimates. In all specifications, standard errors are clustered at the state-module level, the broadest level at which sales taxes are determined. The dependent variable is expenditure in columns (1) and (2), consumer price in columns (3) and (4), and variety in columns (5) and (6). Note that because we use consumer (after-tax) price, a coefficient of one in columns (3) and (4) implies full pass-through of the sales tax to consumers. Columns (1), (3) and (5), present the results for specifications which control for store and module fixed effects. Our estimates across these columns indicate that the tax elasticity of expenditures is -0.683 (s.e. 0.255), and that there is slight overshifting of taxes onto consumer prices with a coefficient of 1.145 (s.e. 0.036). We can reject the null of complete pass-through ($dp/dt = 1$). The elasticity of product variety with respect to sales taxes is -0.848 (s.e. 0.148). To address the concern that sales tax rates and exemptions are spatially correlated across regions of the U.S. in ways that may endogenously reflect the geographic distribution of consumer preferences, we consider a specification that allows module fixed effects to vary across census regions (columns (2), (4), and (6)). All of our results gain in precision with the inclusion of module-by-region fixed effects, and the point estimates are robust.[37]

Panel B of Table 2 presents the results of a border-counties analysis.[38] To account for spatial auto-correlation, standard errors are clustered two-way – by state-module as well as by border-segment-module in all specifications (Cameron et al. 2011). In columns (1), (3) and (5), the regression model is equivalent to the one in Panel A but the sample is restricted to border-counties. This sample restriction has little impact on the magnitude of the coefficient for consumer prices, but yields slightly smaller point estimates for both expenditures and variety. We then allow module fixed effects to vary by county pairs in columns (2), (4) and (6). The implied tax elasticity of expenditures -0.649 (s.e. 0.142) is in line with our baseline

[37] In Table A4, we report estimates from a simplified DD model by restricting the sample to counties where food products are fully exempt from the sales and to modules that are either taxed or exempt in all stores in this subset of counties. Reassuringly, our estimates are qualitatively similar to the estimates reported in Table 2.

[38] There is evidence that border counties set local sales tax rate strategically to compensate for cross-border difference in state-level sales tax rates (Agarwal 2015). In unreported regressions, we verified that results are not driven by this possible source of bias by instrumenting the statutory county-level tax rate with its state-level value.

results, as is the coefficient for prices 1.036 (s.e. 0.016). The effect of taxes on variety −0.304 (s.e. 0.070) is smaller in this specification, but remains statistically different from zero at the 1% level. Both the pooled cross-sectional results and "border county" results suggest large effect of taxes on variety, with an elasticity approximately between −0.3 and −0.8. This estimate can be compared to some the recent estimates of the elasticity of variety to population in Handbury and Weinstein (2014) and Schiff (2015) which range from 0.3 to 0.6. In both cases (change in taxes and change in population), there is a change in aggregate demand which in turn affects variety.[39]

Our estimates indicate that variety, measured as the number of products with positive sales, adjusts to changes in sales tax rates in the long-run. This approach implicitly puts an equal weight on all products that are marginally added or withdrawn from stores as a result of sales taxes. To explore which types of products are driving the effect on variety, we construct a new measure that puts relatively more weight on popular products – those with high national market shares – relative to low-share UPCs. We call this variable *variety shares* (G) and calculate it as follows. First, we obtain each UPC's market share by dividing total national sales of each product by the total national sales of the module it belongs to.[40] Then, for each store-module cell, variety shares are obtained by adding up the UPC-level market shares. For instance, a store that has positive sales of nearly all products that appear in the data set would have a variety share close to 1. Differences in G across stores therefore reflect both differences in the number of products (J) as well as in the market shares of products with positive sales. The empirical relationship between G and J is positive and concave, where stores with few different products tend to mainly sell high market-share products. Intuitively, if the variety effect operates mainly through low-share products, the tax elasticity of variety shares (G) should be smaller than the elasticity of variety counts (J). Estimates of these elasticities are shown in Table 3. In columns (1) and (2) we report our benchmark results for variety counts, and columns (3) and (4) show the corresponding results for variety shares. The estimates for the main specification with store and module fixed effects indicate a variety share elasticity of −0.462 (s.e. 0.149), about half the size of the variety count elasticity. The gap between the two coefficients shrinks when module effects are region-specific, but the pattern is qualitatively similar. In columns (5) through (8), we conduct a border-design analysis for

[39] In Appendix Table A5, we report a variety of robustness checks on our long-run estimates and show they are similar to our baseline estimates. These are described in more detail in the Appendix.

[40] In practice, we calculate these shares separately for each UPC-store cell leaving out the store's sales from national sales, so that the market shares are store-specific. However, since every store only makes up a very small portion of total national shares, leaving out own-store sales only trivially affect the resulting measures.

both measures of variety. In columns (6) and (8) module fixed effects are allowed to vary by county pairs. Again, the variety count elasticity of -0.304 (s.e. 0.070) exceeds the variety share elasticity of -0.209 (s.e. 0.062). Overall, these results suggest that the variety response is mainly driven by entry and exit of low-market share products.

5.2 Empirical Estimates of $\left.\frac{dp}{d\tau}\right|_J$ and $\left.\frac{d(pQ)}{d\tau}\right|_J$

We next report estimates of the short-run effects of taxation using our quarterly panel covering 2006-2014. The main results are reported in Panel A of Table 4. Standard errors are clustered at the state-module level to allow for correlation across stores located in a given state and to adjust for serial correlation. Columns (1) to (2) contain our estimates for expenditures and columns (3) to (4) contain our results for prices. In columns (1) and (3), we include module, store and time fixed effects, module effects interacted with store effects and module-specific linear time trends. By including module effects interacted with store effects, we effectively shut down the cross-sectional variation in taxes and rely exclusively on within-store, across module time variation. Our estimates indicate an expenditure elasticity of -0.342 (s.e. 0.111) and a price elasticity of 1.026 (s.e. 0.038), consistent with full pass-through. We cannot reject the null hypothesis that the coefficient for prices is equal to one at conventional levels of statistical significance. In columns (2) and (4) we control for module-specific trends more flexibly by including module effects interacted with quarter fixed effects instead of linear time trends. The expenditure elasticity remains negative but is smaller in magnitude -0.165 (s.e. 0.103), and the price elasticity is again statistically indistinguishable from one. Finally, as a placebo test we report results using variety as the outcome variable in columns (5) and (6). Reassuringly, the coefficient is statistically insignificant and close to 0, supporting our interpretation of the evidence as representing a short-run effect of taxation. To further assess the robustness of our short-run estimates, we also report results from a state border sample in Panel B. In columns (1), (3) and (5), our main specification is estimated on the restricted state border sample, and columns (2), (4) and (6) include pair-specific module trends. Reassuringly, these results closely mirror our main estimates, with coefficients for expenditures and prices slightly smaller in this sample. Overall, we find larger effects of taxes on variety and expenditures in the long-run, while pass-through estimates are similar in the long-run and in the short-run. The coefficient on price remains relatively close to one and is considerably smaller than

estimates reported in Besley and Rosen (1999).[41]

Testing for Symmetric Pass-through

Section 3 showed that the key assumption for the validity of the sufficient statistics formula for the variety effect in the general model is symmetric pass-through. We now empirically test for symmetric pass-through across products with different price levels. To classify UPCs as either high- or low-price products, we first use our panel dataset from 2006-2014 to estimate UPC fixed effects by regressing pre-tax prices on UPC and store fixed effects, separately for each module. For each module, we define the average value of the UPC fixed effects that is time-invariant. We denote UPCs with a time-invariant fixed effect above their module-specific mean as "high-price" products, and below-mean UPCs as "low-price" products. Next, we regress log prices on UPC fixed effects and store-by-quarter fixed effects interacted with the high/low price dummy variable. This delivers a quarterly panel data set of store-module price indices with two price indices for each module-store-quarter cell. Results are reported in Table 5. In columns (1) and (2) we calculate pass-through rates separately for high- and low-price UPCs controlling for module-specific linear time trends. We fail to reject the null of equal pass-through rates (p-value$>$ 0.112). Similarly, in columns (3) and (4), we control for module fixed effects interacted with quarter fixed effects. We test for equality of the two estimates and again cannot reject symmetric pass-through (p-value$>$ 0.235). In Table A7, we test for symmetric pass-through based on a comparison of products across brands with different market shares and generally fail to reject the null of equal pass-through rates.

5.3 Empirical Estimate of Λ

As a final step in the paper, we numerically implement our formula for the variety effect using our empirical estimates. First, we scale each term in equation (17) by J/pQ so that:

$$\frac{J}{pQ}\Lambda = -\left[\left.\frac{dlog(p)}{d\tau}\right|_J \frac{\frac{dlogQ}{d\tau}}{\left.\frac{dlogQ}{d\tau}\right|_J} - \frac{dlog(p)}{d\tau}\right]\frac{1}{\frac{dlog(J)}{d\tau}} \tag{21}$$

In order to implement this formula, we need estimates of $\frac{dlogQ}{d\tau}$ and $\left.\frac{dlogQ}{d\tau}\right|_J$. We get this by applying the chain rule and using our expenditure elasticity and price elasticity estimates, $\frac{dlogQ}{d\tau} = \frac{dlogpQ}{d\tau} - \frac{dlogp}{dt}$ and $\left.\frac{dlogQ}{d\tau}\right|_J = \left.\frac{dlogpQ}{d\tau}\right|_J - \left.\frac{dlogp}{d\tau}\right|_J$. Our calibrations are contained in Tables 6 and 7. First, consider Table 6. Columns (1) to (4) provide a set of illustrative calibrations for our welfare formulas. The values of the inputs in these columns do not correspond to

[41] Table A6 in the Appendix reports a variety of robustness checks on our short-run estimates and verifies they are similar to our estimates in Table 4.

our empirical estimates and are intended only as illustrations to provide intuition. Across all columns, the variety effect $\frac{J}{pQ}\Lambda$ is the main object of interest. In column (1), we see that the short-run and long-run effects of taxes on prices and output are the same and so the variety effect is 0. In this case, the reduced-form estimates reveal that consumers do not value variety. Since the shift in the aggregate demand curve depends jointly on the variety response to taxes and the variety effect, there is no downward shift in the aggregate demand curve, despite the fact that taxes nevertheless affect varieties ($dJ/d\tau < 0$). Turning to column (2), here we see that the long-run output effect exceeds the short-run output effect giving an estimate of the variety effect of 1. Column (3) uses the same price and output estimates as column (2) but considers a smaller (in magnitude) estimate for $dJ/d\tau$. Intuitively, the variety effect increases in order to rationalize the given shift in aggregate demand.

Using Table 6 as a template, we next implement the same formulas with our preferred set of reduced-form estimates from empirical analysis and report these results in Table 7. Our estimates imply that the variety effect ranges between 1.2 and 2.9, with most estimates around $1.2 - 1.4$, depending on whether we use our preferred pooled cross-sectional results or instead the "border pair" results. The scaling of the variety effect gives a unit-free measure that can be interpreted as the elasticity of the inverse aggregate demand curve with respect to variety. In other words, the magnitude of estimated variety effect suggests that an exogenous decrease in variety of 1 percent will decrease average willingness-to-pay by $1.2 - 1.4$ percent.

To give some intuition behind magnitude of the variety effect estimate, consider the following example: assume that an exogenous 1 percent decrease in variety results in all consumers who previously purchased the (removed) varieties deciding to purchase the outside good. Assuming identical market shares initially, then in this example a 1 percent decrease in variety will reduce quantity demanded by 1 percent (holding price constant). This is identical to an inward shift in the inverse aggregate demand curve by the reciprocal of the price elasticity of demand, or approximately 2 given the short-run estimates in Table 7. Alternatively, we can assume that the individuals who purchased the (removed) varieties instead purchase second choice, which they value 10 percent less on average. In this case, we would calculate an inward shift in inverse aggregate demand curve by 0.1 percent. The large variety effect is consistent with many consumers not choosing an alternative variety when their preferred variety is eliminated due to taxes. The final columns of Table 7 reproduces these calibrations instead imposing the Logit assumption. A common theme across the columns is the implied estimate of the variety effect under the Logit specification is significantly larger than most estimates

using our approach. This can be seen by comparing columns (1) through (4) with columns (5) and (6). The variety effect is significantly larger, which likely relates to the well-known result that the Logit model tends to overstate the welfare gains from new goods (Hausman 1996; Petrin 2002). The sensitivity of these results is examined in Appendix Table A8, in which we use coefficients estimates from several specifications that form the basis of our robustness checks. The magnitude of the variety effect varies between 0.8 and 2 in these cases.

6 Additional Applications

In this section, we consider several applications and highlight the usefulness of our approach to a broad range of topics in Industrial Organization and Public Economics.

6.1 Free entry and social inefficiency

Consider as in Mankiw and Whinston (1986) symmetric single product firms with cost function $c(q) + f$ where $c(q)$ is convex. Let profits be denoted by π and social welfare be the sum of consumer surplus and producer surplus, $W = CS + PS$. Then the marginal welfare gain of variety is given by:

$$\frac{dW}{dJ} = \Lambda + \pi + (p - c)J\frac{dq}{dJ}.$$

Where c is short for $c'(q)$. Under free entry $\pi = 0$, and under the parallel inverse demands assumption $\Lambda = d * Q$ therefore the marginal welfare gain can be written as

$$\frac{dW}{dJ} = \left(d * q + (p - c)\frac{dq}{dJ}\right) J.$$

Which means the following are sufficient statistics for $\frac{dW}{dJ}$:

1. The vertical shift parameter d which captures the average willingness to pay for the new variety.
2. Markup $p - c$.
3. The business stealing term $\frac{dq}{dJ}$.
4. Size of market: output per firm q and variety J.

Furthermore, the welfare formula and the sufficient statistics are the same regardless of whether firms are setting quantities or prices, and it holds in general for any conduct parameter (See the Appendix for details).

6.2 Welfare effects of Taxation

Consider a government that imposes an ad-valorem tax (τ) on each product in the market (but not the outside good). This generates revenue $R = \tau pQ$. We define welfare as $W = CS + PS + R$ where CS and PS are aggregate consumer and producer surplus, respectively. The marginal welfare gain is:

$$\frac{dW}{d\tau} = \frac{dCS}{d\tau} + \frac{dPS}{d\tau} + \frac{dR}{d\tau}.$$

Assume, as in the previous application that firms are symmetric with a convex cost function. Again, under free entry, profits are driven to 0. Therefore the marginal welfare gain can be expressed as:[42]

$$\frac{dW}{d\tau} = \Lambda \frac{dJ}{d\tau} - Q\frac{dp}{d\tau} + \frac{d(\tau pQ)}{d\tau} = Q\left(d - \frac{dp}{d\tau}\right) + \frac{d(\tau pQ)}{d\tau}$$

where $d = \left(\frac{dP}{dQ} - \frac{dP}{dQ}\Big|_J\right)\frac{dQ}{d\tau}$ and the second line follows by the parallel inverse demands assumption. From there we can check the following are sufficient statistics for $\frac{dW}{d\tau}$:

1. The vertical shift parameter d which captures the average change in willingness to pay.
2. Pass-through $\frac{dp}{d\tau}$.
3. The effect of taxes on output $\frac{dQ}{d\tau}$.
4. Output Q, price level p and tax rate τ.

In the Appendix, we implement this formula and show how this approach generalizes the common approach in public economics that does not take into account the effect of taxes on product variety.

7 Conclusion

Understanding how changes in product variety affect consumer welfare is critical for a host of questions. In this paper, we develop a sufficient statistics method for valuing changes in product variety and we consider an empirical application. Central to our application are reduced-form estimates of taxes on prices and quantities where variety is held constant and where variety responds to a change in taxes, along with estimates of taxes on varieties.

[42] In the particular case of competitive pricing and socially optimal variety J we obtain the Harberger (1964) triangle $\frac{dW}{d\tau} = \tau p \left.\frac{dQ}{d\tau}\right|_J$ (see also Chetty (2009)). When a planner sets variety J optimally conditional on pricing decisions being to firms we obtain $\frac{dW}{d\tau} = (p-c)\left.\frac{dQ}{d\tau}\right|_J$ (see Auerbach and Hines (2001)). Finally, the case of homogeneous products (where $\Lambda = 0$) is considered in Besley (1989) and Auerbach and Hines (2001). While these last two papers consider Cournot competition, our formula is valid for a broader class of models. See Appendix for details.

We implement our approach by combining rich retail scanner data from grocery stores in the U.S. with detailed state and county sales tax data. Our empirical results suggest a large effect of sales taxes on product variety, and we find a reasonably large variety effect which is an order of magnitude smaller than what would be obtained by instead assuming a Logit model. These results are of course specific to this setting, but we see this "proof of concept" as demonstrating applicability to other settings in IO, International Trade and Public Economics.

Our analysis can be extended and generalized in several dimensions. Most directly, one could apply the sufficient statistics formula for the variety effect developed here to study the impact of tariffs on welfare or other government policies that result in changes in equilibrium varieties, such as price controls. One could also use a similar approach to study the welfare effects of mergers, taking into account the variety effect in addition to the price effect. Finally, it would be useful to extend the methods developed here to consider optimal product variety allowing for rich firm heterogeneity and multi-product firms.

References

1. Agrawal, David R. (2015), "The Tax Gradient: Spatial Aspects of Fiscal Competition," *American Economic Journal: Economic Policy*, 7(2): 1-29.
2. Anderson, Simon P., Andre de Palma, and Yurii Nesterov (1995), "Oligopolistic Competition and the Optimal Provision of Products," *Econometrica*, 63(6): 1281-1301.
3. Anderson, Simon P., Andre De Palma, and Jacques-Francois Thisse (1987), "The CES is a discrete choice model?," *Economics Letters*, 24(2): 139-140.
4. Anderson, Simon P., Andre De Palma and Brent Kreider (2001a), "The Efficiency of Indirect Taxes under Imperfect Competition," *Journal of Public Economics*, 81(2): 231-251.
5. Anderson, Simon P., Andre De Palma and Brent Kreider (2001b), "Tax Incidence in Differentiated Product Oligopoly," *Journal of Public Economics*, 81(2): 173-192.
6. Arkolakis, Costas, Arnaud Costinot and Andres Rodriguez-Clare (2012), "New Trade Models, Same Old Gains?" *American Economic Review*, 102(1): 94-130.
7. Atkin, D., Benjamin Faber and Marco Gonzalez-Navarro, "Retail Globalization and Household Welfare: Evidence from Mexico," forthcoming *Journal of Political Economy.*
8. Auerbach, Alan J. and James R. Hines Jr. (2001), "Perfect Taxation with Imperfect Competition," NBER Working Paper #8138.
9. Behrens, Kristian, Yoshitsugu Kanemoto and Yasusada Murata (2017), "On measuring welfare changes when varieties are endogenous," CEPR Discussion Paper, DP11774.
10. Benassy, Jean-Pascal (1996), "Taste for Variety and Optimum Production Patterns in Monopolistic Competition," *Economics Letters*, 52, 41-47.

11. Berry, Steven (1994), "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics*, Vol. 25, No. 2, Summer, 1994, 242-262.

12. Berry, Steven, Alon Eizenberg, and Joel Waldfogel (2016), "Optimal Product Variety in Radio Markets," *RAND Journal of Economics* Vol. 47, No. 3, Fall 2016 pp. 463–497.

13. Berry, Steven, and Joel Waldfogel (1999), "Free entry and social inefficiency in radio broadcasting," *RAND Journal of Economics*, 30(3): 397-420.

14. Berry, Steven, James Levinsohn, and Ariel Pakes, 1995, "Automobile Prices in Market Equilibrium," *Econometrica,* 63, 841–890.

15. Besley, Timothy J. (1989), "Commodity taxation and imperfect competition: A note on the effects of entry,*" Journal of Public Economics*, 40(3): 359-367.

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

17. Bhattacharya, Debopam (2015), "Nonparametric Welfare Analysis for Discrete Choice*," Econometrica* 83(2): 617–649.

18. Broda, Christian and David E. Weinstein (2006), "Globalization and the Gains From Variety," *The Quarterly Journal of Economics*, 121(2): 541-585.

19. Bulow, Jeremy I., and Paul Pfleiderer (1983), "A Note on the Effect of Cost Changes on Prices." *Journal of Political Economy*, 91 (1): 182–85.

20. Busso, Matias, Jesse Gregory and Pat Kline (2013). "Assessing the Incidence and Efficiency of a Prominent Place Based Policy," *American Economic Review,* 103, pp. 897-947.

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

22. Card, David, Ana Rute Cardoso, Joerg Heining and Pat Kline (2017). "Firms and Labor Market Inequality: Evidence and Some Theory," *Journal of Labor Economics (forthcoming).*

23. Cardell, N. Scott (1997), "Variance Components Structures for the Extreme-Value and Logistic Distributions with Application to Models of Heterogeneity," *Econometric Theory*, 13(2): 185-213.

24. Chetty, Raj (2006), "A General Formula for the Optimal Level of Social Insurance*," Journal of Public Economics* 90 (November): 1879-1901.

25. Chetty, Raj (2008), "Moral Hazard versus Liquidity and Optimal Unemployment Insurance," *Journal of Political Economy*, 116(2): 173-234.

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

27. Chetty, Raj (2012), "Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply," *Econometrica*, 80(3): 969-1018.

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

29. Diamond, Rebecca (2016). "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000," *American Economic Review,* 106(3): 479-524.

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

31. 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.

32. Einav, Liran, Amy Finkelstein, and Mark R. Cullen (2010), "Estimating Welfare In Insurance Markets Using Variation in Prices," *The Quarterly Journal of Economics*, 125(3): 877-921.

33. Fabinger, Michal and Glen Weyl, (2016), "The Average-Marginal Relationship and Tractable Equilibrium Forms". Working Paper.

34. Feenstra, Robert C. (1994), "New Product Varieties and the Measurement of International Prices," *American Economic Review*, 84(1): 157-177.

35. Gabaix, Xavier, David Laibson, Deyuan Li, Hongyi Li, Sidney Resnick, and Casper G. de Vries. (2016), "The Impact of Competition on Prices with Numerous Firms." *Journal of Economic Theory,* Vol. 165, p. 1-24.

36. Gentzkow, Matthew, Jesse M. Shapiro and Michael Sinkinson (2014), "Competition and Ideological Diversity: Historical Evidence from US Newspapers," *American Economic Review*, 104(10): 3073-3114.

37. Gillitzer, Christian, Henrik J. Kleven and Joel Slemrod (2015), "A Characteristics Approach to Optimal Taxation: Line Drawing and Tax-driven Product Innovation," *Scandinavian Journal of Economics*.

38. Gorman, W. (1953), "Community Preference Fields", *Econometrica*, 21: 63-80.

39. Hagedorn, Marcus, Iourii Manovskii and Kurt Mitman (2016), "The Impact of Unemployment Benefit Extensions on Employment: The 2014 Employment Miracle?", Working paper.

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

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

42. Hart, Oliver (1982). "Reasonable Conjectures". Discussion Paper. London School of Economics.

43. Hausman, Jerry (1996), "Valuation of New Goods under Perfect and Imperfect Competition," in *The Economics of New Goods* (eds. Timothy F. Bresnahan and Robert J. Gordon).

44. Hausman, Jerry and Gregory K. Leonard (2002), "The Competitive Effects of a New Product Introduction: A Case Study," *Journal of Industrial Economics*, 50(3): 237-263.

45. Heckman, James and Edward Vytlacil (2007), "Econometric evaluation of social programs". In *Handbook of Econometrics*, Vol. 6, ed. J. Heckman, E. Leamer, pp. 4779–874. Amsterdam: North Holland.

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

47. Lamadon, Thibaut, Magne Mogstad, and Bradley J. Setzler (2017), "Rent Sharing, Earnings Dynamics and Transmission of Firm and Market Level Shocks." Working Paper.

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

49. Marschak, Jacob (1953), "Economic measurements for policy and prediction". In: Hood, W., Koopmans, T. (Eds.), Studies in Econometric Method. Wiley, New York, pp. 1–26.

50. McFadden, Daniel (1981), "Econometric models of probabilistic choice". In Manski, C. and McFadden, D., editors, Structural Analysis of Discrete Data with Econometric Applications. MIT Press.

51. Melitz, Marc and Stephen Redding (2015), "New Trade Models, New Welfare Implications," *American Economic Review*, 105(3): 1105-1146.

52. Milgrom, Paul and John Roberts (1996). "The LeChatelier Principle". *American Economic Review* 86(1): 173-179.

53. Myles, Gareth D. (1989), "Product Variety and Tax Policy," Working Paper.

54. Nevo, Aviv (2003), "New Products, Quality Changes and Welfare Measures Computed from Estimated Demand Systems," *The Review of Economics and Statistics*, 85(2), 266-275.

55. Nevo, Aviv (2011), "Empirical Models of Consumer Behavior," *Annual Review of Economics*, 3: 51-75.

56. Petrin, Amil (2002), "Quantifying the Benefits of New Products: The Case of the Minivan," *Journal of Political Economy*, 110(4): 705-729.

57. Resnick, Sidney I. (1987), "Extreme Values, Regular Variation and Point Processes". Springer, New York.

58. Riordan, Michael (1985), "Imperfect Information and Dynamic Conjectural Variations". *Rand Journal of Economics* 16: 41-50.

59. Samuelson, Paul A. (1947), "The Foundations of Economic Analysis", *Harvard University Press,* Cambridge, MA

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

61. Sheu, Gloria (2014), "Price, Quality, and Variety: Measuring the Gains from Trade in Differentiated Products," *American Economic Journal: Applied Economics*, 6(4): 66-89.

62. Spence, Michael (1976a), "Product Differentiation and Welfare," *American Economic Review*, 66(2): 407-14.

63. Spence, Michael (1976b), "Product Selection, Fixed Costs, and Monopolistic Competition," *Review* of Economic Studies, Vol. 43, No. 2, pp. 217-235.

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

65. Train, Kenneth (2003), "Discrete Choice Methods with Simulation," Cambridge University Press.

66. Trajtenberg, Manuel (1989), "The Welfare Analysis of Product Innovations, with an Application to Computed Tomography Scanners," *Journal of Political Economy*, 97: 444-479.

67. Vives, Xavier (2001), "Oligopoly Pricing: Old Ideas and New Tools". MIT Press.

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

69. Zidar, Owen and Juan Carlos Suárez Serrato (2016), "Who Benefits from State Corporate Tax Cuts? A Local Labor Market Approach with Heterogeneous Firms," *American Economic Review*, 106(9): 2582-2624.

Table 1: Random Coefficients Logit Model Simulation to Assess Bias from Assuming Parallel Inverse Aggregate Demand Curves

Distribution of product attributes, ξ_j	$\xi_j = 0$			$\xi_j \sim$ N(0,0.2)			$\xi_j \sim$ N(0, 0.2) for $j \leq J$ $\xi_j \sim$ N(-0.5, 0.2) for $j > J$		
Distribution of marginal utility of income, α_i	$\alpha_i = 1$	$\alpha_i \sim N(1,0.2)$	$\alpha_i \sim N(1,0.2)$	$\alpha_i = 1$	$\alpha_i \sim N(1,0.2)$	$\alpha_i \sim N(1,0.2)$	$\alpha_i = 1$	$\alpha_i \sim N(1,0.2)$	$\alpha_i \sim N(1,0.2)$
Random coefficients on product characteristics, β_i	$\beta_i = 0$	$\beta_i = 0$	$\beta_i \sim N(0,0.2)$	$\beta_i = 0$	$\beta_i = 0$	$\beta_i \sim N(0,0.2)$	$\beta_i = 0$	$\beta_i = 0$	$\beta_i \sim N(0,0.2)$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Change in Equivalent Variation (EV) from J to M varieties									
Exact change in EV	0.081	0.081	0.081	0.081	0.081	0.081	0.050	0.050	0.050
Approximate change in EV for the Logit model (ignoring random coefficients (α_i, β_i) and selection in product attributes, ξ_j)	0.081	0.081	0.081	0.081	0.081	0.081	0.081	0.080	0.080
Bias in formula relative to exact change, in %	0.0%	-0.4%	-0.5%	0.0%	-0.4%	-0.5%	61.4%	60.6%	60.6%
Approximate change in EV using the parallel demands revealed preference formula	0.079	0.077	0.077	0.079	0.077	0.077	0.049	0.048	0.048
Bias in formula relative to exact change, in %	-2.8%	-5.3%	-5.3%	-2.8%	-5.4%	-5.4%	-1.8%	-4.4%	-4.4%

Notes: This table reports simulation results following the discrete choice model in Example 6 described in the main text. In this model, there are observed and unobserved product characteristics, creating product heterogeneity in addition to price heterogeneity. In all simulations, initial variety is $J = 50$ and increases to $M = 55$. The price of each product p_j is distributed according to Uniform[0,1] for all products. The rows report the change in Consumer Surplus from this change in variety. In all simulations, there is a distribution of product characteristics distributed as $x_j \sim N(0,0.2)$ and this is multiplied by the random coefficient β_i. The unobserved random utility term is distributed according to standard Gumbel distribution, so that the model is a random coefficients logit. In columns (1) through (3), there is no unobserved product heterogeneity. In columns (4) through (6) there is unobserved product heterogeneity but it is distributed the same across all products. In columns (7) through (9), the unobserved product attributes of the new varieties added are drawn from different distribution, creating unobserved selection. The exact change is calculated numerically across the 50,000 simulated individuals. The approximate change for the Logit model uses average value of marginal utility of income (assumed to be known) and the actual prices (which are assumed to be observed). The change in EV is then given by formula in equation (15), ignoring individual-specific heterogeneity and ignoring the distribution of unobserved product attributes. The final rows implement the parallel demands formula by calculating numerical derivatives of dP/dQ.

Table 2 : The Long-Run Effect of Sales Taxes on Expenditure Shares, Average Prices, and Product Variety

Dependent Variable:	Expenditure shares		Prices		Variety	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full sample						
$\log(1+\tau_{mcs})$	-0.683	-0.716	1.145	1.060	-0.848	-0.813
	(0.255)	(0.198)	(0.036)	(0.025)	(0.148)	(0.123)
Specification:						
Store fixed effects	y	y	y	y	y	y
Module fixed effects	y		y		y	
Module × Region fixed effects		y		y		y
N (observations)	17,320,024	17,320,024	17,320,024	17,320,024	17,320,024	17,320,024
N (modules)	198	198	198	198	198	198
N (stores)	11,487	11,487	11,487	11,487	11,487	11,487
N (counties)	1,625	1,625	1,625	1,625	1,625	1,625
N (county-modules)	308,977	308,977	308,977	308,977	308,977	308,977
R^2	0.774	0.810	0.404	0.503	0.837	0.857
Panel B: Border counties						
$\log(1+\tau_{mcs})$	-0.539	-0.649	1.093	1.036	-0.564	-0.304
	(0.230)	(0.142)	(0.030)	(0.016)	(0.139)	(0.070)
Specification:						
Store fixed effects	y	y	y	y	y	y
Module fixed effects	y		y		y	
Module × Pair fixed effects		y		y		y
N (observations)	14,113,781	14,113,781	14,113,781	14,113,781	14,113,781	14,113,781
N (modules)	198	198	198	198	198	198
N (stores)	4,040	4,040	4,040	4,040	4,040	4,040
N (counties)	636	636	636	636	636	636
N (county-modules)	120,876	120,876	120,876	120,876	120,876	120,876
R^2	0.773	0.888	0.393	0.698	0.831	0.905

Notes: Sales tax rates efffective 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. In Panel A, standard errors are clustered at the state-module level. In Panel B, the sample is restricted to border counties and observations are weighted by the inverse of number of pairs a county belongs to. Standard errors are clustered two-way at the state-module level and at the border segment-module level.

Table 3 : The Long-Run Effect of Sales Taxes on Variety - Alternative Measures

Sample:	Full Sample				Border counties			
Dependent Variable:	Variety count		Variety share		Variety count		Variety share	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\log(1+\tau_{mcs})$	-0.848	-0.813	-0.462	-0.603	-0.564	-0.304	-0.288	-0.209
	(0.148)	(0.123)	(0.149)	(0.116)	(0.139)	(0.070)	(0.138)	(0.062)
Specification:								
Store fixed effects	y	y	y	y	y	y	y	y
Module fixed effects	y		y		y		y	
Module × Region fixed effects		y		y				
Module × Pair fixed effects						y		y
N (observations)	17,320,024	17,320,024	17,320,024	17,320,024	14,113,781	14,113,781	14,113,781	14,113,781
N (modules)	198	198	198	198	198	198	198	198
N (stores)	11,487	11,487	11,487	11,487	4,040	4,040	4,040	4,040
N (counties)	1,625	1,625	1,625	1,625	636	636	636	636
N (county-modules)	308,977	308,977	308,977	308,977	120,876	120,876	120,876	120,876

Notes: Sales tax rates efffective on September 1. Variety counts and shares are measured yearly. All coefficients are linear combinations of nine coefficients -- one for each year from 2006 to 2014. Columns (1) to (4): All standard errors are clustered at the state-module level. Columns (5) to (8): The sample is restricted to border counties. Observations are weighted by the inverse of number of pairs a county belongs to. All coefficients are linear combinations of nine coefficients -- one for each year from 2006 to 2014. All standard errors are clustered two-way at the state-module level and at the border segment-module level.

Table 4: The Short-Run Effect of Sales Taxes on Expenditure Shares, Average Prices, and Variety

Dependent Variable:	Expenditure shares		Prices		Variety	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full sample						
$\log(1+\tau_{mcst})$	-0.342	-0.165	1.026	1.008	-0.025	-0.082
	(0.111)	(0.103)	(0.038)	(0.025)	(0.079)	(0.071)
Specification:						
Store, Time, Module fixed effects	y	y	y	y	y	y
Module × Store fixed effects	y	y	y	y	y	y
Module-specific linear time trend	y		y		y	
Module × Time fixed effects		y		y		y
N (observations)	68,076,928	68,076,928	68,076,928	68,076,928	68,076,928	68,076,928
N (modules)	198	198	198	198	198	198
N (stores)	11,487	11,487	11,487	11,487	11,487	11,487
N (counties)	1,625	1,625	1,625	1,625	1,625	1,625
N (quarters)	36	36	36	36	36	36
N (county-modules)	308,977	308,977	308,977	308,977	308,977	308,977
R^2	0.998	0.971	0.676	0.746	0.962	0.965
Panel B: Border counties						
$\log(1+\tau_{mcst})$	-0.215	-0.131	0.986	0.960	0.006	0.058
	(0.116)	(0.103)	(0.036)	(0.038)	(0.078)	(0.069)
Specification:						
Store, Time, Module fixed effects	y	y	y	y	y	y
Module × Store fixed effects	y	y	y	y	y	y
Module-specific linear time trend	y		y		y	
Module × Pair-specific linear time trend		y		y		y
N (observations)	55,419,758	55,419,758	55,419,758	55,419,758	55,419,758	55,419,758
N (modules)	198	198	198	198	198	198
N (stores)	4,040	4,040	4,040	4,040	4,040	4,040
N (counties)	636	636	636	636	636	636
N (quarters)	36	36	36	36	36	36
N (county-modules)	120,876	120,876	120,876	120,876	120,876	120,876
R^2	0.951	0.953	0.645	0.669	0.954	0.957

Notes: A unit of time is a quarter, and the period covered is 2006-2014. In Panel A, standard errors are clustered at the state-module level. In Panel B, the sample is restricted to border counties, observations are weighted by the inverse of number of pairs a county belongs to, and standard errors are clustered two-way at the state-module level and at the border segment-module level.

Table 5: Symmetric Pass-Through

Sample of UPCs:	High-price	Low-price	High-price	Low-price
	(1)	(2)	(3)	(4)
Dependent Variable:				
Log of Average Consumer Price	1.051	0.993	1.034	0.994
	(0.040)	(0.042)	(0.028)	(0.030)
Symmetric pass-through (p-value)	0.112		0.235	
Specification:				
Store, Time, Module fixed effects	y		y	
Module × Store fixed effects	y		y	
Module-specific linear time trend	y			
Module × Time fixed effects			y	
N (observations)	134,556,758		134,556,758	
N (modules)	198		198	
N (stores)	11,487		11,487	
N (counties)	1,625		1,625	
N (quarters)	36		36	

Notes: A unit of time is a quarter, and the period covered is 2006-2014. All standard errors are clustered at the state-module level. All fixed effects are fully interacted with indicators for each UPC categories. The p-value reported is for a test of equality of coefficients.

Table 6: Illustrative Calibrations of the Variety Effect

	(1)	(2)	(3)	(4)
"Long-run" estimates (endogenous J, free entry)				
Pass-through rate, dlog(p)/dt	1.0	1.0	1.0	1.0
Quantity response, dlog(Q)/dt	-1.0	-1.6	-1.6	-1.3
Variety response, dlog(J)/dt	-0.6	-0.6	-0.3	-0.6
"Short-run" estimates (fixed J)				
Pass-through rate, dlog(p)/dt |J	1.0	1.0	1.0	1.0
Quantity response, dlog(Q)/dt |J	-1.0	-1.0	-1.0	-1.0
Welfare estimates				
Variety Effect Parameter (scaled by J/(pQ) = 1/q)	0.0	1.0	2.0	0.5

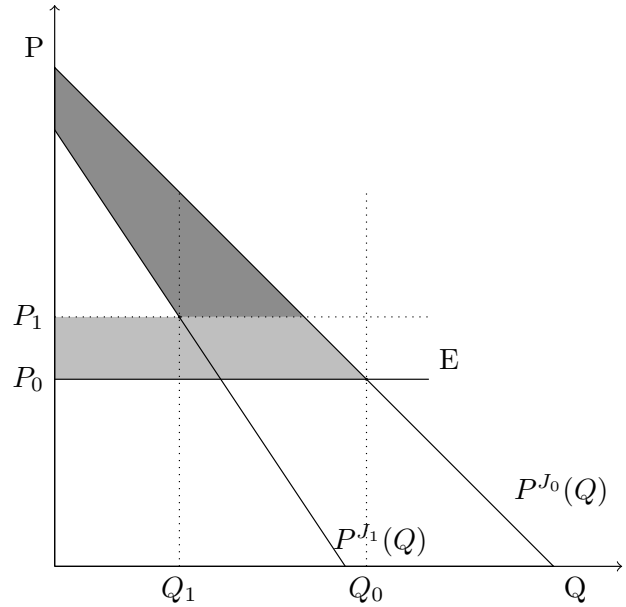
Notes: This table reports calibration of the variety effect formula in main text. In all columns, the results are based on assuming current tax rate is $0.063 per dollar. All columns report illustrative calibrations for different scenarios making different assumption about short-run and long-run effects of taxes. The first two columns illustrate the scenarios where either consumers do not value variety or average willingness-to-pay moves one-to-one with variety. The remaining columns show how the magnitude of the variety effect is determined by the reduced-form estimates of taxes (in short-run and long-run).

Table 7: Model-based estimates of the variety effect

	Baseline results				Results under logit/CES	
Long-run estimates	Pooled cross-section	Pooled cross-section	Border pair	Border pair	Pooled cross-section	Border pair
Short-run estimates	Time series, full sample	Time series, full sample	Time series, full sample	Time series, full sample	Time series, full sample	Time series, full sample
	(1)	(2)	(3)	(4)	(5)	(6)
"Long-run" estimates (endogenous J, free entry)						
Pass-through rate, dlog(p)/dt	1.15	1.06	1.09	1.04	1.15	1.09
Quantity response, dlog(Q)/dt	-0.83	-0.78	-0.63	-0.69	-0.83	-0.63
Variety response, dlog(J)/dt	-0.85	-0.81	-0.56	-0.30	-0.85	-0.56
"Short-run" estimates (fixed J)						
Pass-through rate, dlog(p)/dt |J	1.03	1.03	1.03	1.03	1.03	1.03
Quantity response, dlog(Q)/dt |J	-0.37	-0.37	-0.37	-0.37	-0.37	-0.37
Welfare estimates						
Variety Effect Parameter (scaled by J/(pQ) = 1/q)	1.374	1.360	1.187	2.870	2.788	2.788

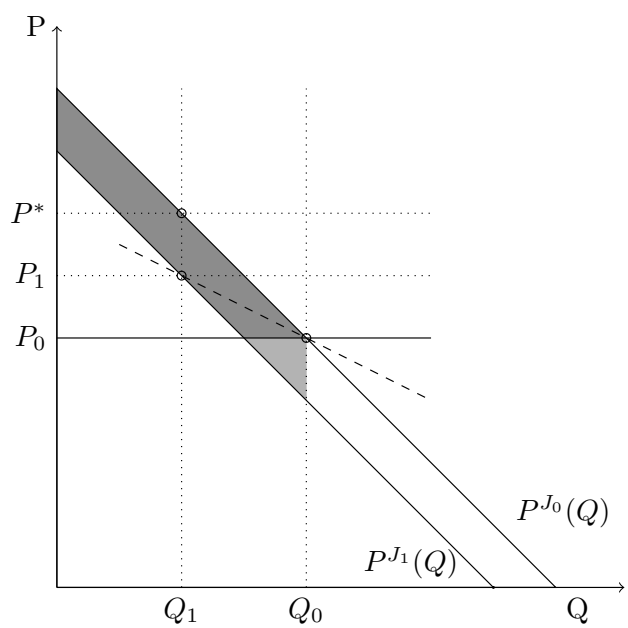
Notes: This table reports calibration of the marginal welfare gain formulas in main text. In all columns, the results are based on assuming current tax rate is $0.063 per dollar. All columns use actual short-run estimates reported in Table 3 and long-run estimates reported in Table 1. To do this, we first translate the reduced-form empirical estimates to the pass-through rate and quantity response using simple algebra. The key results in this table are the implied variety effects. The final two columns compare these results to analogous calibration results that assume logit/CES demand (as opposed to the more general "parallel demands" model described in main text).

Figure 1: Price Effect and Variety Effect



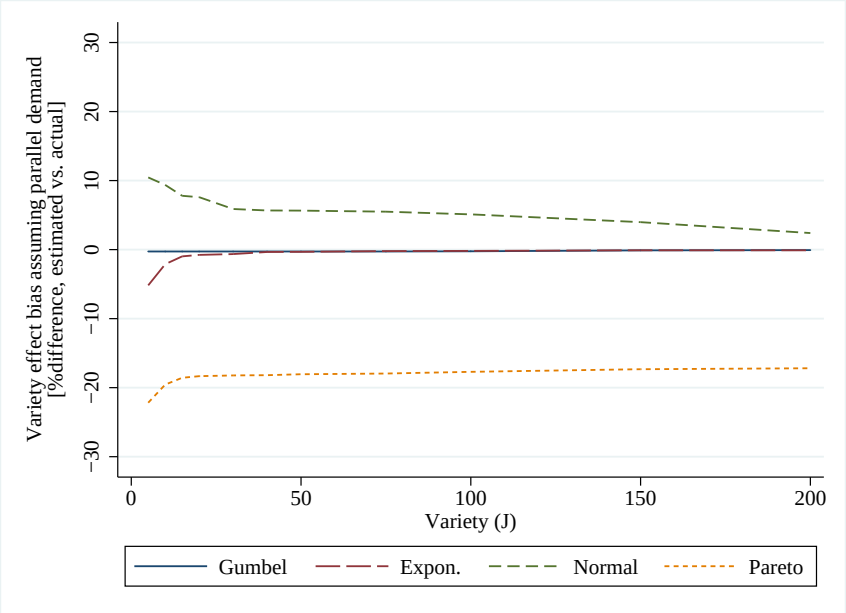
Notes: This figure shows graphically the result of a decrease in variety ($\Delta J < 0$). The price effect represents the rectangle formed by the base of the pre-existing output level and the height is the change in prices. The variety effect is the area between the two aggregate demand curves.

Figure 2: Variety Effect Under Parallel Demands



Notes: This figure shows graphically the variety effect in the case of parallel aggregate demand curves (i.e., ΔJ shifts aggregate demand curves in parallel). In this case, the variety effect is the area of the shaded parallelogram.

Figure 3: Approximate Parallel Demand Curves



Notes: This figure reports results from numerical simulations that are designed to evaluate the quality of the key approximation theorem (Theorem 2) in the main text. By simulating simple discrete choice models under different assumptions about distirbution of the iid error terms and increasing number of varieties in the market, we calculate (exact) variety effect numerically and compare it to variety effect we would infer from assuming parallel demands. Consistent with result of Theorem 2, for distributions that satisfy assumptions of theorem, as J increases, the bias in variety effect from assuming parallel demand approaches zero.

Figure 4: Long Run and Short Run Demand

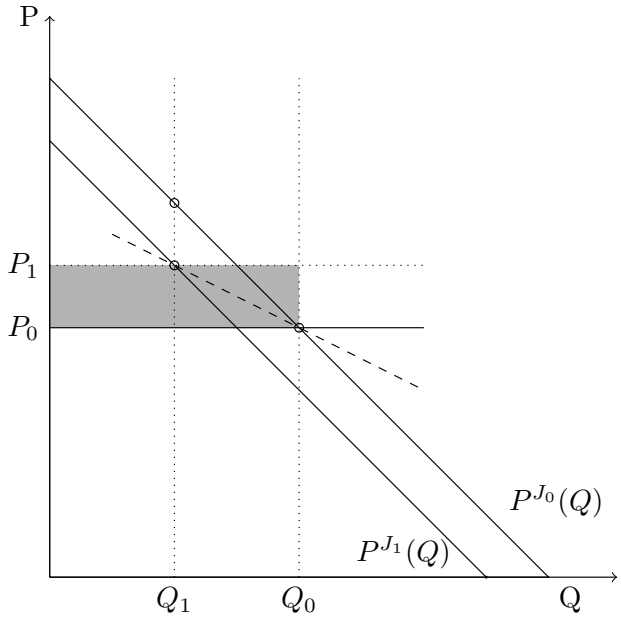
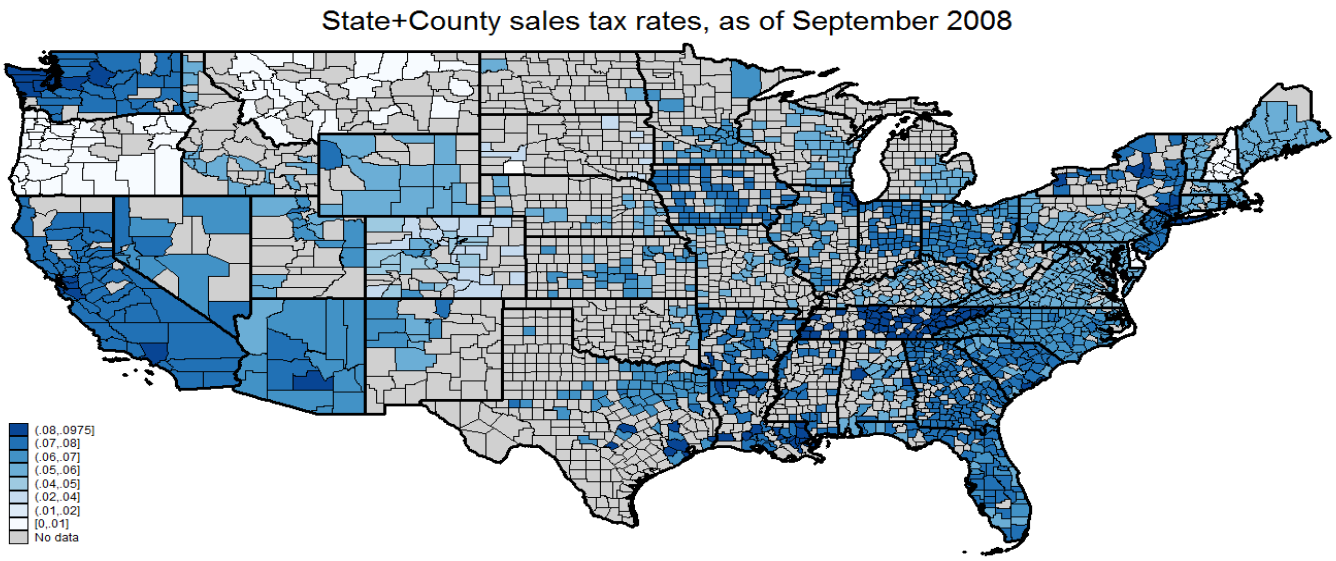


Figure 5: Map of Cross-Sectional Variation in Salex Tax Rates



Note: No data indicates counties for which no grocery store sales were recorded in Nielsen's Retail Scanner data in 2008.