Disclosure Statement

John List was Chief Economist at Walmart during the conduct of this research, remains in that role, and retains equity. Ian Muir was an employee of Walmart during the conduct of this research, remains so, and retains equity. Ariel Goldszmidt was an employee of Walmart during part of the research period, including at its outset, but is no longer employed by Walmart. Magne Mogstad and Lancelot Henry de Frahan received consulting fees from Walmart during a period overlapping with this research.

On the Identification and Interpretation of Price Elasticities in a Class of Discrete-Choice Demand Models*

Ariel Goldszmidt Lancelot Henry de Frahan John A. List
Magne Mogstad Ian Muir Alexander Torgovitsky

January, 2026

Abstract

We provide theoretical and econometric characterizations of price elasticities in commonly used discrete choice demand models for markets in which all (inside) products have small shares. Our theoretical characterizations reveal that differences in own-price elasticities across products are entirely determined by differences in their price levels, up to approximation error. It is well-known that the own-price elasticities in the standard logit model can only vary linearly in prices across products. We show that in the random coefficient logit model with an independent random coefficient on price, the own-price elasticities can only vary *quadratically* in prices across products. Allowing for correlation between random coefficients on product characteristics and price does not fundamentally change this result. Our econometric characterizations show how the linear and quadratic relationships between elasticities and prices are determined from specific moments of the data. We apply our characterization results to a large-scale randomized pricing experiment, in which we find that the approximation error is negligible and that own-price responses are grossly at odds with the elasticities from the standard and random coefficient logit model. In contrast, a logit model with product-specific price coefficients provides sufficient flexibility to rationalize the experimental estimates.

*Goldszmidt, Henry de Frahan, and Torgovitsky: Kenneth C. Griffin Department of Economics, University of Chicago. List: Kenneth C. Griffin Department of Economics, University of Chicago; NBER. Mogstad: Kenneth C. Griffin Department of Economics, University of Chicago, NBER; Statistics Norway. Muir: Walmart. We thank Aureo De Paula, Jean-Pierre Dube and Jesse Shapiro for comments.

1 Introduction

Discrete choice demand models are central to empirical work in industrial organization and related fields. A considerable body of methodological work has considered the (in)flexibility of substitution effects (Berry, 1994; Berry et al., 1995; McFadden and Train, 2000), the choice and validity of instruments (Reynaert and Verboven, 2014; Berry and Haile, 2014; Gandhi and Houde, 2019; Andrews et al., 2025), and computational challenges in estimation (Knittel and Metaxoglou, 2014; Conlon and Gortmaker, 2020; Salanié and Wolak, 2022). The focus of our paper is different: we consider the flexibility, interpretation, and identification of price elasticities in commonly-used specifications.

We study a family of utility specifications containing several models of interest:

$$v_{ij} = \beta_i' x_{jc} + \xi_{jc} + \alpha_{ij}(m_i - p_{jc}) + \varepsilon_{ij} \tag{1}$$

where v_{ij} is individual i's (indirect) utility from purchasing product j, p_{jc} is the price of j in market c, m_i is income, α_{ij} is a price sensitivity parameter, x_{jc} is a vector of observed product characteristics, ξ_{jc} is a scalar "unobserved" characteristic, and ε_{ij} is an idiosyncratic preference shock.[1] Throughout the paper, we assume that ε_{ij} are i.i.d. standard type-I extreme value.

We focus on three special cases of (1) that have been estimated in previous research: the logit model, which restricts α_{ij} to be a constant; the product-specific logit model, which allows α_{ij} to vary across products j but not across individuals i; and the random coefficient logit model, which allows α_{ij} to vary across individuals i but not across products j. In Section 2, we describe the economic assumptions that motivate these specifications.[2] We show that complementarity between inside and outside goods naturally leads α_{ij} to vary across products j, while heterogeneity in preferences leads α_{ij} to vary across individuals i.

For each of these models, Section 3 provides approximate characterizations of price elasticities for markets in which each product has a small market share compared to the total share of all the other products. These theoretical characterizations reveal that in the standard and random coefficient logit models, differences in own-price elasticities across products are entirely determined by differences in their price levels, up to approximation error. The own-price elasticities in the standard logit model are well-known to only vary linearly in prices across products. In the widely used logit model with independent random coefficients on prices, we show that the own-price elasticities can only vary *quadratically* in prices across products. Allowing for correlation between random coefficients on product characteristics and price does not fundamentally change this result: elasticities still follow a common quadratic form in prices, just shifted up and down by the product characteristics. In contrast, the product-specific model allows own-price elasticities to vary flexibly across products, though the elasticity of a given product still varies linearly in its own price.

[1] In (1), we assume that $m_i - p_{jc}$ enters the indirect utility linearly, as is common in the literature. In Appendix D, we show that the essential insights of our paper do not change if we instead assume the indirect utility is logarithmic in $m_i - p_{jc}$.

[2] Discrete choice demand models with product-specific price coefficients are the least common, but have been estimated in the marketing literature (e.g., Krishnamurthi and Raj (1988)), though without a formal micro-foundation. Allenby and Rossi (1991) develop a nonhomothetic logit model with product-specific parameters that allows for greater flexibility in price effects across products.

Our approximate characterization of own-price elasticities for the standard and product-specific logit models depends only on shares being sufficiently small. The characterization of own-price elasticities for the random coefficient logit model also makes a second-order approximation in α_i. We show that the approximation will be relatively accurate in three distinct cases: if the variance of α_i is relatively small; if the distribution of α_i is symmetric and thin-tailed, as in the empirically-common normal case; and if the product's price does not differ too much from the average price in the market.

While we focus on own-price elasticities, Section 3 also shows that most of the insights carry over to cross-price elasticities. In the standard logit model, cross-price elasticities vary linearly in the cross price across different pairs of products. Moving from the standard to the random coefficient logit model changes cross-price elasticities in two ways. First, a random coefficient on price makes cross-price elasticities approximately quadratic in the cross price; this result parallels our result for own-price elasticities. Second, random coefficients on non-price characteristics make cross-price elasticities larger or smaller according to how similar or different the two products are in characteristic space; this result has no parallel for own-price elasticities.

In Section 4, we show how the linear and quadratic relationships between price elasticities and prices are determined from the data. For concreteness and interpretability, we first consider a simple case in which the moment conditions used to identify the demand models are based on J randomized binary instruments, one for the price of each product. Each binary instrument gives rise to a product-specific linear IV estimand of log quantity for the jth product on the instrument-induced change in the price of the jth product.

We show three approximate results for own-price elasticities. The logit estimator constructs own-price elasticities by taking a weighted average of the IV estimands across all products. The product-specific logit model constructs elasticities using the IV estimand for the particular product. The random coefficient logit model constructs elasticities by taking a weighted regression of the IV estimands on a constant and average price levels. These results are useful because they highlight which features of the data approximately pin down the price elasticities of each product under each model. At the end of Section 4, we generalize this analysis to the case of an arbitrary vector of L instruments. The conclusions are entirely analogous, but with instrument-specific IV estimands replacing product-specific IV estimands.

We also offer analog approximation results for cross-price elasticities. We show that cross-price elasticities are constructed by projecting product-specific IV estimates of demand onto a constant (standard logit model), a product-specific constant (product-specific logit model), or a constant and the average price level (random coefficient logit model). Importantly, the IV estimates used in these projections reflect own-price responses to exogenous changes in prices, not substitution patterns or cross-price responses across products.

In Section 5, we illustrate the results in Sections 3 and 4 through simulations based on the (synthetic) cereal data of Nevo (2000). The results suggest that our approximations to price elasticities are accurate: the standard and random coefficient logit models produce elasticities that are close to linear and quadratic in prices and these linear and quadratic relationships are pinned down almost perfectly from the linear IV estimands. Simulations also suggest that our elasticity approximations are relatively accurate even when the distribution of α_i is skewed or thick-tailed.

In Section 6, we consider the implications of our approximation results for the widespread use of price elasticities to draw inferences about economic quantities such as marginal costs, markdowns, and pass-through rates. These quantities are of independent economic interest and are key inputs in counterfactual analyses such as merger simulation and tax policy. We consider a parsimonious model of the supply side: Each product is produced by a single-product firm with constant marginal cost that engages in Bertrand-Nash competition in prices. We show that when all products have small market shares, several key inferences about marginal costs, markdowns, and pass-through rates follow directly from the specification of the demand model, not from the data.

For example, if one uses the random coefficient logit model to estimate price elasticities, the inferred pass-through rate is necessarily larger than one, implying that costs, by assumption, are over-shifted to consumer prices. The exact magnitude of the pass-through rate is inferred by equating the curvature of demand of a given product with heterogeneity in the slope of demand across other products. Use of the random coefficient logit model also implies that the inferred markups (marginal costs) are necessarily (inverted) U-shaped in the observed prices across products. This means that, by assumption, marginal costs are low and markups are high both for firms that sell low- and high price products. The magnitudes of the marginal costs and markups for a given product are determined by the observed consumer responses to price changes of other products.

As each specification of (1) that we consider restricts heterogeneity in demand across price levels, markets, and/or products, it is natural to ask how the estimators can be interpreted under misspecification. In Section 7, we interpret the estimands from Section 4 under a model that imposes minimal structure on demand and nests the specifications considered in Sections 2–4. We show that the logit and product-specific logit estimators produce convex combinations of the true price elasticities even if the models are misspecified. These convex combinations can be used to bound quantities such as marginal costs and markups. By contrast, misspecification of the random coefficient logit estimator yields an affine combination of the true semi-elasticities with weights that may not all be positive, and, therefore, cannot be used to bound quantities such as marginal costs or markups.

In Section 8, we apply our theoretical and econometric results to a large-scale randomized pricing experiment. We work with a large US retailer operating thousands of brick-and-mortar establishments. Our instruments come from an exogenous pricing test run by the retailer's economics team for nine weeks in late 2022. The test included a total of 1,884 products across 23 categories. Since the discrete choice models we consider require that all products are substitutes with one another, we restrict focus to a single category from the field experiment: toothpaste. Importantly for our analysis, the experiment provides independent and exogenous variation in the price of each toothpaste product.

We use this variation to estimate each of the demand models. We find that the error in our approximations of the own-price elasticities is empirically negligible. The estimates further suggest that the own-price responses are grossly at odds with predictions from the standard and random coefficient logit models. We also find that the product-specific logit model provides sufficient flexibility in own-price responses across products to rationalize the experimental estimates.

In Section 9, we provide some concluding remarks and recommendations on how to specify discrete choice demand models in settings where all (inside) products have small shares.

In practice, practitioners will necessarily restrict the preferences in (1) for identification and estimation; the question is how to do so. We discuss alternative restrictions of preferences within and between consumers, illustrating how each affects the interpretation and identification of demand elasticities. The discussion is guided by data from the pricing experiment on consumer responses to multiple exogenous price changes on the same product.

Our paper relates to the literature on identification of models of differentiated product demand (Berry, 1994; Berry et al., 1995; Berry and Haile, 2014, 2024). We bridge a gap between existing theoretical identification analyses—which usually consider highly flexible nonparametric specifications—and empirical applications, which usually consider tightly specified parametric specifications due to the practical difficulties of nonparametric estimation (Compiani, 2022). Our results highlight that the typical assumptions researchers impose when bringing such models to data are not always innocuous. These assumptions severely restrict the own-price elasticities the model can produce and constrain the ability of estimators to rationalize the observed own-price responses in the data.

We are not the first to point out that some discrete choice models imply restrictive own-price elasticities. In their handbook chapter, Gandhi and Nevo (2021) note that with many products, the logit model implies own-price elasticities that are approximately linear in price. They describe this relationship as being driven by modeling assumptions, not data.[3] Gandhi and Nevo (2021) go on to argue that the own-price elasticities in the random coefficient logit model are not driven solely by functional form, but also by heterogeneity across individuals.[4] Our results show that in the random coefficient logit specifications researchers typically bring to data, own-price elasticities are *quadratic* in price levels. This relationship is driven by functional form and distributional assumptions, not data. This result confirms the conjecture of Björnerstedt and Verboven (2016) that a tight link between price elasticities and price levels may also be present in the random coefficient model.[5] Up to approximation error, the random coefficient logit model replaces the linear functional form relationship implied by the standard logit model with a quadratic functional form relationship.

Our paper also builds a bridge between the literature on discrete choice demand models and the literature on interpreting IV estimands (Imbens and Angrist, 1994; Angrist et al., 2000; Mogstad et al., 2021; Blandhol et al., 2025; Andrews et al., 2025). We show that standard estimators for discrete choice demand models can be approximately interpreted as linear IV regressions. The approximating linear IV regressions are not artificial, but instead intimately connected to the structure of own-price elasticities in these models. Moreover, the linear IV regressions lead to simple interpretations of the price sensitivity estimands as weighted aggregates of observed own-price responses in the data.

[3] "Since we have many products, generally, the market share of any given product is small, and therefore $\alpha_0(1 - s_{jt})$ is nearly constant. Therefore, the own-price elasticities are proportional to price: the lower the price, the lower the elasticity (in absolute value). ... The key is not whether this pattern is correct or not but that it is driven completely by modeling assumptions and not informed in any meaningful way by the data" (Gandhi and Nevo, 2021, pg. 78-79).

[4] "[O]wn price elasticities are not driven solely by functional form, but by the heterogeneity in the price sensitivity across consumers who purchase the various products" (Gandhi and Nevo, 2021, pg. 80).

[5] "The price elasticities of different products are quasi-linearly increasing in prices: if product A is twice as expensive as product B, it also tends to have a price elasticity that is twice as high. This property does not only hold in the logit and nested logit model; it may also be present to some extent in the random coefficient logit model" (Björnerstedt and Verboven, 2016, pg. 138).

2 Discrete choice models of differentiated product demand

2.1 A random utility model

Our starting point is the random utility model (RUM) as defined in Manski (1977). His definition of the RUM starts with the following probabilistic structure:

1. An individual $i \in \mathcal{I}$ is drawn from the population $\mathcal{I}$ of all consumers according to some distribution. Each individual has attributes a_i. In our application, these include the identity of the market c_i to which the individual belongs.

2. Conditional on the individual i, a choice set $\mathcal{J}_i \subseteq \{0, 1, \ldots, J\}$ is drawn according to some distribution over nonempty subsets of $\{0, 1, \ldots, J\}$. In our application, we assume that $\mathcal{J}_i = \{0, 1, \ldots, J\}$ with probability 1.

3. Individual i selects a single option $j_i^* \in \mathcal{J}_i$.

This structure generates various (conditional) choice probabilities; for example, the share of consumers in market c that choose option $k \in \{0, 1, \ldots, J\}$ is given by

$$s_{kc} \equiv \mathbb{P}\{j_i^* = k \mid c_i = c\}.$$

The model above is a RUM if it satisfies the *random utility assumption*: There exist variables v_{ij} for all $i \in \mathcal{I}$ and $j \in \{0, 1, \ldots, J\}$ such that choices j_i^* satisfy $v_{ij_i^*} \geq v_{ik}$ for all $k \in \mathcal{J}$. This means that each individual assigns a utility value to each alternative, and chooses an alternative in their choice set with the highest utility. The random utility assumption itself imposes few restrictions on how these utilities are formed.

Not all variables of the model are observed by the econometrician; in particular, individual choices j_i^* are not observed. Instead, we observe shares s_{jc}, prices p_{jc}, and (observable) product characteristics x_{jc} for each product j and market c. The object of interest is the share function σ, defined by

$$\sigma_j(p, x, \xi) \equiv \mathbb{P}\left\{v_{ij} \geq \max_k v_{ik} \;\middle|\; p_{c(i)} = p, x_{c(i)} = x, \xi_{c(i)} = \xi\right\} \tag{2}$$

and related to observed shares by $s_{jc} = \sigma_j(p_c, x_c, \xi_c)$. The random vector $\xi_c = (\xi_{1c}, \ldots, \xi_{Jc})'$ contains unobserved characteristics.

2.2 Specific cases of interest

Nonparametric estimation of a general RUM is challenging when the number of products J is large (Compiani, 2022). Parametric assumptions are almost always imposed in practice. For now, we focus on three specific members of the family of parametric utility specifications (1), each of which generates shares through (2). However, we consider several extensions in the paper.

Case 1: The standard logit model is obtained from (1) by restricting the price sensitivity parameter to be constant across products j and individuals i: $\alpha_{ij} = \alpha$. This assumption leads to shares

$$s_{jc} = \frac{\exp(\beta' x_{jc} - \alpha p_{jc} + \xi_{jc})}{1 + \sum_{k=1}^{J} \exp(\beta' x_{kc} - \alpha p_{kc} + \xi_{kc})} \tag{3}$$

and own-price elasticities

$$\eta_{jjc}(p) \equiv \frac{p_{jc}}{s_{jc}} \frac{\mathrm{d}s_{jc}}{\mathrm{d}p_j} = -\alpha p_j (1 - s_{jc}). \tag{4}$$

The cross-price elasticity of product k's share with respect to product j's price is

$$\eta_{kjc}(p) = \alpha p_j s_{jc}. \tag{5}$$

Case 2: The product-specific logit model relaxes the logit model by allowing the price sensitivity parameter in (1) to vary across *products* j but not individuals i: $\alpha_{ij} = \alpha_j$. This assumption leads to shares

$$s_{jc} = \int \frac{\exp(\beta' x_{jc} + \alpha_j (m_i - p_{jc}) + \xi_{jc})}{1 + \sum_{k=1}^{J} \exp(\beta' x_{kc} + \alpha_k (m_i - p_{kc}) + \xi_{kc})} \, \mathrm{d}F_{mc}(m_i) \tag{6}$$

and own-price elasticities

$$\eta_{jjc}(p) = \frac{-\alpha_j p_j \int (1 - s_{jc}(m_i)) s_{jc}(m_i) \, \mathrm{d}F_{mc}(m_i)}{\int s_{jc}(m_i) \, \mathrm{d}F_{mc}(m_i)} \tag{7}$$

where $s_{jc}(m_i) \equiv \exp(\beta' x_{jc} + \alpha_j (m_i - p_{jc}) + \xi_{jc}) / [1 + \sum_{k=1}^{J} \exp(\beta' x_{jc} + \alpha_k (m_i - p_{kc}) + \xi_{kc})]$. The cross-price elasticity of product k's share with respect to product j's price is

$$\eta_{kjc}(p) = \frac{\alpha_j p_j \int s_{kc}(m_i) s_{jc}(m_i) \, \mathrm{d}F_{mc}(m_i)}{\int s_{kc}(m_i) \, \mathrm{d}F_{mc}(m_i)}. \tag{8}$$

Case 3: The random coefficient logit model relaxes the logit model by allowing the price sensitivity parameter in (1) to vary across *individuals* i but not products j: $\alpha_{ij} = \alpha_i$. This assumption leads to shares

$$s_{jc} = \int \frac{\exp(\beta_i' x_{jc} - \alpha_i p_{jc} + \xi_{jc})}{1 + \sum_{k=1}^{J} \exp(\beta_i' x_{kc} - \alpha_i p_{kc} + \xi_{kc})} \, \mathrm{d}F_{\alpha,\beta}(\alpha_i, \beta_i) \tag{9}$$

and own-price elasticities

$$\eta_{jjc}(p) = \frac{\int -\alpha_i p_j (1 - s_{jc}(\alpha_i, \beta_i)) s_{jc}(\alpha_i, \beta_i) \, \mathrm{d}F_{\alpha,\beta}(\alpha_i, \beta_i)}{\int s_{jc}(\alpha_i, \beta_i) \, \mathrm{d}F_{\alpha,\beta}(\alpha_i, \beta_i)} \tag{10}$$

where $s_{jc}(\alpha_i, \beta_i) \equiv \exp(\beta_i' x_{jc} - \alpha_i p_{jc} + \xi_{jc}) / [1 + \sum_{k=1}^{J} \exp(\beta_i' x_{jc} - \alpha_i p_{kc} + \xi_{kc})]$. The cross-price elasticity of product k's share with respect to product j's price is

$$\eta_{kjc}(p) = \frac{\int \alpha_i p_j s_{kc}(\alpha_i, \beta_i) s_{jc}(\alpha_i, \beta_i) \, \mathrm{d}F_{\alpha,\beta}(\alpha_i, \beta_i)}{\int s_{kc}(\alpha_i, \beta_i) \, \mathrm{d}F_{\alpha,\beta}(\alpha_i, \beta_i)}. \tag{11}$$

There are several possible specifications of the random coefficient model. Some allow β_i to be correlated with α_i, while others assume β_i is independent or constant.

Other cases. While our paper focuses on these three cases because they impose economically distinct restrictions on consumer behavior, we also consider other specific cases of interest. One, discussed in Appendix E, is the nested logit model. Because the nested logit model can be written as a random coefficient logit model with an appropriate choice of x_j and F_β (Cardell, 1997), our results easily extend to this model. Another very flexible model, discussed in Section 9, is a product-specific random coefficient logit model. This model combines the flexibility of the product-specific and random coefficient logit models. In Appendix D, we also consider specifications in which indirect utility is logarithmic in income minus price rather than linear in price. These specifications allow for income eff8ects by making utility non-quasilinear in the outside good. All of our results easily extend to this case.

2.3 Implied restrictions on consumer preferences and behavior

What implicit restrictions does specification (1) impose on consumer preferences and behavior and what further assumptions are needed to produce the three special cases above?

To answer these questions, an explicit model of consumer behavior is necessary. Following Dubé (2019), suppose individual i in market c allocates income m_i over J indivisible "inside" products $q_1, \ldots, q_J$ and a divisible "outside" product q_0 to maximize utility U_i:

$$\max_{(q_1,\ldots,q_J)\in\{0,1\}^J, q_0\geq 0} U_i(q_0, q_1, \ldots, q_J) \quad \text{s.t.} \quad \sum_{j=0}^{J} p_{jc} q_j \leq m_i, \quad \sum_{j=1}^{J} q_j \leq 1. \tag{12}$$

The solution to (12) generates a RUM with utilities

$$\begin{aligned} v_{ij}(p_c, m_i) &\equiv U_i((m_i - p_{jc})/p_{0c}, 0, \ldots, 1, \ldots, 0) \\ v_{i0}(p_c, m_i) &\equiv U_i(m_i/p_{0c}, 0, \ldots, 0). \end{aligned}$$

The following proposition gives necessary and sufficient conditions for these utilities to take the form given in (1).

Proposition 1. *The utility maximization problem* (12) *generates the RUM* (1) *if and only if preferences can be written as*

$$U_i(q_0, q_1, \ldots, q_J) = \left(\sum_{j=1}^{J} q_j(\beta' x_{jc} + \xi_{jc} + \varepsilon_{ij} - \varepsilon_{i0})\right) + \left(\sum_{j=1}^{J} (\alpha_{ij} - \alpha_{i0}) q_j\right) q_0 + \alpha_{i0} q_0 \tag{13}$$

for any non-negative integer quantities such that $\sum_{j=1}^{J} q_j \leq 1$.

Proof. See Appendix A.1. □

The logit model can be derived from the utility maximization problem (12) by assuming (13) and imposing two further restrictions on U_i.

The first is that $\alpha_{ij} = \alpha_i$ does not depend on j, which is assumed in the standard logit model and in the random coefficient logit model, but not in the product-specific logit model. Inspecting (13), we see that this restriction implies that inside goods cannot differ in their complementarity to the outside good. Under this restriction, the RUM model generated by

(13) becomes a *linear* random utility model (LRUM), as defined by Anderson et al. (1992). In a LRUM, a common additive increase to all prices will not induce any substitution between inside goods, even though relative prices have changed.

The second restriction is that $\alpha_{ij} = \alpha_j$ does not depend on i, which is imposed in the standard logit model and the product-specific logit model, but not in the random coefficient logit model. This restriction implies that the marginal distribution of indirect utility for product j does not depend on j (up to a location shift). Under this restriction, the RUM model in (13) remains a *non-linear* random utility model. A common additive increase to all prices may induce substitution between inside goods, as relative prices have changed.

Relaxing separability to derive the product-specific logit model. The restriction that $\alpha_{ij} = \alpha_i$ implies that preferences (13) are *separable* between the inside and outside products. Relaxing separability allows the price coefficients to differ across products because the marginal utility of outside goods (and hence price sensitivity) depends on which inside good is chosen. This is natural when inside goods complement specific components of the outside bundle. For example, if the inside goods are cars and the outside good includes gasoline, less fuel-efficient cars are more complementary to this outside good.

In (13), where preferences are linear in q_0, product-specific price coefficients arise solely from such differences in complementarity. In Appendix B, we present a model with a utility function that is concave in multiple outside goods. In that case, product-specific coefficients may be generated either by differences in complementarity or by income effects, since variation in prices across inside goods alters the marginal utility of the outside goods. Although these income effects become negligible when prices are small relative to income, they could be important in certain markets, such as automobiles or housing.

Relaxing exchangeability to derive the random coefficient logit model. The restriction that $\alpha_{ij} = \alpha_j$ for all j implies that the stochastic components of the utility of each product are *exchangeable* across products j. Exchangeability may be unrealistic when tastes are more dispersed across individuals for some products than for others.

One way to relax exchangeability is to allow the price coefficients to vary across individuals i. In this case, price elasticities take the form (10) if $\alpha_{ij} = \alpha_i$ does not depend on j. The own-price elasticity of product j then reflects a weighted average of the price sensitivities of individuals who choose to purchase j. Lower-priced products will tend to be purchased by more price-sensitive consumers and will therefore have larger own-price semi-elasticities.

3 Price elasticity approximations

For each of the models in Section 2, we now provide an approximate characterization of price elasticities for markets in which all inside products have small shares. We first consider own-price elasticities before shifting attention to cross-price elasticities. For the random coefficient logit model, we first consider the instructive case in which β_i is constant, and return to the more general model in Section 3.2.2.

3.1 Basic results

The own-price elasticity expressions given in Section 2.2 can be difficult to interpret because they are complicated, nonlinear functions of the model primitives. In applications with many inside products (large J), however, shares are often small and it becomes possible to obtain interpretable approximations of the own-price elasticities.

The approximations for the logit and product-specific logit models are immediate; the approximation for the random coefficient logit model is more involved. Below, we explain when the approximation error is large or small.

Proposition 2. *(a) In the logit model,*

$$\begin{aligned}\eta_{jjc}(p_c) &= -\alpha p_{jc}(1 - s_{jc}) \\ &= -\alpha p_{jc} + O(s_{jc}).\end{aligned} \tag{14}$$

(b) In the product-specific logit model,

$$\begin{aligned}\eta_{jjc}(p_c) &= -\alpha_j p_{jc}\left(1 - \frac{\mathbb{E}[s_{jc}(m_i)^2]}{\mathbb{E}[s_{jc}(m_i)]}\right) \\ &= -\alpha_j p_{jc} + O\left(\frac{\mathbb{E}[s_{jc}(m_i)^2]}{\mathbb{E}[s_{jc}(m_i)]}\right).\end{aligned} \tag{15}$$

(c) In the random coefficient logit model,

$$\begin{aligned}\eta_{jjc}(p_c) &= -p_{jc}\frac{\mathbb{E}[\alpha_i s_{jc}(\alpha_i)(1 - s_{jc}(\alpha_i))]}{\mathbb{E}[s_{jc}(\alpha_i)]} \\ &= -\mathbb{E}[\alpha_i]p_{jc} + \mathrm{Var}[\alpha_i](p_{jc} - \bar{p}_c(\mathbb{E}[\alpha_i]))p_{jc} + O(\mathrm{Var}[\alpha_i]^2) \\ &\quad + O(\mathbb{E}[|\alpha_i - \mathbb{E}[\alpha_i]|^3]) + O\left(\frac{\mathbb{E}[\alpha_i s_{jc}(\alpha_i)^2]}{\mathbb{E}[s_{jc}(\alpha_i)]}\right).\end{aligned} \tag{16}$$

Proof. See Appendix A.2. □

Proposition 2 shows that under the logit model, each product's own-price elasticity is a linear function of its price. The coefficient on price, α, is common across all products. As such, any two products with the same price level must have the same own-price elasticity.

The product-specific logit model relaxes the elasticities of the logit model by allowing α_j to vary across products. While elasticities are still linear in price levels, the coefficient multiplying price differs by product. The reason is that some products may be more or less complementary to the outside good. For inside products that are more complementary to the outside good, a given price increase leads to a larger decrease in conditional indirect utility, because the marginal value of consuming the outside good is greater.

The random coefficient logit model relaxes the elasticities of the logit model in a different way. The expression in (16) shows that own-price elasticities in the random coefficient model can only vary *quadratically* in price.[6] The intuition for the quadratic term $\mathrm{Var}[\alpha_i](p_{jc} -$

[6]Prior work (e.g., Miravete et al. (2025)) has noted a relationship between the variance of price sensitivity and the curvature of the demand function, but the approximation in (16) appears to be novel. Moreover, our paper is the first to connect this curvature relationship to the identification of the random coefficient logit model (Section 4.5).

$\bar{p}_c(\mathbb{E}[\alpha_i]))p_{jc}$ is that when price sensitivity varies across individuals, products with above-average prices will be purchased more frequently by consumers with below-average price sensitivities (and vice versa). Because the price semi-elasticity of a product is a weighted average of the price sensitivities of the consumers purchasing that product, this selection of less-sensitive consumers into higher priced products will tend to make above-average-price products have smaller semi-elasticities and below-average-price products have larger semi-elasticities. This monotone relationship between prices and *semi*-elasticities leads to a quadratic relationship between elasticities and prices.

Approximations (14) and (15) rely only on shares s_{jc} being sufficiently small, and the errors in the approximations are linear in s_{jc}. The errors in approximation (16) are more complex. To gain more insight into the accuracy of the second-order (in α_i) approximation in (16), we consider higher-order approximations. Specifically, taking a fourth-order approximation adds terms

$$\begin{aligned}
&-\frac{1}{2}\mathbb{E}[(\alpha_i-\mathbb{E}[\alpha_i])^3]\left[(p_{jc}-\bar{p}_c(\mathbb{E}[\alpha_i]))^2-(\overline{p_c^2}(\mathbb{E}[\alpha_i])-\bar{p}_c(\alpha_i)^2)\right]p_{jc}\\
&-\frac{1}{6}\mathbb{E}[(\alpha_i-\mathbb{E}[\alpha_i])^4]\Bigg\{2(\bar{p}_c(\mathbb{E}[\alpha_i])-p_{jc})\left[(p_{jc}-\bar{p}_c(\mathbb{E}[\alpha_i]))^2-(\overline{p_c^2}(\mathbb{E}[\alpha_i])-\bar{p}_c(\mathbb{E}[\alpha_i])^2)\right] \quad (17)\\
&\qquad\qquad -\left[3\bar{p}_c(\mathbb{E}[\alpha_i])\overline{p_c^2}(\mathbb{E}[\alpha_i])-\overline{p_c^3}(\mathbb{E}[\alpha_i])-2\bar{p}_c(\mathbb{E}[\alpha_i])^3\right]\Bigg\}p_{jc}.
\end{aligned}$$

The first term will be zero if α_i has 0 skewness or if p_{jc} is one standard deviation above or below the average price $\bar{p}_c(\mathbb{E}[\alpha_i])$, while the second term will be small when α_i has small kurtosis. The fourth-order approximation therefore suggests the approximation in (16) is likely to be relatively accurate in three distinct cases: if the variance of α_i is relatively small; if the distribution of α_i is symmetric and thin-tailed, as in the empirically-common normal case; and for products with prices that do not differ too much from the average price.

The additional terms in (17) also suggest the general pattern that emerges as the order of the approximation increases. We summarize the general patten in the following proposition.

Proposition 3. *The Mth order approximation to the own-price elasticity in the random coefficient logit model is*

$$\eta_{jjc}(p_c)\approx\sum_{m=1}^{M}\frac{\mathbb{E}[(\alpha_i-\mathbb{E}[\alpha_i])^m]}{(m-1)!}L_{m-1}(p_{jc}-\bar{p}_c;\bar{p}_c(\mathbb{E}[\alpha_i]),\ldots,\overline{p_c^{m-1}}(\mathbb{E}[\alpha_i]))p_{jc}$$

where L_{m-1} is an $(m-1)$th order polynomial in $p_{jc}-\bar{p}_c$, the coefficients of which depend only on the first $m-1$ moments $\bar{p}_c(\mathbb{E}[\alpha_i]),\ldots,\overline{p_c^{m-1}}(\mathbb{E}[\alpha_i])$ of the distribution of prices of products purchased by an individual with average price sensitivity.

Proof. See Appendix A.4. □

The Mth order approximation to the own-price elasticity for product j is a sum of M polynomials in p_{jc}, the parameters of which depend on the distribution of prices in the market. The coefficient on the mth polynomial is the mth (centered) moments of the

distribution of α_i. Thus even as the order of the approximation in Proposition 2 increases, it remains the case that the random coefficient logit model produces the same relationship between price levels and own-price elasticities across all products, with the shape of the relationship determined by the distribution of α_i. Own-price elasticities may vary flexibly across price levels, but are equal across products at the same price level.

3.2 Extensions

3.2.1 Models with demographics

In a random coefficient logit model with demographics, the sensitivity α_i is assumed to be a linear function of a mean-0, variance-1 random draw ν_i and a demographic variable d_{ic}, such as income:

$$\alpha_i = \mu + \lambda\nu_i + \pi d_{ic}. \tag{18}$$

In this model, the mean and variance of α_i depend on the distribution of income in the market:

$$\begin{aligned}\mathbb{E}_{i\in c}[\alpha_i] &= \mu + \pi\,\mathbb{E}_{i\in c}[d_{ic}] \\ \mathrm{Var}_{i\in c}[\alpha_i] &= \lambda^2 + \pi^2\,\mathrm{Var}_{i\in c}[d_{ic}]\end{aligned}$$

where $\mathbb{E}_{i\in c}[d_{ic}]$ and $\mathrm{Var}_{i\in c}[d_{ic}]$ are the mean and variance of income across individuals in market c.

Apart from the assumed form of α_i, this model is equivalent to the random coefficient logit model without demographics for a given market c. The random coefficient logit market share expression (9) and elasticity expression (10) hold for this model, and we have the following analogue to Proposition 2:

Proposition 4. *In the random coefficient logit model with demographics,*

$$\begin{aligned}\eta_{jjc}(p_c) = &-\mathbb{E}_{i\in c}[\alpha_i]p_{jc} + \mathrm{Var}_{i\in c}[\alpha_i](p_{jc} - \bar{p}_c(\mathbb{E}_{i\in c}[\alpha_i]))p_{jc} + O(\mathrm{Var}_{i\in c}[\alpha_i]^2) \\ &+ O(\mathbb{E}_{i\in c}[|\alpha_i - \mathbb{E}_{i\in c}[\alpha_i]|^3]) + O\left(\frac{\mathbb{E}_{i\in c}[\alpha_i s_{jc}(\alpha_i)^2]}{\mathbb{E}_{i\in c}[s_{jc}(\alpha_i)]}\right)\end{aligned} \tag{19}$$

with $\mathbb{E}_{i\in c}[\alpha_i] = \mu + \pi\,\mathbb{E}_{i\in c}[d_{ic}]$ *and* $\mathrm{Var}_{i\in c}[\alpha_i] = \lambda^2 + \pi^2\,\mathrm{Var}_{i\in c}[d_{ic}]$.

Proposition 4 shows that for the random coefficient logit model with demographics, elasticities are approximately quadratic in price levels. In this model, however, the coefficients $-\mathbb{E}_{i\in c}[\alpha_i]$ and $\mathrm{Var}_{i\in c}[\alpha_i]$ of the quadratic vary across markets because they depend on the distribution of demographics. In markets with higher average values of the demographic (higher $\mathbb{E}_{i\in c}[d_{ic}]$), the slope of the quadratic will be more or less negative according to whether π is positive or negative. In markets with more dispersion in the demographic (higher $\mathrm{Var}_{i\in c}[d_{ic}]$), the curvature of the quadratic will be greater.

3.2.2 Multiple random coefficients

We next consider a model with multiple random coefficients:

$$v_{ij} = \beta_i' x_{jc} - \alpha_i p_{jc} + \xi_{jc} + \varepsilon_{ij}. \tag{20}$$

The same Taylor approximation used to derive Proposition 2 yields

Proposition 5.

$$\begin{aligned}\eta_{jjc}(p_c) = &- \mathbb{E}[\alpha_i] p_{jc} + \Big(\operatorname{Var}[\alpha_i](p_{jc} - \bar{p}_c(\mathbb{E}[\alpha_i], \mathbb{E}[\beta_i]))p_{jc} \\ &\qquad - \operatorname{Cov}[\alpha_i, \beta_i'](x_{jc} - \bar{x}_c(\mathbb{E}[\alpha_i], \mathbb{E}[\beta_i]))p_{jc} \Big) + O(\| \operatorname{Var}[(\alpha_i, \beta_i')] \|^2) \\ &+ O(\mathbb{E}[\|(\alpha_i, \beta_i') - \mathbb{E}[(\alpha_i, \beta_i')]\|^3]) + O\left(\frac{\mathbb{E}[\alpha_i s_{jc}(\alpha_i, \beta_i)^2]}{\mathbb{E}[s_{jc}(\alpha_i, \beta_i)]} \right)\end{aligned} \tag{21}$$

where $\bar{x}_c(\alpha_i, \beta_i) \equiv \sum_{k=1}^{J} s_{kc}(\alpha_i, \beta_i) x_{kc}$.

Proof. The proof is similar to that of Proposition 2 and is omitted. □

Proposition 5 shows that for the multiple random coefficient logit model, elasticities vary both with price levels p_j and product characteristics x_{jc}. But if $\operatorname{Cov}[\alpha_i, \beta_i'] = 0$—as is commonly assumed in applications; see (Gandhi and Nevo, 2021, pg. 84)—then the approximate elasticity (21) is the same as (16): elasticities are again approximately quadratic in price levels.

Moreover, even if α_i and β_i covary, Proposition 5 shows that conditional on product characteristics, elasticities are quadratic in price levels. The covariance between α_i and β_i allows the coefficient on the linear term in price to differ across products. However, the coefficient $\operatorname{Var}[\alpha_i]$ on the quadratic term remains common across products; the presence of covariance between α_i and β_i allows the level and slope of the quadratic in price to differ across products, but its curvature remains the same across products.

If $\operatorname{Cov}[\alpha_i, \beta_{ir}] > 0$, so that more price-sensitive consumers also assign more value to the rth product characteristic, then the quadratic for products with a high value of the rth characteristic will be more negatively-sloped. This pattern again reflects selection: products with high values of the rth characteristic will tend to be purchased by individuals with a taste for this characteristic, and the own-price elasticities of these products reflect the sensitivities of individuals who purchase them.

3.2.3 Flexible random coefficients

Inspection of Proposition 5 also suggests that the richer the space of product characteristics x_{jc}, the more own-price elasticities can vary across products. The richest possible space is obtained when x_{jc} consists of a full set of product indicators each with random coefficients, leading to the model

$$v_{ij} = \beta_{ij} - \alpha_i p_{jc} + \xi_{jc} + \varepsilon_{ij}. \tag{22}$$

The model in (22) is likely to be infeasible to use in practice because of the large number of parameters to estimate. Nonetheless, it is useful to consider as a limiting case of the flexibility one could achieve with a random-coefficient logit model. Proposition 6 characterizes the own-price elasticities in this infeasibly-flexible random coefficient model.

Proposition 6.

$$\eta_{jjc}(p_c) = -\mathbb{E}[\alpha_i]p_{jc} + \left(\operatorname{Var}[\alpha_i](p_{jc} - \bar{p}_c(\mathbb{E}[\alpha_i], \mathbb{E}[\beta_i]))p_{jc} - \operatorname{Cov}[\alpha_i, \beta_{ij}]p_{jc}\right) + O(\|\operatorname{Var}[(\alpha_i, \beta_i')]\|^2)$$
$$+ O(\mathbb{E}[\|(\alpha_i, \beta_i') - \mathbb{E}[(\alpha_i, \beta_i')]\|^3]) + O\left(\max_k s_{kc}(\mathbb{E}[\alpha_i], \mathbb{E}[\beta_i])\right) + O\left(\frac{\mathbb{E}[\alpha_i s_{jc}(\alpha_i, \beta_i)^2]}{\mathbb{E}[s_{jc}(\alpha_i, \beta_i)]}\right).$$

Proposition 6 shows that in the infeasibly-flexible random coefficient logit model, the own-price elasticity of product j is still approximately quadratic in the own price, and the coefficient on the quadratic term is still $\operatorname{Var}[\alpha_i]$ for all products. However, the coefficient on the linear term, $-(\mathbb{E}[\alpha_i] + \operatorname{Cov}[\alpha_i, \beta_{ij}])$, now varies across products j with $\operatorname{Cov}[\alpha_i, \beta_{ij}]$. Intuitively, when $\operatorname{Cov}[\alpha_i, \beta_{ij}] > 0$, individuals with a taste for the jth product are more price sensitive, shifting down the own-price elasticity for this product at every price level.

It is instructive to compare the system of own-price elasticities in Proposition 6 to those generated by the product-specific logit model. Like the product-specific logit model, the flexible random coefficient logit model allows the level of semi-elasticities to vary arbitrarily across products. However, the two models differ in their implications for how semi-elasticities vary across price levels for a given product. In the product-specific logit model, semi-elasticities are constant across price levels, and hence elasticities are linearly increasing in price. In the flexible random coefficient logit model, semi-elasticities are linearly increasing in price, and hence elasticities are quadratic in price. Semi-elasticities increase with price at the same rate for all products under the flexible random coefficient logit model.

3.3 Cross-price elasticities

The same approximations used above to derive closed-form expressions for own-price elasticities can also be applied to cross-price elasticities. For the logit and product-specific logit models, we easily obtain the exact expressions $\eta_{kjc}(p_c) = \alpha p_{jc} s_{jc}$ and $\eta_{kjc}(p_c) = \alpha_j p_{jc} \frac{\mathbb{E}[s_{kc}(m_i)s_{jc}(m_i)]}{\mathbb{E}[s_{kc}(m_i)]}$. For the random coefficient logit model, we have the following approximation.

Proposition 7. *In the random coefficient logit model, when all random coefficients are*

pairwise uncorrelated,

$$\begin{aligned}\eta_{kjc}(p_c) = \mathbb{E}[\alpha_i] p_{jc} \mathbb{E}[s_{jc}(\alpha_i, \beta_i)] \Bigg[& 1 \\ & + \operatorname{Var}[\alpha_i](p_{jc} - \bar{p}_c(\mathbb{E}[\alpha_i], \mathbb{E}[\beta_i]))(p_{kc} - \bar{p}_c(\mathbb{E}[\alpha_i], \mathbb{E}[\beta_i])) \\ & + \sum_{r=1}^{R} \operatorname{Var}[\beta_{ir}](x_{jrc} - \bar{x}_{rc}(\mathbb{E}[\alpha_i], \mathbb{E}[\beta_i]))(x_{krc} - \bar{x}_{rc}(\mathbb{E}[\alpha_i], \mathbb{E}[\beta_i])) \Bigg] \\ & + \left(\operatorname{Var}[\alpha_i](2\bar{p}_c(\mathbb{E}[\alpha_i], \mathbb{E}[\beta_i]) - p_{kc} - p_{jc}) \right) p_{jc} \mathbb{E}[s_{jc}(\alpha_i, \beta_i)] \\ & + O(\| \operatorname{Var}[(\alpha_i, \beta_i')] \|^2) + O(\mathbb{E}[\|(\alpha_i, \beta_i') - \mathbb{E}[(\alpha_i, \beta_i')]\|^3]). \end{aligned} \tag{23}$$

Proof. See Appendix A.5. □

Proposition 7 shows that adding random coefficients to the logit model makes the cross-price elasticities more flexible in two distinct ways. First, allowing α to be random makes the cross-price elasticity quadratic rather than linear in the cross price p_j; this flexibility exactly parallels the flexibility that a random α adds to the own-price elasticities. Second, allowing β to be random allows the level of the cross-price elasticity between products k and j to be greater or lesser according to how similar or different k and j are in their product characteristics. This flexibility has no parallel in our expressions for own-price elasticities.

To parse the expression in (23), start with the second term,

$$\operatorname{Var}[\alpha_i](2\bar{p}_c(\mathbb{E}[\alpha_i], \mathbb{E}[\beta_i]) - p_{kc} - p_{jc}) p_{jc} \mathbb{E}[s_{jc}(\alpha_i, \beta_i)].$$

This term has the same form as the quadratic term $\operatorname{Var}[\alpha_i](p_{jc} - \bar{p}_c)p_{jc}$ in the own-price elasticity approximation (16), with two differences. First, $p_{jc} - \bar{p}_c$ is replaced by $(\bar{p}_c - p_{jc}) + (\bar{p}_c - p_{kc})$; this reflects that the cross-price elasticity between products j and k depends both on which consumers select into j and which consumers select into k. Second, the entire term is multiplied by the share of product j; this reflects that the total amount of substitution into k when the price of j changes is proportional to the number of individuals consuming j before the price change.

Next, consider the first term in (23),

$$\begin{aligned}\mathbb{E}[\alpha_i] p_{jc} \mathbb{E}[s_{jc}(\alpha_i, \beta_i)] \Bigg[& 1 + \operatorname{Var}[\alpha_i](p_{jc} - \bar{p}_c(\mathbb{E}[\alpha_i], \mathbb{E}[\beta_i]))(p_{kc} - \bar{p}_c(\mathbb{E}[\alpha_i], \mathbb{E}[\beta_i])) \\ & + \sum_{r=1}^{R} \operatorname{Var}[\beta_{ir}](x_{jrc} - \bar{x}_{rc}(\mathbb{E}[\alpha_i], \mathbb{E}[\beta_i]))(x_{krc} - \bar{x}_{rc}(\mathbb{E}[\alpha_i], \mathbb{E}[\beta_i])) \Bigg]. \end{aligned}$$

The first part of the term, $\mathbb{E}[\alpha_i] p_{jc} s_{jc}(\mathbb{E}[\alpha_i], \mathbb{E}[\beta_i])$, has the same form as the linear term $-\mathbb{E}[\alpha_i] p_{jc}$ in the own-price elasticity approximation (16), with two differences. First, it is multiplied by the share of product j, again since the total amount of substitution when the price of product j changes is proportional to the mass of individuals consuming j before the price change. Second, the term is also multiplied by a long expression in brackets. This

bracketed expression is a measure of the correlation between the characteristics of products j and k. The term $(x_{jrc} - \bar{x}_{rc}(\mathbb{E}[\alpha_i], \mathbb{E}[\beta_i]))(x_{krc} - \bar{x}_{rc}(\mathbb{E}[\alpha_i], \mathbb{E}[\beta_i]))$ will be positive and large when products k and j either both have above-average or both have below-average levels of the rth characteristic. Each of these terms enters a sum weighted by $\text{Var}[\beta_{ir}]$, the variance of taste for the rth characteristic. Overall, then, the bracketed expression makes the cross-price elasticity between k and j more positive the more k and j share characteristics and the more dispersed are the tastes for the characteristics they share.

The bracketed term reflects the effects of selection that arise when individual tastes are heterogeneous. Suppose taste for characteristic r is highly dispersed in the population, and that products j and k both have above-average levels of characteristic r. Individuals who choose to consume product j will on average have a strong taste for characteristic r. When the price of j increases, these consumers will tend to substitute relatively more to product k than to other products, since k has a high level of the characteristic r they enjoy.

4 Identification of price sensitivity parameters

The exact own-price elasticities in Section 2 can be difficult to interpret for two reasons. The first challenge is that they are complicated, nonlinear functions of the model primitives. The approximation results in Section 3 address this issue by re-expressing the elasticities in terms of the price levels and the price sensitivity parameters. A second challenge remains: the standard estimators of these parameters cannot be expressed as closed-form functions of the data. This makes it hard to determine which moments of the data get used to pin down the price sensitivity parameters. The goal of this section is to address this second challenge by providing formulas that link the price sensitivity parameters to moments of the data. The results are approximate because they apply to an estimator that approximates the price sensitivity parameters, rather than the standard estimators. For concreteness and interpretability, we consider first the simple case where the moments conditions used to identify the demand models are based on J randomized binary instruments, one for the price of each product.

4.1 Data and moment conditions

The data consist of observations of shares s_{jc} (based on quantities q_{jc}), prices p_{jc}, product characteristics x_{jc}, and instruments z_{jc} for each product j and market c. To simplify the presentation, we start by assuming the characteristics x_{jc} consist only of a constant; we consider richer product characteristics in an extension below.

Identification of the models discussed in Section 2 with such data presents two challenges. The first is that the share s_{jc} is a function of both prices p_{jc} and unobserved product characteristics ξ_{jc}, and these variables are likely correlated. The reason is that, all else equal, higher-quality products will have higher prices. Identification requires instruments z_{jc} that are correlated with prices but uncorrelated with ξ_{jc}. The second challenge is that s_{jc} depends not only on ξ_{jc}, but on the entire vector of unobservables $\xi_c = (\xi_{1c}, \ldots, \xi_{Jc})$. We therefore need to isolate (the average of) the indirect utility terms before we can form moment conditions. This process is often referred to as "inverting" the market shares (Berry,

1994; Berry et al., 2013).

To invert the shares and specify a moment condition, it is useful to first regroup terms in the indirect utility v_{ij}:

$$v_{ij} = \delta_{jc}(\theta_1) + \mu_{ij}(\theta_2) + \varepsilon_{ij}.$$

$\delta_{jc}(\theta_1)$ denotes the mean (non-individual-specific) component of v_{ij}, ε_{ij} is the iid T1EV shock, and $\mu_{ij}(\theta_2)$ denotes the remaining stochastic component of v_{ij} (if any). For example, in the logit model, $\delta_{jc}(\theta_1) = \beta_0 - \alpha p_{jc} + \xi_{jc}$ and $\mu_{ij}(\theta_2) = 0$; in the random coefficient logit model with normally-distributed α_i, $\delta_{jc}(\theta_1) = \beta_0 - \mathbb{E}[\alpha_i]p_{jc} + \xi_{jc}$ and $\mu_{ij}(\theta_2) = \sqrt{\text{Var}[\alpha_i]}\nu_i p_{jc}$ where ν_i is standard normal.

The mean and stochastic components depend on unknown "linear" parameters $\theta_1 \in \Theta_1$ and unknown "nonlinear" parameters $\theta_2 \in \Theta_2$. (In the logit model, $\theta_1 = (\beta_0, \alpha)$ and θ_2 is empty; in the random coefficient logit model with normally-distributed α_i, $\theta_1 = (\beta_0, \mathbb{E}[\alpha_i])$ and $\theta_2 = \text{Var}[\alpha_i]$.) Define the function $\sigma(\cdot;\cdot) : \mathbb{R}^J \times \Theta_2 \to \mathbb{R}^J$ by

$$\sigma_j(\delta_c(\theta_1); \theta_2) = \int \frac{\exp(\delta_{jc}(\theta_1) + \mu_{ij}(\theta_2))}{1 + \sum_{k=1}^J \exp(\delta_{kc}(\theta_1) + \mu_{ik}(\theta_2))} \, \mathrm{d}F_{\mu(\theta_2)}(\mu_i(\theta_2))$$

and note that at the true values of θ_1, θ_2, we have $\sigma_j(\delta_c(\theta_1); \theta_2) = s_{jc}$. The function σ is invertible in its first argument in many cases of interest (Berry, 1994; Berry et al., 2013), though the inverse must usually be calculated numerically. Using σ^{-1} and instruments z_{jc}, we construct a moment condition

$$\mathbb{E}\left[J^{-1}\sum_{j=1}^{J}\begin{pmatrix}1\\ z_{jc}\end{pmatrix}(\sigma_j^{-1}(s_c;\theta_2) - \delta_{jc}(\theta_1))\right] = 0 \tag{24}$$

to identify the parameters θ_1 and θ_2. Estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ are constructed by solving (or, in the overidentified case, minimizing the norm of) the sample analogue of (24).

4.2 Definition of instruments and IV estimands

For concreteness and interpretability, we focus first on the case of a J-dimensional vector of randomized price instruments, deferring results for general instruments to Section 4.7. Our J-dimensional instrument is constructed to reflect the exogenous variation induced by the pricing experiment we study in Section 8. For each product, the experiment randomly varies prices up and down across markets. We represent this variation as a randomized binary variable $\dot{z}_{jc}$ that is one if the price was changed (increased or decreased) and zero otherwise. Because price changes were conditionally randomized across stores for each product, we interact $\dot{z}_{jc}$ with a vector of product indicators and residualize the interactions on the product indicators. The final instrument consists of J centered binary instruments, one for the price of each product. Each instrument averages to 0 within each product, has the same variance within each product, and is orthogonal to each other instrument.[7]

[7]Most of our analysis does *not* involve explicit instruments for shares. As we show below, many models we consider can be identified with only price instruments. We discuss share instruments further in Section 4.5.

To simplify notation in the results to follow, we define ρ_j^y as the estimand from an IV regression of an outcome variable y_{jc} (e.g., $y_{jc} = \log(q_{jc}/q_{0c})$) on p_{jc} with instrument $\dot{z}_{jc}$, using only observations for the jth product:

$$\rho_j^y \equiv \frac{\text{Cov}[\dot{z}_{jc}, y_{jc}]}{\text{Cov}[\dot{z}_{jc}, p_{jc}]} = \frac{\mathbb{E}[y_{jc} \mid \dot{z}_{jc} = 1] - \mathbb{E}[y_{jc} \mid \dot{z}_{jc} = 0]}{\mathbb{E}[p_{jc} \mid \dot{z}_{jc} = 1] - \mathbb{E}[p_{jc} \mid \dot{z}_{jc} = 0]}.$$

We also define ω_j to be a weight proportional to the squared first-stage relationship between $\dot{z}_{jc}$ and p_{jc} for the jth product and normalized so that $\sum_{j=1}^J \omega_j = 1$:

$$\omega_j \equiv \frac{\text{Cov}[\dot{z}_{jc}, p_{jc}]^2}{\sum_{k=1}^J \text{Cov}[\dot{z}_{kc}, p_{kc}]^2} = \frac{\big(\mathbb{E}[p_{jc} \mid \dot{z}_{jc} = 1] - \mathbb{E}[p_{jc} \mid \dot{z}_{jc} = 0]\big)^2}{\sum_{k=1}^J \big(\mathbb{E}[p_{kc} \mid \dot{z}_{kc} = 1] - \mathbb{E}[p_{kc} \mid \dot{z}_{kc} = 0]\big)^2}.$$

4.3 Logit

We first consider the standard logit model. In this model, the share inversion σ^{-1} has a closed form, $\sigma_j^{-1}(s_c) = \log(s_{jc}/s_{0c})$, and the moment condition (24) becomes

$$\mathbb{E}\left[J^{-1}\sum_{j=1}^J \begin{pmatrix} 1 \\ z_{jc} \end{pmatrix} (\log(s_{jc}/s_{0c}) - \beta_0 + \alpha p_{jc})\right] = 0.$$

The solution to this moment problem is the vector of coefficients from a (pooled) 2SLS regression of $\log(s_{jc}/s_{0c})$ on an intercept and $-p_{jc}$, with instruments z_{jc}. The corresponding estimator is

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\alpha} \end{pmatrix} = \left[C^{-1}J^{-1}\sum_{c=1}^C\sum_{j=1}^J \begin{pmatrix} 1 \\ -\hat{p}_{jc} \end{pmatrix} (1, -\hat{p}_{jc})\right]^{-1} \left[C^{-1}J^{-1}\sum_{c=1}^C\sum_{j=1}^J \begin{pmatrix} 1 \\ -\hat{p}_{jc} \end{pmatrix} \log(s_{jc}/s_{0c})\right] \quad (25)$$

where $\hat{p}_{jc}$ are the fitted values from a regression of p_{jc} onto $1, z_{jc}$.

The $-\alpha$ estimand can be expressed as a positively-weighted average of product-specific IV estimands $\rho_j^{\log(q/q_0)} = \frac{\mathbb{E}[\log(q_{jc}/q_{0c})|\dot{z}_{jc}=1] - \mathbb{E}[\log(q_{jc}/q_{0c})|\dot{z}_{jc}=0]}{\mathbb{E}[p_{jc}|\dot{z}_{jc}=1] - \mathbb{E}[p_{jc}|\dot{z}_{jc}=0]}$:

Proposition 8. *Let $\hat{\alpha}$ be the estimator in (25) with instruments z_{jc} as defined in Section 4.2. Then*

$$\operatorname*{plim}_{C\to\infty}(-\hat{\alpha}) = \sum_{j=1}^J \omega_j \rho_j^{\log(q/q_0)}. \quad (26)$$

Proof. Follows from Proposition 13 proven below. □

Proposition 8 and the small-share approximation (14) immediately imply:

Corollary 1. *Let $\hat{\eta}_{j'j'c'}(p_{c'})$ denote the own-price elasticity of product j' in market c' from the logit model estimated using (25) with instruments z_{jc} as defined in Section 4.2. Then*

$$\operatorname*{plim}_{C\to\infty} \hat{\eta}_{j'j'c'}(p_{c'}) = \left(\sum_{j=1}^J \omega_j \rho_j^{\log(q/q_0)}\right) p_{j'c'} + O(s_{j'c'}). \quad (27)$$

Corollary 1 shows that the logit own-price semi-elasticity is approximately a weighted average of product-specific IV estimands of log quantity (relative to the outside option) on price. Proposition 2 showed that logit own-price elasticities can only vary across products linearly in their price levels, up to approximation error. Corollary 1 shows that the linear relationship is determined from the data by a weighted average of the IV estimands $\rho_j^{\log(q/q_0)}$.

The intuition behind this result is connected to our small-share approximation of the logit model's own-price elasticities. If each product has an own-price elasticity that is linear in its own price, with a constant coefficient $-\alpha$, then the own-price semi-elasticity of demand is $-\alpha$ for *all* products at *all* price levels. Each product-specific IV estimand is an estimate of that product's semi-elasticity of demand and hence of $-\alpha$. The logit estimator simply takes a weighted average of all these product-specific estimators.

The weight on the jth estimate, ω_j, is proportional to the square of the change in average price induced by the instrument. These weights arise for reasons of statistical efficiency. More weight is given to those products for which the instrument varies price more, as the semi-elasticities for such products will be more precisely estimated (at least under homoskedasticity).

4.4 Product-specific logit

We next consider the product-specific logit model. For simplicity, we focus on the case where income m_i is constant across individuals, deferring the variable-income case to the Appendix. The share inversion has a closed form, $\sigma^{-1}(s_c) = \log(s_{jc}/s_{0c})$, and the moment condition (24) becomes

$$\mathbb{E}\left[J^{-1}\sum_{j=1}^{J}\begin{pmatrix} I_{jc} \\ z_{jc}\end{pmatrix}\left(\log(s_{jc}/s_{0c}) - \beta_0 - (m\alpha)'I_{jc} + \alpha' I_{jc}p_{jc}\right)\right]$$

where I_{jc} is a J-dimensional vector of product indicators. The corresponding estimator is

$$\begin{pmatrix}\widehat{\beta_0 + m\alpha} \\ \hat{\alpha}\end{pmatrix} = \left[C^{-1}J^{-1}\sum_{c=1}^{J}\sum_{j=1}^{J}\begin{pmatrix} I_{jc} \\ \widehat{Ip}_{jc}\end{pmatrix}(I_{jc}', \widehat{Ip}_{jc}')\right]^{-1}\left[C^{-1}J^{-1}\sum_{c=1}^{J}\sum_{j=1}^{J}\begin{pmatrix} I_{jc} \\ \widehat{Ip}_{jc}\end{pmatrix}\log(s_{jc}/s_{0c})\right] \tag{28}$$

where $\widehat{Ip}_{jc}$ are the fitted values from a projection of $I_{jc}p_{jc}$ onto I_{jc} and z_{jc}.

The $-\alpha_j$ estimand can be expressed in terms of the IV estimand for the jth product $\rho_j^{\log(q/q_0)} = \frac{\mathbb{E}[\log(q_{jc}/q_{0c})|\dot{z}_{jc}=1] - \mathbb{E}[\log(q_{jc}/q_{0c})|\dot{z}_{jc}=0]}{\mathbb{E}[p_{jc}|\dot{z}_{jc}=1] - \mathbb{E}[p_{jc}|\dot{z}_{jc}=0]}$:

Proposition 9. *Let $(\hat{\alpha}_1, \ldots, \hat{\alpha}_J)'$ be the estimators in (28) with z_{jc} as defined in Section 4.2. Then*

$$\operatorname*{plim}_{C\to\infty}(-\hat{\alpha}_j) = \rho_j^{\log(q/q_0)}.$$

Proof. Follows immediately from the observation that (28) is an IV regression with fully interacted product fixed effects. □

Proposition 9 and the small-share approximation (15) immediately imply:

Corollary 2. *Let $\hat{\eta}_{j'j'c'}(p_{c'})$ denote the own-price elasticity of product j' in market c' from the product-specific logit model estimated using (28) with instruments z_{jc} as defined in Section 4.2. Then*

$$\operatorname*{plim}_{C\to\infty} \hat{\eta}_{jjc}(p) = \left[\rho_j^{\log(q/q_0)}\right] p_{jc} + O(s_{jc}). \tag{29}$$

This corollary shows that the product-specific logit elasticity for the jth product is a linear function of the product's price, with the linear relationship determined by the IV estimand $\rho_j^{\log(q/q_0)}$. Unlike the logit model, the product-specific logit model allows the coefficient α_j to vary across products. In contrast, the logit model restricts all α_j to be equal, and estimates the common α by a weighted average of IV estimands across all products.

The above results are derived under the assumption that income m_i is constant across all consumers.[8] We show in Appendix C that the essential insights from the results above generalize to the variable-income case.

4.5 Random coefficient logit

4.5.1 A simple case of the random coefficient logit model

We now consider the random coefficient logit model (9). We first focus on the simple case where the distribution of price sensitivities F_α is known up to a location parameter μ and a scale parameter λ:

$$F_\alpha(\alpha_i) = \Phi\left(\frac{\alpha_i - \mu}{\lambda}\right)$$

where Φ is a known CDF with $\int \nu \,\mathrm{d}\Phi(\nu) = 0$ and $\int \nu^2 \,\mathrm{d}\Phi(\nu) = 1$.[9] For simplicity, we also assume the vector of characteristics x_{jc} consists only of a constant.

The share inversion σ^{-1}, where

$$\sigma_j(\delta_c; \lambda) \equiv \int \frac{\exp(\delta_{jc} + \lambda\nu_i p_{jc})}{1 + \sum_{k=1}^J \exp(\delta_{kc} + \lambda\nu_i p_{kc})} \,\mathrm{d}\Phi(\nu_i), \tag{30}$$

does not have a closed form, but can be computed numerically. The moment condition (24) becomes

$$\mathbb{E}\left[J^{-1} \sum_{j=1}^J \begin{pmatrix} 1 \\ z_{jc} \end{pmatrix} \left(\sigma_j^{-1}(s_c; \lambda) - \beta_0 + \mu p_{jc}\right)\right] = 0. \tag{31}$$

[8] Even if total income is highly variable across individuals, it may be the case that all individuals allocate roughly the same budget to a particular product category like toothpaste.

[9] The most common example of such a Φ is the standard normal CDF, but we do not assume normality for any of the results below.

The Berry et al. (1995) (BLP) estimator of the random coefficient logit model is the vector $(\hat{\beta}_0, \hat{\mu}, \hat{\lambda})'$ that solves the sample analogue of (31):

$$C^{-1}J^{-1}\sum_{c=1}^{C}\sum_{j=1}^{J}\begin{pmatrix}1\\ z_{jc}\end{pmatrix}(\sigma_j^{-1}(s_c;\hat{\lambda}) - \hat{\beta}_0 + \hat{\mu}p_{jc}) = 0. \tag{32}$$

Let $\bar{p}_c \equiv \sum_{j=1}^{J} s_{jc}p_{jc}$ be the share-weighted average price in market c. We show below that the $-\mu$ and λ^2 estimands can be expressed as the coefficients γ_0, γ_1 in the ω_j-weighted least squares regression of $\rho_j^{\log(q/q_0)}$ on $\tilde{p}_j \equiv \rho_j^{\frac{1}{2}p(p-2\bar{p})}$:

$$\rho_j^{\log(q/q_0)} = \gamma_0 + \gamma_1\tilde{p}_j + u_j \tag{33}$$

where u_j satisfies

$$\sum_{j=1}^{J}\omega_j\begin{pmatrix}1\\ \tilde{p}_j\end{pmatrix}u_j = 0.$$

Define a weighting function ω_j^p by

$$\omega_j^p(p_j) \equiv \frac{\mathbb{P}\{p_{jc} \geq p_j \mid \dot{z}_{jc} = 1\} - \mathbb{P}\{p_{jc} \geq p_j \mid \dot{z}_{jc} = 0\}}{\int_0^\infty \mathbb{P}\{p_{jc} \geq p_j' \mid \dot{z}_{jc} = 1\} - \mathbb{P}\{p_{jc} \geq p_j' \mid \dot{z}_{jc} = 0\}\,\mathrm{d}p_j'}. \tag{34}$$

Appendix A.8 shows that $\tilde{p}_j$ equals a ω_j^p-weighted average price level for the jth product, plus an approximately constant term:

$$\tilde{p}_j = \underbrace{\int_0^\infty p_j\omega_j^p(p_j)\,\mathrm{d}p_j}_{\text{weighted average price of } j} - \underbrace{\frac{\mathbb{E}[p_{jc}\bar{p}_c \mid \dot{z}_{jc} = 1] - \mathbb{E}[p_{jc}\bar{p}_c \mid \dot{z}_{jc} = 0]}{\mathbb{E}[p_{jc} \mid \dot{z}_{jc} = 1] - \mathbb{E}[p_{jc} \mid \dot{z}_{jc} = 0]}}_{\approx\mathbb{E}[\bar{p}_c]}. \tag{35}$$

The first term in (35) is a weighted average price level using the weighting function ω_j^p. This function gives more weight to points of the price distribution that are more shifted by the instrument. The weighting function is nonnegative if the distribution of p_{jc} given $\dot{z}_{jc} = 1$ first-order stochastically dominates the distribution of p_{jc} given $\dot{z}_{jc} = 0$ (or vice versa). The second term in (35) will approximately equal $\mathbb{E}[\bar{p}_c]$ unless the instrument $\dot{z}_{jc}$ has a large effect on $\bar{p}_c$. Since $\dot{z}_{jc}$ only affects the price of j, a large effect on $\bar{p}_c$ is unlikely when s_{jc} is small.

The following proposition states the key result of this subsection:

Proposition 10. *Let $(\hat{\mu}, \hat{\lambda})$ be the estimators in (32) with instruments z_{jc} as defined in Section 4.2. Then*

$$\operatorname*{plim}_{C\to\infty} -\hat{\mu} = \gamma_0 + O(|\lambda|^3)$$
$$\operatorname*{plim}_{C\to\infty} \hat{\lambda}^2 = \gamma_1 + O(|\lambda|^3)$$

where (γ_0, γ_1) are defined in (33). The $O(|\lambda|^3)$ convergence of the plim*s to (γ_0, γ_1) holds along a sequence of DGPs satisfying the regularity conditions specified in Lemma 5.*

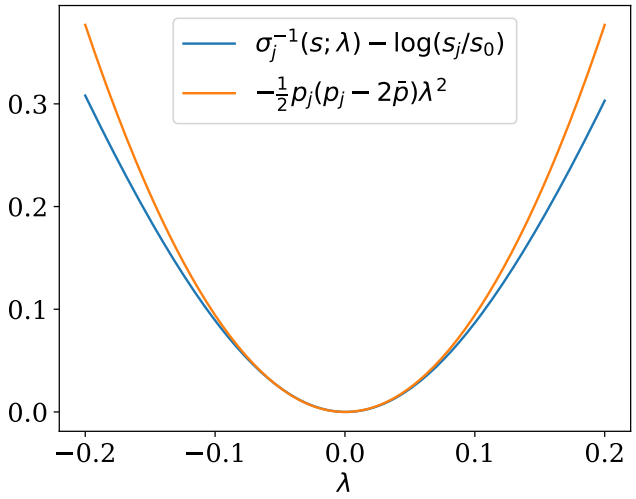


Figure 1: Numerical evaluation of the quadratic approximation to $\lambda \mapsto \sigma_j^{-1}(s;\lambda) - \log(s_j/s_0)$.

Proof. See Appendix A.9. □

The proof of Proposition 10 involves two key lemmas. The first (Lemma 4) computes the second-order Taylor expansion of the share inversion function σ^{-1} in the parameter λ:

$$\sigma_j^{-1}(s_c;\lambda) = \log(s_{jc}/s_{0c}) - \frac{1}{2}p_{jc}(p_{jc} - 2\bar{p}_c)\lambda^2 + O(\lambda^3). \tag{36}$$

Figure 1 shows the accuracy of this approximation in a simulated example, comparing the function $\lambda \mapsto \sigma_j^{-1}(s;\lambda) - \log(s_j/s_0)$ (calculated using a contraction-mapping iteration) against the quadratic approximation $\lambda \mapsto -\frac{1}{2}p_j(p_j - 2\bar{p})\lambda^2$. Lemma 4 suggests approximating the solution of the nonlinear moment condition (31) by the solution of the linear moment condition[10]

$$\mathbb{E}\left[J^{-1}\sum_{j=1}^{J} z_{jc}\left(\log(s_{jc}/s_{0c}) - \beta_0 + \mu p_{jc} - \frac{1}{2}p_{jc}(p_{jc} - 2\bar{p}_c)\lambda^2\right)\right] = 0. \tag{37}$$

The second key lemma (Lemma 5) is an epi-convergence result, establishing conditions under which the solution to the linear moment condition (37) is close to the solution to the nonlinear moment condition (31). Under these relatively weak conditions, the relative error in the approximation goes to 0 as the value of λ solving (31) approaches 0.

We note a connection between the linear moment condition (37) and the small-share approximation (16). Consider the model

$$\log(s_{jc}(p_c)) - \log(s_{0c}(p_c)) = \beta_0 - \mu p_{jc} + \lambda^2\frac{1}{2}p_{jc}(p_{jc} - 2\bar{p}_c) + \xi_{jc},$$

[10] Salanié and Wolak (2022) previously derived the same approximation to the BLP moment conditions. They used the approximation to obtain a computationally-efficient approximate estimator.

which would naturally suggest linear IV regression of $\log(s_{jc}/s_{0c})$ on $1, -p_j, \frac{1}{2}p_{jc}(p_{jc} - 2\bar{p}_c)$ as an estimator. Assume that $s_{jc}(p) \approx 0$, so that $\bar{p}$ and $\log(s_{0c}(p))$ both respond negligibly to a change in p_j. The own-price elasticity of s_{jc} under this model is

$$\eta_{jjc}(p) = p_j \frac{\partial \log(s_{jc})}{\partial p_j}(p_c) = -\mu p_{jc} + \lambda^2 p_{jc}(p_{jc} - \bar{p}_c),$$

which is exactly the small-share elasticity approximation (16). The small-share approximation highlights one observable implication of the random coefficient logit model: own-price semi-elasticities are (approximately) linearly increasing in $p_{jc} - \bar{p}_c$. Proposition 10 shows that this observable implication is (approximately) what the BLP estimator uses to pin down the variance of α_i.

The linear moment condition (37) also sheds light on the role of "instruments for shares" in the identification and estimation of random coefficient logit models. The vector of shares s_c enter the specification through the share-weighted average price $\bar{p}_c$ which appears in the endogenous regressor $\frac{1}{2}p_{jc}(p_{jc} - 2\bar{p}_c)$. In our approximation, the variance parameter λ^2 can be identified *without* explicit instruments for shares: instruments that vary both p_{jc} and p_{jc}^2 will generally suffice to estimate both μ and λ^2. Such identification is made possible by the functional form assumption that price enters the indirect utilities linearly; if price instead entered nonparametrically, one would vary the share-weighted average price $\bar{p}_c$ while holding p_{jc} fixed to pin down λ^2.

Proposition 10 and the small-share approximation (16) immediately imply:

Corollary 3. *Let $\hat{\eta}_{j'j'c'}(p_{c'})$ denote the own-price elasticity of product j' in market c' from the random coefficient logit model estimated using (32) with instruments z_{jc} as described in Section 4.2. Then*

$$\operatorname*{plim}_{C\to\infty} \hat{\eta}_{j'j'c'}(p_{c'}) = \left[\gamma_0 + \gamma_1 \left(p_{j'c'} - \bar{p}_{c'}\right)\right] p_{j'c'} + O(|\lambda|^3) + O(s_{j'c'}) \tag{38}$$

where (γ_0, γ_1) are defined in (33).

This corollary shows that the random coefficient logit own-price elasticity is a (common) quadratic function of the product's price, and that the quadratic relationship is determined by a weighted *regression* of product-specific IV estimands onto a constant and average price levels. Adding a random coefficient on price to the logit model provides one additional degree of freedom to the system of own-price elasticities, but own-price elasticities still vary across products only with their price levels.

4.5.2 Extension: models with demographics

We next consider the model (18) that allows the random coefficient α_i to vary with a demographic feature d_{ic} drawn from a known distribution F_{dc}. The results of the previous subsections generalize to this model in a straightforward way. To simplify the presentation, we assume $\lambda = 0$, so α_i varies only due to variation in d_{ic} across individuals.

The unknown parameters are the average price sensitivity μ and the coefficient π that measures how price sensitivity α_i varies with the demographic d_{ic}. The moment condition

(31) becomes

$$\mathbb{E}\left[J^{-1}\sum_{j=1}^{J} z_{jc}(\sigma_j^{-1}(s_c, F_{dc}; \pi) - \beta_0 + \alpha p_{jc})\right] = 0. \tag{39}$$

The random coefficient logit estimator solves the sample analogue of the moment conditions (39).

We show below that μ and π can be expressed in terms of the coefficients γ_0, γ_1 of the ω_j-weighted least squares regression of $\rho_j^{\log(q/q_0)}$ on $\tilde{d}_j \equiv \rho_j^{p\bar{d}}$, where $\bar{d}_c \equiv \int d_{ic}\, \mathrm{d}F_{dc}(d_{ic})$ is the average demographic in market c:

$$\rho_j^{\log(q/q_0)} = \gamma_0 + \gamma_1 \tilde{d}_j + u_j \tag{40}$$

where u_j satisfies

$$\sum_{j=1}^{J} \omega_j \begin{pmatrix} 1 \\ \tilde{d}_j \end{pmatrix} u_j = 0.$$

A derivation similar to that in Appendix A.8 shows that $\tilde{d}_j$ equals a weighted average of $\bar{d}_c$:

$$\tilde{d}_j = \int_0^\infty \mathbb{E}[\bar{d}_c \mid p_{jc}(0) \le p_j \le p_{jc}(1)]\omega_j^{p+}(p_j)\,\mathrm{d}p_j - \int_0^\infty \mathbb{E}[\bar{d}_c \mid p_{jc}(1) \le p_j \le p_{jc}(0)]\omega_j^{p-}(p_j)\,\mathrm{d}p_j \tag{41}$$

where ω_j^{p+} and ω_j^{p-} are the positive and negative parts of ω_j^p defined in (34).

We now state the main result of this subsection:

Proposition 11. *Let $(\hat{\mu}, \hat{\pi})$ be the estimators solving the sample analogue of (39) with instruments z_{jc} as defined in Section 4.2. Then*

$$\operatorname*{plim}_{C\to\infty} -\hat{\mu} = \gamma_0 + O(\pi^2)$$
$$\operatorname*{plim}_{C\to\infty} \hat{\pi} = \gamma_1 + O(\pi^2)$$

where (γ_0, γ_1) are defined in (40).

Proof. See Appendix A.10. □

Proposition 11 and the small-share approximation (19) immediately imply:

Corollary 4. *Let $\hat{\eta}_{j'j'c'}(p_{c'})$ denote the own-price elasticity of product j' in market c' from the random coefficient logit model with demographics estimated using the sample analogue of (39) with instruments z_{jc} as described in Section 4.1. Then*

$$\operatorname*{plim}_{C\to\infty} \hat{\eta}_{j'j'c'}(p_{c'}) = \left[\gamma_0 + \gamma_1^2 \operatorname{Var}_{i\in c}[d_{ic}]\,(p_{j'c'} - \bar{p}_{c'})\right] p_{j'c'} + O(\pi^2) + O(s_{j'c'}) \tag{42}$$

where (γ_0, γ_1) are defined in (40).

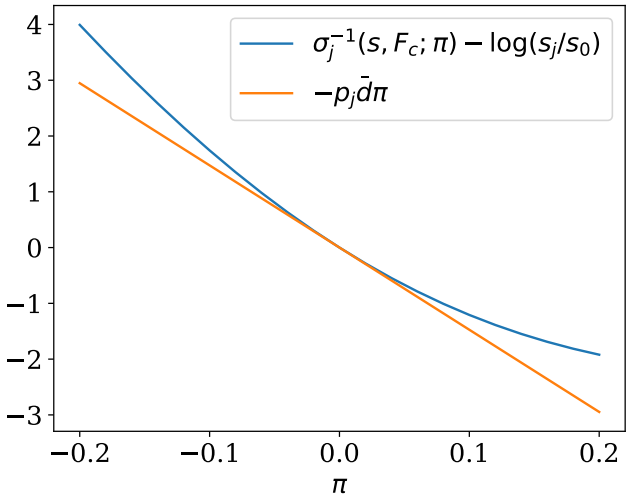


Figure 2: Numerical evaluation of the linear approximation to $\pi \mapsto \sigma_j^{-1}(s, F; \pi) - \log(s_j/s_0)$.

This corollary shows that in the random coefficient logit model with demographics, own-price elasticities remain quadratic functions of price. The coefficient on the quadratic term, γ_1^2, measures how different the average semi-elasticity of demand is between markets with higher and lower average values of the demographic. The intuition behind this comparison is that under the random coefficient logit model with demographics, price elasticities will vary across markets with different average demographics, as shown by the small-share approximation (19).

Like that of Proposition 10, the proof of Proposition 11 involves computing a Taylor approximation of the share inversion function σ^{-1}:

$$\sigma_j^{-1}(s_c, F_{dc}; \pi) = \log(s_{jc}/s_{0c}) - p_{jc}\bar{d}_c\pi + O(\pi^2).$$

Figure 2 illustrates this approximation.

4.5.3 Extension: multiple random coefficients

We next consider the model (20) that adds a random coefficient β_i on an R-dimensional vector of product characteristics x_{jc}. The results of the previous subsections generalize to this model in a straightforward way. To simplify the presentation, we assume that product characteristics x_{jc} do not vary across markets c (as is common in applications).

The unknown parameters are the mean μ and variance-covariance matrix $\Lambda \equiv \text{Var}[(\alpha_i, \beta_i')]$. Since β_i is random, and its distribution must also be estimated, we use instruments z_{jc}^x in addition to the price instruments z_{jc} described in Section 4.2. The moment condition (24) becomes

$$\mathbb{E}\left[J^{-1}\sum_{j=1}^{J}\begin{pmatrix}1\\ z_{jc}\\ z_{jc}^x\end{pmatrix}(\sigma_j^{-1}(s_c;\Lambda) - \mu_{1:R}'x_j + \mu_0 p_{jc})\right] = 0. \tag{43}$$

The random coefficient logit estimator solves the sample analogue of the moment conditions (43).

We show below that μ and Λ can be expressed in terms of the coefficients $\gamma_1, \gamma_p, \gamma_x$ in the ω_j-weighted least squares regression of $\rho_j^{\log(q/q_0)}$ on $\tilde{p}_j \equiv \rho_j^{\frac{1}{2}p(p-2\bar{p})}$ and $\tilde{x}_{jr} \equiv \rho_j^{px_r - p\bar{x}_r - x_r\bar{p}}$ $(r = 1, \ldots, R)$:

$$\rho_j^{\log(q/q_0)} = \gamma_1 + \gamma_p \tilde{p}_j + \gamma_x' \tilde{x}_j + u_j \tag{44}$$

where u_j satisfies

$$\sum_{j=1}^{J} \omega_j \begin{pmatrix} 1 \\ \tilde{p}_j \\ \tilde{x}_j \end{pmatrix} u_j = 0.$$

A derivation similar to that in Appendix A.8 shows that $\tilde{x}_{jr}$ equals x_{jr} minus two approximately constant terms:

$$\tilde{x}_{jr} = x_{jr} - \underbrace{\frac{\mathbb{E}[p_{jc}\bar{x}_{rc} \mid \dot{z}_{jc} = 1] - \mathbb{E}[p_{jc}\bar{x}_{rc} \mid \dot{z}_{jc} = 0]}{\mathbb{E}[p_{jc} \mid \dot{z}_{jc} = 1] - \mathbb{E}[p_{jc} \mid \dot{z}_{jc} = 0]}}_{\approx \mathbb{E}[\bar{x}_{rc}]} - \underbrace{\frac{\mathbb{E}[x_{jr}\bar{p}_c \mid \dot{z}_{jc} = 1] - \mathbb{E}[x_{jr}\bar{p}_c \mid \dot{z}_{jc} = 0]}{\mathbb{E}[p_{jc} \mid \dot{z}_{jc} = 1] - \mathbb{E}[p_{jc} \mid \dot{z}_{jc} = 0]}}_{\approx 0}. \tag{45}$$

Intuitively, the second term in (45) will approximately equal $\mathbb{E}[\bar{x}_{rc}]$ unless $\dot{z}_{jc}$ has a large effect on $\bar{x}_{rc}$. Since $\dot{z}_{jc}$ only affects the price of j, a large effect is unlikely when s_{jc} is small. Similarly, the third term in (45) will approximately equal 0 unless the instrument $\dot{z}_{jc}$ has a large effect on $\bar{p}_c$.

We now state the main result of this subsection:

Proposition 12. *Let $(\hat{\mu}, \hat{\Lambda})$ be the estimators solving the sample analogue of (43) with instruments z_{jc} as defined in Section 4.2 and additional instruments z_{jc}^x. Suppose that the matrix of coefficients from a regression of $p_{jc}, x_{jc}, \frac{1}{2}p_{jc}(p_{jc} - 2\bar{p}_c), \frac{1}{2}x_{jrc}(x_{jrc} - 2\bar{x}_{rc})$ $(1 \leq r \leq R)$, $p_{jc}x_{jsc} - p_{jc}\bar{x}_{sc} - x_{jsc}\bar{p}_c$ $(1 \leq s \leq R)$, and $x_{jrc}x_{jsc} - x_{jrc}\bar{x}_{sc} - x_{jsc}\bar{x}_{rc}$ $(0 \leq r < s \leq R)$ onto $1, z_{jc}, z_{jc}^x$ has a block diagonal structure, such that regressands involving prices load only onto z_{jc} and regressands not involving prices load only onto z_{jc}^x. Then*

$$\begin{aligned} \operatorname*{plim}_{C\to\infty} -\hat{\mu}_0 &= \gamma_1 + o(\|\Lambda\|) \\ \operatorname*{plim}_{C\to\infty} \hat{\Lambda}_{0,0} &= \gamma_p + o(\|\Lambda\|) \\ \operatorname*{plim}_{C\to\infty} \hat{\Lambda}_{0,1:R} &= \gamma_x' + o(\|\Lambda\|) \end{aligned}$$

where $(\gamma_1, \gamma_p, \gamma_x')$ are defined in (44).

Proof. See Appendix A.11. □

Proposition 12 and the small-share approximation (21) immediately imply:

Corollary 5. *Let $\hat{\eta}_{j'j'c'}(p_{c'})$ denote the own-price elasticity of product j' in market c' from random coefficient logit model estimated using the sample analogue of (43). Assume the conditions in Proposition 12. Then*

$$\operatorname*{plim}_{C\to\infty} \hat{\eta}_{j'j'c'}(p_{c'}) = \left[\gamma_1 + \gamma_p(p_{j'c'} - \bar{p}_{c'}) + \gamma_x'\left(x_{j'} - \bar{x}_{c'}\right)\right] p_{j'c'} + o(\|\Lambda\|) + O(s_{j'c'})$$

where $(\gamma_1, \gamma_p, \gamma_x')$ are defined in (44).

This corollary shows that the multiple random coefficient logit own-price elasticity is a quadratic function of price conditional on product characteristics. Adding a random coefficient on a non-price characteristic—and allowing this random coefficient to be correlated with the random coefficient on price—adds this characteristic as an additional regressor in the weighted regression of IV estimands on average price levels. Corollary 5 thus provides a basis for the common intuition that characteristic-space demand models "project elasticities onto characteristics," though we stress that allowing for non-zero correlation between α_i and β_i is essential for this interpretation.

4.6 Cross-price elasticities

We now shift our attention to the identification and interpretation of cross-price elasticities in the models. We first consider logit, product-specific logit, and simple (i.e., constant β) random coefficient logit models. Recall from Proposition 7 that cross-price elasticities in these models depend on the *same* parameters as own-price elasticities: α for the standard logit model, α_j $(j = 1, \ldots, J)$ for the product-specific logit model, and $\mathbb{E}[\alpha_i]$ and $\text{Var}[\alpha_i]$ for the simple random coefficient logit model. The approximate identification results obtained above for these parameters thus lead immediately to the following corollary on cross-price elasticities in these models.

Corollary 6. *(a) In the standard logit model,*

$$\operatorname*{plim}_{C\to\infty} \hat{\eta}_{kj}(p) = -\left(\sum_{j=1}^{J} \omega_j \rho_j^{\log(q/q_0)}\right) p_j s_j.$$

(b) In the product-specific logit model,

$$\operatorname*{plim}_{C\to\infty} \hat{\eta}_{kj}(p) = -\rho_j^{\log(q/q_0)} p_j s_j.$$

(c) In the simple random coefficient logit model,

$$\operatorname*{plim}_{C\to\infty} \hat{\eta}_{kj}(p) = -\gamma_0 p_j s_j \left[1 + \gamma_1 (p_j - \bar{p})(p_k - \bar{p})\right] + \gamma_1 (2\bar{p} - p_k - p_j) p_j s_j + O(|\lambda^3|)$$

where (γ_0, γ_1) *are defined in* (33).

Corollary 6 shows that cross-price elasticities are constructed by projecting product-specific IV estimates of demand onto a constant (standard logit model), a product-specific constant (product-specific logit model), or a constant and the average price level (random coefficient logit model). The IV estimates used in these projections reflect *own* price responses to exogenous changes in prices, not substitution or cross-price responses across products.

The construction of cross-price elasticities becomes more complicated if we allow the coefficient on product characteristics β to be a random coefficient. As shown in Proposition 7, the cross-price elasticities in this model depend on the variance of the non-price random coefficients β_i, even if α is constant or uncorrelated with β_i. For simplicity, consider a random

coefficient logit model with a random α_i and an uncorrelated random β_i on a single scalar characteristic x_{jc}. The estimands of μ_α, λ_α^2, μ_β, and λ_β^2 are approximately the coefficients in the population IV regression

$$\log(s_{jc}/s_{0c}) = \beta_0 - \mu_\alpha p_{jc} + \lambda_\alpha^2 \frac{1}{2}(p_{jc}^2 - 2p_{jc}\bar{p}_c) + \mu_\beta x_{jc} + \lambda_\beta^2 \frac{1}{2}(x_{jc}^2 - 2x_{jc}\bar{x}_c) + \xi_{jc} \tag{46}$$

and the cross-price elasticity of product k with respect to the price of product j is

$$\begin{aligned}\operatorname*{plim}_{C\to\infty} \hat{\eta}_{kj}(p) = \mu_\alpha p_j s_j \left[1 + \lambda_\alpha^2(p_j - \bar{p})(p_k - \bar{p}) + \lambda_\beta^2(x_j - \bar{x})(x_k - \bar{x})\right] \\ + \lambda_\alpha^2(2\bar{p} - p_j - p_k)p_j s_j + O(\|(\lambda_\alpha, \lambda_\beta)\|^3).\end{aligned}$$

The coefficient on the regressor $\frac{1}{2}(x_{jc}^2 - 2x_{jc}\bar{x}_c)$ in (46) determines the cross-price elasticities: the larger this coefficient is, the greater the implied substitution between products with similar levels of characteristic x.

It is useful to observe that the regressor $\frac{1}{2}(x_{jc}^2 - 2x_{jc}\bar{x}_c)$ is endogenous in (46), even if x_{jc} is assumed exogenous, since $\bar{x}_c = \sum_{k=1}^J s_{kc}x_{kc}$ is a function of market shares. Identification of (46) thus requires instruments both for price and for $\frac{1}{2}(x_{jc}^2 - 2x_{jc}\bar{x}_c)$. If x_{jc} takes more than two values and is assumed exogenous, then x_{jc}^2 can serve as an instrument for $\frac{1}{2}(x_{jc}^2 - 2x_{jc}\bar{x}_c)$. Otherwise, "instruments for shares" will be required to estimate (46). The resulting estimand of λ_β^2 will in general depend on the choice of these instruments. However, for standard choices, such as the Berry et al. (1995) instruments or the Gandhi and Houde (2019) differentiation instruments, the λ_β^2 estimand will not depend directly on any cross-price IV estimands, and thus neither will the cross-price elasticities η_{kj}. We refer to Appendix ??? for a proof of this claim.

4.7 Results for general instruments

The previous subsections give particularly simple expressions for the logit, the product-specific logit, and the random coefficient logit estimands under particular assumptions about the form of the instrument z_{jc}. These assumptions are not essential for the substantive results. We next present closed-form approximations to these estimates for an arbitrary L-dimensional instrument $z_{jc} = (z_{jc1}, \ldots, z_{jcL})'$. Without loss of generality, we assume that the instruments are pooled orthonormal in the sense that

$$\mathbb{E}\left[J^{-1}\sum_{j=1}^J (z_{jc} - \bar{z})(z_{jc} - \bar{z})'\right] = I_{L\times L},$$

where $\bar{z} \equiv J^{-1}\sum_{j=1}^J \mathbb{E}[z_{jc}]$ and $I_{L\times L}$ is the $L \times L$ identity matrix.

To simplify notation in the results to follow, we define

$$\rho_\ell^y \equiv \frac{\mathbb{E}[J^{-1}\sum_{j=1}^J (z_{jc\ell} - \bar{z}_\ell)y_{jc}]}{\mathbb{E}[J^{-1}\sum_{j=1}^J (z_{jc\ell} - \bar{z}_\ell)p_{jc}]}, \qquad \omega_\ell \equiv \mathbb{E}\left[J^{-1}\sum_{j=1}^J (z_{jc\ell} - \bar{z}_\ell)p_{jc}\right].$$

ρ^y_ℓ is the estimand from a pooled regression of y_{jc} on p_{jc} using $z_{jc\ell}$ as an instrument, while ω_ℓ is the pooled covariance between $z_{jc\ell}$ and p_{jc}. We also define the associated within- and between-product IV estimands

$$\rho^y_{j\ell} \equiv \frac{\mathbb{E}[(z_{jc\ell} - \mathbb{E}[z_{jc\ell}])y_{jc}]}{\mathbb{E}[(z_{jc\ell} - \mathbb{E}[z_{jc\ell}])p_{jc}]} \qquad (j = 1, \ldots, J)$$
$$\rho^y_{0\ell} \equiv \frac{J^{-1}\sum_{j=1}^J (\mathbb{E}[z_{jc\ell}] - \bar{z}_\ell)\,\mathbb{E}[y_{jc}]}{J^{-1}\sum_{j=1}^J (\mathbb{E}[z_{jc\ell}] - \bar{z}_\ell)\,\mathbb{E}[p_{jc}]}$$

and weights

$$\omega_{j\ell} \equiv \frac{\mathbb{E}[(z_{jc\ell} - \mathbb{E}[z_{jc\ell}])p_{jc}]}{\mathbb{E}[J^{-1}\sum_{j=1}^J (z_{jc\ell} - \bar{z}_\ell)p_{jc}]} \qquad (j = 1, \ldots, J)$$
$$\omega_{0\ell} \equiv \frac{J^{-1}\sum_{j=1}^J (\mathbb{E}[z_{jc\ell}] - \bar{z}_\ell)\,\mathbb{E}[p_{jc}]}{\mathbb{E}[J^{-1}\sum_{j=1}^J (z_{jc\ell} - \bar{z}_\ell)p_{jc}]}.$$

Note that $\sum_{j=0}^J \omega_{j\ell} = 1$ and the following decomposition holds:

$$\rho^y_\ell = \sum_{j=0}^J \omega_{j\ell}\rho^y_{j\ell}.$$

4.7.1 Logit

For the logit model, Proposition 8 expressed the logit $-\alpha$ estimand as a weighted average of product-specific IV estimands. Proposition 13 below shows that in the general instrument case, the α estimand can be expressed as a weighted average of instrument-specific IV estimands.

Proposition 13. *Let $\hat{\alpha}$ be the estimator in (25) with pooled orthonormal instruments z_{jc}. Then*

$$\operatorname*{plim}_{C\to\infty}(-\hat{\alpha}) = \sum_{\ell=1}^L \frac{\omega_\ell^2}{\sum_{\ell'=1}^L \omega_{\ell'}^2}\rho_\ell^{\log(q/q_0)} = \sum_{\ell=1}^L \frac{\omega_\ell^2}{\sum_{\ell'=1}^L \omega_{\ell'}^2}\left(\sum_{j=0}^J \omega_{j\ell}\rho_{j\ell}^{\log(q/q_0)}\right). \tag{47}$$

Proof. See Appendix A.12. □

Proposition 13 involves two sets of weights: instrument weights $\omega_\ell^2 / \sum_{\ell'=1}^L \omega_{\ell'}^2$ and product weights $\omega_{j\ell}$ for each instrument. Both sets of weights sum to 1. While the instrument weights are nonnegative, the product weights $\omega_{j\ell}$ may vary in sign across products j unless a uniformity assumption is imposed (i.e., unless the correlation between $z_{jc\ell}$ and p_{jc} has the same sign for all j). The plausibility of this assumption depends on the choice of instrument.

When $\mathbb{E}[z_{jc}]$ varies with j, the probability limit of $-\hat{\alpha}$ involves not just product-specific IV estimands, but also *between*-product IV estimands $\rho^y_{0\ell}$. This term reflects that under the standard logit model, the common own-price semi-elasticity parameter α can be identified not just by cross-market comparisons of the same product at different prices, but also within-market comparisons of different products at different prices. These terms drop out of the $-\alpha$ estimand if product fixed effects are included in the model.

4.7.2 Product-specific logit

For the product-specific logit model, Proposition 9 expressed the product-specific logit $-\alpha_j$ estimand as the coefficient in a regression of product-specific IV estimands onto J mutually-exclusive product indicators. Proposition 14 below shows that in the general instrument case, the $-\alpha_j$ estimand is the coefficient in a regression of instrument-specific IV estimands onto J regressors that sum to 1. To state the proposition, define $\gamma_1, \ldots, \gamma_J$ to the coefficients in the ω_ℓ^2-weighted regression

$$\rho_\ell^{\log(q/q_0)} = \sum_{k=1}^{J} \gamma_k \frac{\mathbb{E}[J^{-1}(z_{kc\ell} - \bar{z}_\ell)p_{kc}]}{\mathbb{E}[J^{-1}\sum_{k'=1}^{J}(z_{k'c\ell} - \bar{z}_\ell)p_{k'c}]} + u_\ell. \tag{48}$$

Note that if $L < J$, the product-specific logit model is under-identified, and the coefficients γ_j cannot be defined.

Proposition 14. *Let $-\hat{\alpha}_j$ be the estimator in (28) with pooled orthonormal instruments z_{jc}. Then*

$$\operatorname*{plim}_{C\to\infty} -\hat{\alpha}_j = \gamma_j,$$

where γ_j is defined in (48).

Proof. See Appendix A.13. □

Like Proposition 9, Proposition 14 expresses the product-specific logit estimand $-\alpha_j$ as the coefficient in a weighted least squares regression of IV estimands onto J regressors. While the regression in Proposition 9 involved product-specific IV estimands, the regression in Proposition 14 involves instrument-specific estimands $\rho_\ell^{\log(q/q_0)}$. Moreover, the J regressors in Proposition 14 are no longer a set of mutually-exclusive product indicators: for a given value of ℓ, multiple regressors $\frac{\mathbb{E}[J^{-1}(z_{kc\ell}-\bar{z}_\ell)p_{kc}]}{\mathbb{E}[J^{-1}\sum_{k'=1}^{J}(z_{k'c\ell}-\bar{z}_\ell)p_{k'c}]}$ may be nonzero if instrument ℓ shifts prices for multiple products. Like mutually-exclusive indicators, the regressors always sum to 1 for each value of ℓ: $\sum_{k=1}^{J} \frac{\mathbb{E}[J^{-1}(z_{kc\ell}-\bar{z}_\ell)p_{kc}]}{\mathbb{E}[J^{-1}\sum_{k'=1}^{J}(z_{k'c\ell}-\bar{z}_\ell)p_{k'c}]} = 1$.

The intuition behind Proposition 14 is as follows. Each IV estimand $\rho_\ell^{\log(q/q_0)}$ measures a semi-elasticity. When each instrument ℓ varies the price of only a single product $j(\ell)$, then $\rho_\ell^{\log(q/q_0)}$ is a semi-elasticity for product $j(\ell)$, the regressors in (48) collapse to a set of J mutually exclusive indicators, and the coefficient γ_j recovers a weighted average of the estimands $\rho_\ell^{\log(q/q_0)}$ for all the instruments ℓ that vary the price of product j. More generally, each instrument ℓ could vary the price of multiple products. In this case, the regression (48) "portions out" the IV estimand $\rho_\ell^{\log(q/q_0)}$ among all the products j whose prices are varied by instrument ℓ.

4.7.3 Random coefficient logit

For the random coefficient logit model, Proposition 10 expressed the $-\mathbb{E}[\alpha_i]$ and $\mathrm{Var}[\alpha_i]$ estimands as coefficients in a weighted least squares regression of product-specific IV estimands on a constant and average price levels. Proposition 15 below shows that in the general

instrument case, these estimands can be expressed as coefficients in a weighted least squares regression of instrument-specific IV estimands on a constant and average price levels.

Proposition 15. *Let* $(\hat{\mu}, \hat{\lambda}^2)$ *be the estimators in* (32) *with pooled orthonormal instruments* $z_{jc} = (z_{jc1}, \ldots, z_{jcL})'$. *Then*

$$\operatorname*{plim}_{C\to\infty} -\hat{\mu} = \gamma_0 + O(|\lambda|^3)$$
$$\operatorname*{plim}_{C\to\infty} \hat{\lambda}^2 = \gamma_1 + O(|\lambda|^3)$$

where (γ_0, γ_1) *are coefficients from the the* ω_ℓ^2*-weighted least squares regression*

$$\rho_\ell^{\log(q/q_0)} = \gamma_0 + \gamma_1 \rho_\ell^{\frac{1}{2}p(p-2\bar{p})} + u_\ell.$$

Proof. The proof is similar to that of Propositions 10 and 13 and is omitted. □

Like Proposition 10, Proposition 15 approximately expresses the BLP estimands as coefficients in a weighted least squares regression of IV estimands. While the regression in Proposition 10 involved product-specific IV estimands, the regression in Proposition 15 involves instrument-specific IV estimands $\rho_\ell^{\log(q/q_0)}$, each of which is itself a weighted average of within- and between-product estimands $\rho_{j\ell}^{\log(q/q_0)}$.

5 Numerical simulation

For each model and estimator, we derived expressions relating estimands (i.e., the parameter values that solve the moment conditions) to underlying data moments. These expressions clarify how model elasticities relate to underlying data moments. Table 1 summarizes the theoretical results. We now illustrate these results using three simulated DGPs.

The DGPs we use are based on the synthetic cereal data generated by Nevo (2000). We use the same distribution of products, markets, and prices in the synthetic cereal data, and construct indirect utilities

$$v_{ij} = \hat{\gamma}_j - \alpha_{ij} p_{jc} + \varepsilon_{ij}.$$

The product fixed effects $\hat{\gamma}_j$ are obtained by fitting a random coefficient logit model to the original data. Product-specific instruments for estimation are constructed from Nevo's instruments; details are in Appendix A.14.

We consider three different DGPs, one for each of the three main models we consider:

1. Logit: α_{ij} is a constant determined by estimating a logit model on the Nevo (2000) data.

2. Product-specific (PS) logit: $\alpha_{ij} = \alpha_j$, where α_j is a linear function of the characteristics of product j:

$$\alpha_j = -20 + 0.2\text{sugar}_j - 0.6\text{mushy}_j.$$

Table 1: Summary of own-price approximation results.

Logit

$$\begin{aligned}\eta_{j'j'c'}(p_{c'}) &= -\alpha p_{j'c'}(1 - s_{j'c'}) \\ &\approx -\alpha p_{j'c'} \\ &= \left(\sum_{j=1}^{J} \omega_j \rho_j^{\log(q/q_0)}\right) p_{j'c'}\end{aligned}$$

Product-specific logit

$$\begin{aligned}\eta_{j'j'c'}(p_{c'}) &= -\alpha_j p_{j'c'}\left(1 - \frac{\mathbb{E}[s_{j'c'}(m_i)^2]}{\mathbb{E}[s_{j'c'}(m_i)]}\right) \\ &\approx -\alpha_j p_{j'c'} \\ &= \rho_{j'}^{\log(q/q_0)} p_{j'c'}\end{aligned}$$

Random coefficient logit

$$\begin{aligned}\eta_{j'j'c'}(p_{c'}) &= \frac{\int -\alpha_i p_{j'c'}(1 - s_{j'c'}(\alpha_i)) s_{j'c'}(\alpha_i)\, \mathrm{d}F_\alpha(\alpha_i)}{\int s_{j'c'}(\alpha_i)\, \mathrm{d}F_\alpha(\alpha_i)} \\ &\approx -\mathbb{E}[\alpha_i] p_{j'c'} + \mathrm{Var}[\alpha_i](p_{j'c'} - \bar{p}_{c'}) p_{j'c'} \\ &\approx \gamma_0 p_{j'c'} + \gamma_1 (p_{j'c'} - \bar{p}_{c'}) p_{j'c'}\end{aligned}$$

where (γ_0, γ_1) are coefficients from ω_j-weighted regression $\rho_j^{\log(q/q_0)} = \gamma_0 + \gamma_1 \tilde{p}_j + u_j$

Multiple random coefficient logit

$$\begin{aligned}\eta_{j'j'c'}(p_{c'}) &= \frac{\int -\alpha_i p_{j'c'}(1 - s_{j'c'}(\alpha_i, \beta_i)) s_{j'c'}(\alpha_i, \beta_i)\, \mathrm{d}F_{\alpha,\beta}(\alpha_i, \beta_i)}{\int s_{j'c'}(\alpha_i, \beta_i)\, \mathrm{d}F_{\alpha,\beta}(\alpha_i, \beta_i)} \\ &\approx -\mathbb{E}[\alpha_i] p_{j'c'} + \mathrm{Var}[\alpha_i](p_{j'c'} - \bar{p}_{c'}) p_{j'c'} - \mathrm{Cov}[\alpha_i, \beta_i'](x_{j'} - \bar{x}_{c'}) \\ &\approx -\gamma_1 p_{j'c'} + \gamma_p (p_{j'c'} - \bar{p}_{c'}) p_{j'c'} + \gamma_x'(x_{j'} - \bar{x}_{c'})\end{aligned}$$

where $(\gamma_1, \gamma_p, \gamma_x')$ are coefficients from ω_j-weighted regression $\rho_j^{\log(q/q_0)} = \gamma_1 + \gamma_p \tilde{p}_j + \gamma_x' \tilde{x}_j + u_j$

3. Random coefficient (RC) logit: $\alpha_{ij} = \alpha_i$, where α_i follows a normal distribution determined by estimating a random coefficient logit model on the Nevo (2000) data.

We then calculate shares and own-price elasticities for each product under each of the three DGPs.

Own-price semi-elasticities

Figure 3 plots the true own-price semi-elasticity for each product, averaged across markets, against that product's average price level for each of the three DGPs described above. The

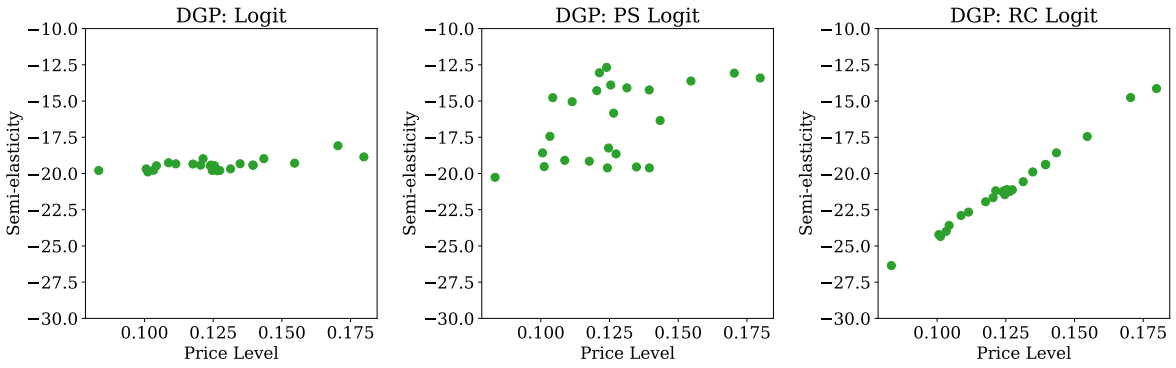


Figure 3: Product own-price semi-elasticities against price levels for each DGP.

patterns closely match those suggested by Proposition 2. Under the logit DGP (left panel), semi-elasticities are approximately constant across products, and hence elasticities are approximately linear in prices. Under the product-specific logit DGP (middle panel), there is no systematic relationship between price levels and semi-elasticities across products. Under the random coefficient logit DGP (right panel), semi-elasticities are approximately linear in price, and hence elasticities are approximately quadratic in prices. The linear relationship between prices and semi-elasticities holds even though the standard deviation of the price coefficient is relatively large in this DGP (close to half the mean price coefficient).

Identification

We now use the three DGPs to illustrate the relationship between model semi-elasticity estimates and IV moments in the data. The green points in Figure 4 show the IV estimates $\hat{\rho}_j^{\log(q/q_0)}$ for each product j in each of the three DGPs. The IV estimates naturally follow patterns similar to the true semi-elasticities: they are constant under the logit DGP, unrelated to price under the product-specific DGP, and linearly increasing in price under the random coefficient logit DGP.

Next, we illustrate how the own-price elasticities of the different models can be approximated from the IV estimates. The dashed lines in Figure 4 shows the Proposition 8 approximation to the logit semi-elasticity estimator: an $\hat{\omega}_j$-weighted average of the IV estimates $\hat{\rho}_j^{\log(q/q_0)}$. The blue crosses show the exact semi-elasticities from a logit model fit to the DGP. Across all three DGPs, the exact logit estimates lie close to the approximating line, deviating only because shares are not exactly 0.

Figure 4 highlights that the logit model always generates approximately constant semi-elasticities, whether or not semi-elasticities are approximately constant under the true DGP. Differences across DGPs may cause the constant semi-elasticity to shift up or down, but only insofar as the weighted average IV estimate differs across DGPs.

We next estimate a product-specific logit model on each DGP. The dashed lines in Figure 5 show the Proposition 9 approximation to the product-specific logit semi-elasticity estimator. The approximation coincides exactly with the IV estimates themselves. The exact semi-elasticities from the estimated model lie close to the approximate semi-elasticities and hence to the IV estimates.

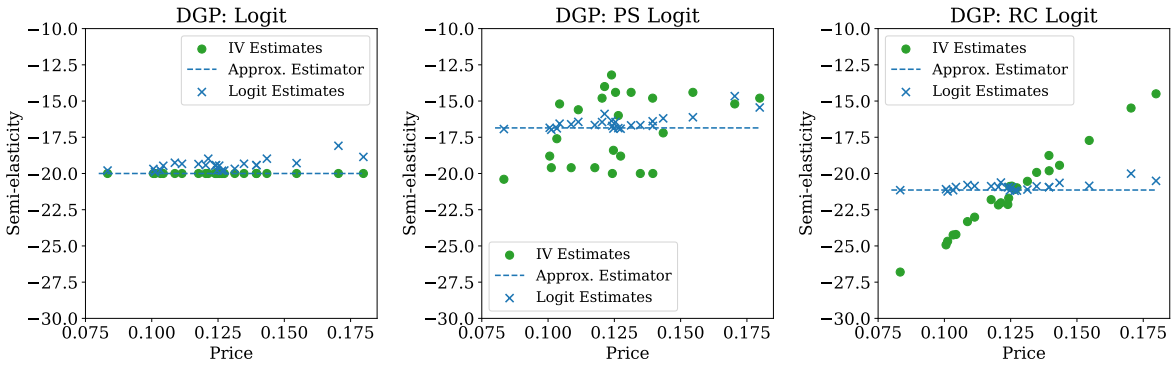


Figure 4: Semi-elasticity IV estimates and logit model-based estimates against product price levels for each of the three DGPs.

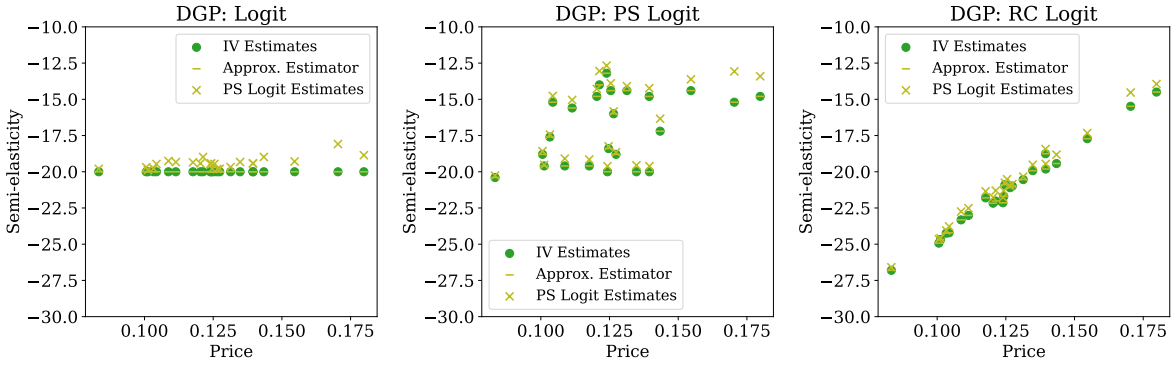


Figure 5: Semi-elasticity IV estimates and product-specific logit model-based estimates against product price levels for each of the three DGPs.

Figure 5 highlights that the product-specific logit model, being just-identified, will always be able to reproduce the IV estimates. This gives the product-specific model significant flexibility in fitting the true semi-elasticities. Under the logit DGP, the model generates approximately constant semi-elasticities; under the random coefficient logit DGP, the model generates semi-elasticities that are linearly increasing in price; and under the product-specific DGP, the model generates semi-elasticities that closely track the true semi-elasticities for each product.

We turn finally to the random coefficient logit estimator. The dashed lines in Figure 6 show the Proposition 10 approximation to the random coefficient logit estimator: an $\hat{\omega}_j$-weighted least squares regression of the IV estimates $\hat{\rho}_j^{\log(q/q_0)}$ onto a constant and an average price level. The orange dots shows the exact semi-elasticities from a random coefficient logit model fit to the data using the BLP estimator. Again, the exact semi-elasticities from the model closely follow the approximation.

Figure 6 highlights that the random coefficient logit model always generates semi-elasticities that are approximately linear in price, whether or not semi-elasticities are approximately linear in price under the true DGP. Differences across DGPs vary the intercept and slope of the line that the model draws through the IV estimates. When semi-elasticities are approx-

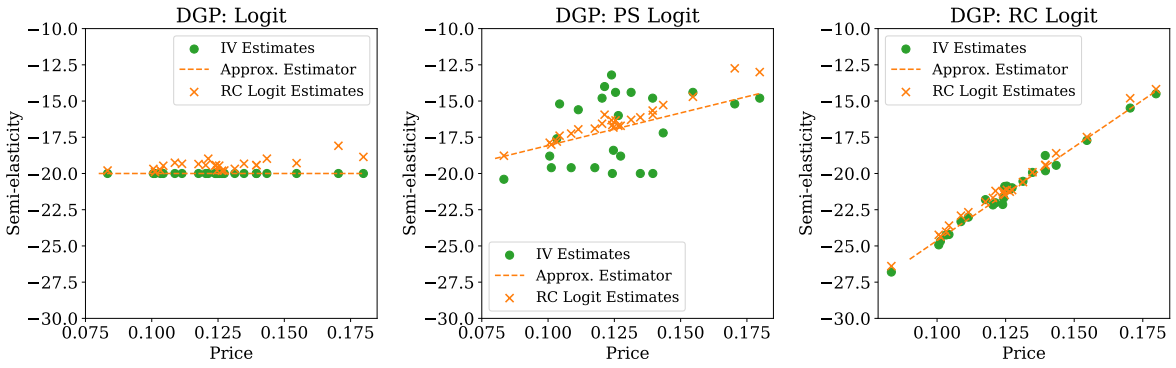


Figure 6: IV and random coefficient logit estimates of semi-elasticities against product price levels for each of the three DGPs.

imately linear in price, as under the logit and random coefficient logit DGPs, the random coefficient logit model produces semi-elasticities that closely track the IV estimates. However, when semi-elasticities are not linear in price, as under the product-specific logit DGP, the semi-elasticities from the model track price levels more than IV estimates.

Thick-tailed/skewed distribution of α_i

The right-hand panel of Figure 6 also highlights that when the random coefficient on price is normally distributed, the approximate semi-elasticities are close to the exact semi-elasticities from the estimated model. As noted in Section 3, we expect the approximation to be less accurate when the distribution of the random coefficient α_i on price is thick-tailed and/or skewed.

To investigate the effect of tail thickness on the accuracy of the approximation, we consider random coefficient logit DGPs with two different thick-tailed distributions of α_i: a Student's t distribution with 5 degrees of freedom and a Cauchy distribution. Figure 7 plots the results. When α_i follows a $t(5)$ distribution, the estimated elasticities of the random coefficient logit model still closely follow the approximating weighted least squares regression line. When α_i follows a Cauchy distribution, the exact semi-elasticities begin to diverge from the approximate semi-elasticities, particularly for low and high prices. Nonetheless, the approximation still captures much of variation in semi-elasticities across products.

To investigate the effects of skewness on the accuracy of the approximation, we consider random coefficient logit DGPs with two different skewed distributions of α_i: a log-normal distribution and an exponential distribution. Figure 8 plots the results. When α_i follows a log-normal distribution, the estimated elasticities of the random coefficient logit model still closely follow the approximating weighted least squares regression line, though some divergence is apparent for medium- to high-priced products. The pattern is similar when α_i follows an exponential distribution.

Finally, we consider a random coefficient logit DGP with a skewed and thick-tailed Pareto(1) distribution of α_i. The results are in Figure 9. Once again, the elasticity estimates closely follow the approximating line.

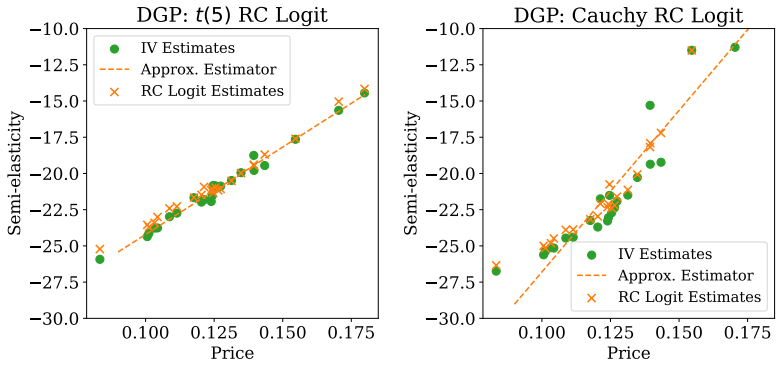


Figure 7: IV and random coefficient logit estimates of semi-elasticities against product price levels for DGPs with thick-tailed α_i distributions.

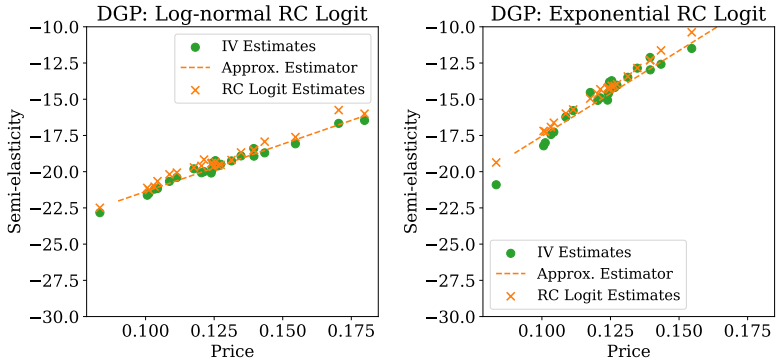


Figure 8: IV and random coefficient logit estimates of semi-elasticities against product price levels for DGPs with log-normal and exponential α_i distributions.

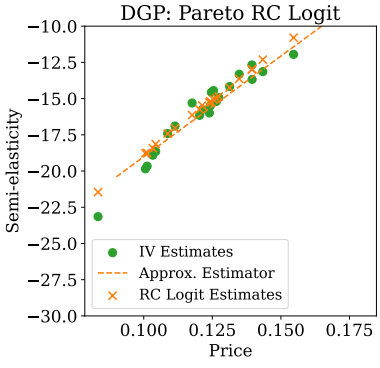


Figure 9: IV and random coefficient logit estimates of semi-elasticities against product price levels for DGP with Pareto(1) α_i distribution.

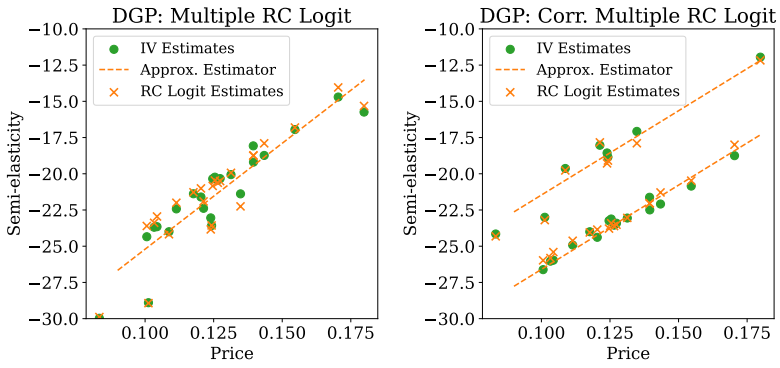


Figure 10: IV and random coefficient logit estimates of semi-elasticities against product price levels for DGPs with multiple random coefficients.

Multiple random coefficients

In Section 3.2.2, we noted that when multiple random coefficients are included in the logit model, own-price elasticities remain approximately quadratic in price conditional on product characteristics. Further, when the other random coefficients are uncorrelated with the random coefficient on price, the approximate quadratic is common across all products.

To investigate the accuracy of these approximations, we modify the random coefficient logit DGP to include a random coefficient on a non-price characteristic ("mushy"):

$$v_{ij} = \hat{\gamma}_j - \alpha_i p_{jc} + \beta_i \text{mushy}_j + \varepsilon_{ij}$$

with (α_i, β_i) bivariate normal. We simulate data from two different versions of the DGP, one in which α_i and β_i are uncorrelated and one in which they have correlation -0.5. We then estimate the corresponding random coefficient logit model on each version.

Figure 10 plots the results. When α_i and β_i are uncorrelated (left panel), the estimated elasticities of the multiple random coefficient logit model still closely follow the approximating weighted least squares regression line. Differences between the exact and approximate elasticities are greater than for the single random coefficient logit model because the presence of multiple random coefficients in the model reduces the accuracy of the small share approximation; nonetheless, the single regression line captures most of the variation in own-price elasticities across products.

When α_i and β_i are negatively correlated (right panel), the approximate estimator is a weighted least squares regression with an intercept shift for the product characteristic "mushy"; in other words, the approximate estimator consists of two *parallel* regression lines, one for mushy products and one for non-mushy products. The exact elasticities closely follow the two approximating lines. The approximating line for mushy products lies above that for non-mushy products; this is because taste for mushy products is negatively correlated with price sensitivity in this DGP; thus low-sensitivity consumers tend to select mushy products, making demand for these products less elastic.

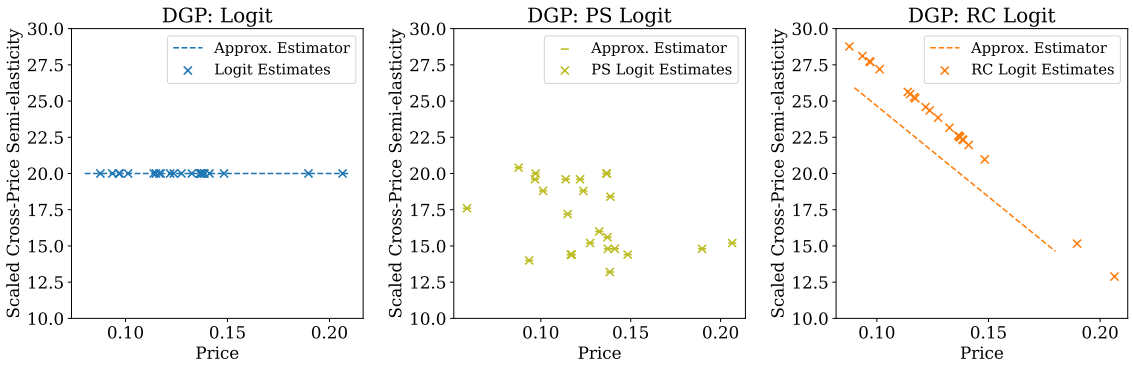


Figure 11: Actual and approximate estimates of cross-price elasticities in standard, product-specific, and random coefficient logit models.

Cross-price elasticities

We conclude with an illustration of our approximations to cross-price elasticities in the standard, product-specific, and random coefficient logit models. For all three models, our approximation to the cross-price semi-elasticity of product k with respect to the price of product j in market c, $\hat{\eta}_{kjc}/p_{jc}$, is proportional to the share s_{jc}; additionally, under the random coefficient logit model, the approximate cross-price semi-elasticity also depends on the price of product k. For these reasons, we focus below on scaled cross-price semi-elasticities $\hat{\eta}_{kjc}/(p_{jc}s_{jc})$ for a single product k in a single market c. Scaling by the share of product j and fixing the focal product k and the market c allows us to illustrate the extent to which the three models do or do not allow cross-price effects to vary flexibly across products.

Figure 11 plots exact and approximated scaled cross-price semi-elasticities for each of the three models. The first panel shows scaled semi-elasticities for the standard logit model under the standard logit DGP. The approximate estimator is exactly the same as in the first panel of Figure 4 (mirrored about the y-axis) since the logit estimator uses the same own-price IV estimates $\hat{\rho}_j^{\log(q/q_0)}$ to determine both own- and cross-price effects; cross-price IV estimates play no role in determining the cross-price elasticities. The approximation is exact: the scaled cross-price semi-elasticity $\hat{\eta}_{kj}/(p_j s_j)$ equals $\hat{\alpha}$ exactly.

The second panel of Figure 11 shows scaled semi-elasticities for the product-specific logit model under the product-specific logit DGP. As with the logit model, the approximate estimator is exactly the same as the approximate estimator for the own-price semi-elasticities shown in the second panel of Figure 5 (mirrored about the y-axis); cross-price effects are determined by the same IV estimands as own-price effects. The approximation is exact: the scaled cross-price semi-elasticity $\hat{\eta}_{kj}/(p_j s_j)$ equals $\hat{\rho}_j^{\log(q/q_0)}$ exactly. Unlike the standard logit model, however, the scaled cross-price semi-elasticities vary across products j.

Finally, the third panel of Figure 11 shows scaled semi-elasticities for the random coefficient logit model under the random coefficient logit DGP. The approximate estimator is the own-price approximate estimator in the third panel of Figure 6 (mirrored about the y-axis). While the approximation slightly underestimates the scale of the cross-price effects, it correctly captures how these vary across products j.

6 Implications for Other Target Parameters

In this section, we consider the implications of our approximation results for the widespread use of discrete choice demand models to draw inferences about economic quantities such as marginal costs, markdowns, and pass-through rates. These parameters are of independent economic interest and are also key inputs into the analysis of firm policy counterfactuals and merger simulation.

6.1 Profit Margin, Markup, and Inferred Marginal Costs

We assume each product j is produced by a single-product firm with constant marginal cost c_j and that firms engage in Bertrand-Nash competition in prices. The optimal pricing condition for product j is

$$p_j - c_j = -\frac{p_j}{\eta_{jj}(p)} \equiv \mu_j(p_j), \tag{49}$$

which implicitly defines the profit margin $\mu_j(p_j)$. Re-arranging equation (49) allows marginal costs and markups $((p_j - c_j)/p_j)$ to be recovered from observed prices and own-price elasticities. Inference on markups is a key motivation for demand estimation (Berry et al., 1995).

The following proposition summarizes the implications of the results in Section 3 for marginal costs.

Proposition 16. *Given observed price p_j, the marginal cost c_j implied by each model is*

(a) Logit:

$$c_j \approx p_j - \underbrace{\frac{1}{\alpha}}_{\textit{profit margin}}$$

(b) Product specific logit:

$$c_j \approx p_j - \underbrace{\frac{1}{\alpha_j}}_{\textit{profit margin}}$$

(c) Random coefficient logit (with uncorrelated random coefficients on product characteristics):

$$c_j \approx p_j - \underbrace{\frac{1}{\mathbb{E}[\alpha_i] - \text{Var}[\alpha_i](p_j - \bar{p})}}_{\textit{profit margin}}$$

Remark 1. Under the standard logit model, differences in marginal costs across products are mechanically equal to differences in prices, up to small-share approximation error. This mapping follows directly from the model's structure and does not depend on data, as the model implies a constant profit margin across products and price levels. As a consequence, the markup—defined as the ratio of the profit margin to price—is decreasing in the price level. In our setting, for example, a firm that produces a cheap toothpaste would be assumed to have the same profit margin and hence a higher markup than a firm that produces a high-end toothpaste. In contrast, the product-specific logit model allows marginal costs, profit margins, and markups to vary across firms through variation in product-specific price coefficients.

In the random coefficient logit model, inferred marginal costs are likewise entirely determined by prices, but through a common functional relationship that is inverted-U-shaped in price. This means that, by *assumption*, marginal costs are low and markups are high both for firms that sell low-price toothpaste and those that sell high-price toothpaste. The high markups for the low-price firms reflect that the own-price elasticity of a given consumer type is increasing in price. The high markups for the high-price firms reflect selection of less-sensitive consumers into higher-priced products.

Remark 2. By applying the identification results in Section 4 to Proposition 16, we can determine which features of the data approximately pin down the marginal cost and markup of a given firm. The logit estimator recovers α by taking a weighted average of IV estimands across all products; the product-specific logit model recovers α_j by taking the IV estimand for the particular product; and the random coefficient logit model recovers $\mathbb{E}[\alpha_i]$ and $\text{Var}[\alpha_i]$ by taking a weighted regression of the IV estimands on a constant and average price levels. This means that both the logit and random coefficient model are inferring the marginal cost and markup of a given product from consumer responses to price changes of other products. The product-specific logit model does not extrapolate in this way.

6.2 Pass-through of Cost Shocks

We next consider the pass-through of firm-specific cost shocks. The (absolute) pass-through of a marginal increase in c_j is given by:

$$\frac{dp_j}{dc_j} = \frac{1}{1 - \mu_j'(p_j)}. \tag{50}$$

This expression abstracts from equilibrium feedback through rival firms adjusting prices. Under the small-share assumption, such feedback is negligible, as a unilateral change in p_j has little effect on the prices of competing products. Pass-through is governed by the curvature of demand and serves as a central sufficient statistic in imperfectly competitive markets; it determines how cost or policy shocks translate into prices, markups, and welfare (Weyl and Fabinger, 2013).

Proposition 17. *The passthrough of firm specific shocks to prices according to each model is*

(a) Logit:

$$\frac{1}{1-\mu_j'(p_j)} \approx 1$$

(b) Product specific logit:

$$\frac{1}{1-\mu_j'(p_j)} \approx 1$$

(c) Random coefficient logit (with uncorrelated random coefficients on product characteristics):

$$\frac{1}{1-\mu_j'(p_j)} \approx \frac{1}{1-\frac{\text{Var}[\alpha_i]}{\left(\mathbb{E}[\alpha_i]-\text{Var}[\alpha_i](p_j-\bar{p})\right)^2}} \geq 1$$

Remark 1. Under the standard- and product-specific logit models, pass-through is mechanically equal to one, as own-price elasticities are approximately linear in prices. By assumption, cost shocks are fully borne by consumers. The random coefficient logit model instead directly imposes overshifting: pass-through exceeds one. This property follows from the selection mechanism that governs demand curvature: as a firm raises its price in response to a cost shock, the most price-sensitive consumers substitute away, causing the semi-elasticity of demand to decline linearly.

A further implication of the random coefficient logit model is that pass-through is monotonically increasing in the price level. Higher-priced products therefore exhibit greater overshifting than lower-priced products. In our setting, for example, firms selling low-priced generic toothpaste are assumed to have lower pass-through than firms selling higher-priced branded toothpaste. Intuitively, as price rises and the semi-elasticity falls, profit margins increase, and the rate at which profit margins rise itself increases with price.

While over-shifting is empirically and theoretically plausible in some settings, the assumptions that pass-through must exceed one and that it rises monotonically with price are strong restrictions relative to the range of patterns discussed in the pass-through literature (Fabinger and Weyl, 2018; Miravete et al., 2025).

Remark 2. By applying the identification results in Section 4 to Proposition 17, we can determine which features of the data would approximately pin down pass-through under each model. Under the standard- and product-specific logit models, data play no role in determining the pass-through. The random coefficient logit model identifies the curvature of demand—and hence pass-through—by projecting product-specific IV estimates of demand responses onto average price levels. As a result, the pass-through for a given product is inferred from systematic relationships between demand responses and price levels across products, rather than from variation specific to that product.

7 Interpretation under misspecification

The estimators studied in the previous sections can be evaluated under any data generating process, regardless of whether the motivating model is correctly specified. Because each model restricts heterogeneity in price responses across price levels, markets, and products, it is natural to ask how the resulting estimands should be interpreted under misspecification.

In this section, we interpret the estimands of discrete choice demand models within a general potential outcomes framework. This framework imposes minimal structure on demand, allows for arbitrary heterogeneity across markets, products, and price levels, and nests the models considered in Sections 2–4.

In the first subsection, we show that estimators of own-price elasticities can be interpreted as weighted averages of true price responses, even under misspecification, and we characterize how these weights differ across models. In the second subsection, we derive implications for marginal costs and markups, showing that their estimates retain a meaningful economic interpretation under misspecification only in the logit and product-specific logit models.

We begin by introducing the potential outcomes framework used throughout this section. We consider the following structure. For each product j in market c, let $q_{jc}(p)$ denote the counterfactual quantity demanded under price vector p. Counterfactual demand does not depend on prices in other markets, nor does it directly depend on the instrument z_{jc} (*exclusion*). We assume $q_{jc}(\cdot)$ is differentiable.

The instrument z_{jc} is a J-dimension binary vector, as described in Section 4.2. Let $p_{jc}(0)$ and $p_{jc}(1)$ denote the counterfactual prices of product j in market c under $\dot{z}_{jc} = 0$ and $\dot{z}_{jc} = 1$, respectively. The observed price is $p_{jc} = p_{jc}(\dot{z}_{jc})$, and the observed quantity is $q_{jc} = q_{jc}(p_{1c}(\dot{z}_{1c}), \ldots, p_{Jc}(\dot{z}_{Jc}))$. We impose the standard IV assumptions of exogeneity (z_{jc} is independent of $(p_{jc}(0), p_{jc}(1), q_{jc}(\cdot))$), monotonicity ($p_{jc}(0) \leq p_{jc}(1)$ with probability 1), and relevance ($p_{jc}(0) < p_{jc}(1)$ with positive probability).

7.1 Own-price elasticities under misspecification

Under this framework, an application of Theorem 1 of Angrist et al. (2000) implies

$$\rho_j^{\log(q/q_0)} = \int_0^\infty \mathbb{E}\left[\frac{\partial \log q_{jc}}{\partial p_j}(p_j, p_{-jc}) - \frac{\partial \log q_{0c}}{\partial p_j}(p_j, p_{-jc}) \,\middle|\, p_{jc}(0) \leq p_j \leq p_{jc}(1)\right] \omega_j^p(p_j)\, \mathrm{d}p_j$$

where $\omega_j^p(p_j) \propto \mathbb{P}\{p_{jc}(0) \leq p_j \leq p_{jc}(1)\}$ and $\int_0^\infty \omega_j^p(p_j)\, \mathrm{d}p_j = 1$. Thus, the IV estimand from regressing $\log(q_{jc}/q_{0c})$ on p_{jc} with instrument z_{jc} (using only observations for product j) recovers a positively-weighted average of own-price semi-elasticities. Building on this result and those in Section 4, we show below that the logit and product specific logit estimators produce convex combinations of the true semi-elasticities, even if the models are misspecified. By contrast, misspecification of the random coefficient logit estimator yields an affine combination of semi-elasticities, the weights of which may vary in sign.

Proposition 18. *Let z_{jc} denote the J-dimensional instrument defined in Section 4.2. As*

$C \to \infty$ and $s_{j'c'} \to 0$, the predicted own-price elasticity of product j' in market c' is

$$\begin{aligned}
&\operatorname*{plim}_{C\to\infty} \hat{\eta}_{j'j'c'}(p_{c'}) \\
&= -p_{j'c'} \sum_{j=1}^{J} \Psi_{j'c'}(j, p_{j'c'}) \\
&\qquad \cdot \int_0^\infty \mathbb{E}\left[\frac{\partial \log q_{jc}}{\partial p_j}(p_j, p_{-jc}) - \frac{\partial \log q_{0c}}{\partial p_j}(p_j, p_{-jc}) \,\middle|\, p_{jc}(0) \le p_j \le p_{jc}(1) \right] \omega_j^p(p_j)\, \mathrm{d}p_j \\
&\qquad + R_{j'c'},
\end{aligned}$$

where the weights $\Psi_{j'c'}(j)$ are estimator-specific and satisfy $\sum_{j=1}^{J} \Psi_{j'c'}(j, p_{j'c'}) = 1$ and $R_{j'c'}$ collects remainder terms. The weights and remainders for each estimator are

(a) Logit:

$$\Psi_{j'c'}(j) = \omega_j, \qquad R_{j'c'} = O(s_{j'c'}).$$

(b) Product specific logit:

$$\Psi_{j'c'}(j) = \mathbb{1}\{j = j'\}, \qquad R_{j'c'} = O(s_{j'c'}).$$

(c) Random coefficient logit:

$$\begin{aligned}
\Psi_{j'c'}(j, p_{j'c'}) &= \omega_j \left(1 + \frac{(p_{j'c'} - \overline{p}_{c'} - \mu_\omega)\,(p_{jc} - \overline{p}_c - \mu_\omega)}{\sigma_\omega^2} \right), \\
R_{j'c'} &= O(|\lambda|^3) + O(s_{j'c'}),
\end{aligned}$$

where

$$\mu_\omega = \sum_j \omega_j (p_{jc} - \overline{p}_c), \qquad \sigma_\omega^2 = \sum_j \omega_j (p_{jc} - \overline{p}_c - \mu_\omega)^2.$$

Proof. The results for (a) and (b) directly follow by substituting the Angrist et al. (2000) result into equations (27) and (29) respectively. For (c), see Appendix A.15. □

Even under misspecification, the estimators studied in Section 4 can still be computed in the data. Proposition 18 shows that their elasticities admit interpretation in terms of averages of potential outcomes.

Proposition 18(a) shows that the logit semi-elasticity involves three level of averaging of semi-elasticities $\frac{\partial \log q_{jc}}{\partial p_j}(p_j, p_{-jc}) - \frac{\partial \log q_{0c}}{\partial p_j}(p_j, p_{-jc})$ specific to a particular product j, price level p_j, and market c. The first level of averaging (the conditional expectation) is over markets for the given product and price level. The second level of averaging (the integral) is over price levels for a given product. The third and final level of averaging (the sum) is over products. Each level of averaging involves non-negative weights that sum to 1, so that the logit semi-elasticity will always lie in the convex hull of the product-, price-, and market-specific semi-elasticities $\frac{\partial \log q_{jc}}{\partial p_j}(p_j, p_{-jc}) - \frac{\partial \log q_{0c}}{\partial p_j}(p_j, p_{-jc})$.

Proposition 18(b) shows that the product-specific logit estimator averages over markets and over price levels in the same way as the logit estimator. However, the product-specific logit estimator involves no averaging over products: the estimator for the semi-elasticity of product j' only puts positive weight on semi-elasticities for product j'.

Proposition 18(c) shows that the random coefficient logit estimator again averages across markets and price levels in the same way as the logit estimator. However, the random coefficient logit estimator also *extrapolates* across products using weights $\Psi_{j'c'}(j, p_{j'c'})$ that vary with the price level $p_{j'c'}$. These weights sum to 1, but may be negative. This extrapolation is related to the weighted least squares regression (33) implicitly run by the BLP estimator. The BLP estimator fits a line through semi-elasticities for different products at different price levels, and uses this line to extrapolate semi-elasticities at other price levels. There will typically exist values of $p_{j'c'}$ such that some product weights $\Psi_{j'c'}(j, p_{j'c'})$ are negative. More precisely, if product j has prices above the market average when the focal product's price $p_{j'c'}$ is below its market average (or vice-versa), then the deviations have opposite signs producing a negative term within the weight $\Psi_{j'c'}$. If that term is large enough, the weight can become negative. Thus the random coefficient logit semi-elasticity may be outside the range of semi-elasticities of any product in any market.

7.2 Profit Margins and Marginal Costs under Misspecification

We now consider the implications of misspecification for estimating marginal costs and profit margins. Given price p_{jc} and a predicted own-price semi-elasticity $\hat{\eta}_{jjc}(p_c)$, the implied profit margin is

$$\hat{\mu}_{jjc}(p_c) = \frac{p_{jc}}{\hat{\eta}_{jjc}(p_c)}.$$

Because this expression is nonlinear in $\hat{\eta}_{jjc}(p_c)$, the mapping from elasticity estimates to profit margins may distort average relationships under misspecification. To characterize these relationships, we define the following weighted averages of true profit margins:

$$\bar{\mu}_j \equiv \int_0^\infty \mathbb{E}\left[\mu_{jc}(p_j, p_{-jc}) \,|\, p_{jc}(0) \leq p_j \leq p_{jc}(1)\right] \omega_j^p(p_j)\, dp_j,$$
$$\bar{\mu} \equiv \sum_j \omega_j \bar{\mu}_j.$$

Although these weights are statistical objects shaped by the instrument used in estimation, they may still have direct economic relevance. For instance, in a field experiment, the instrument may induce price variation over a support that reflects the firm's decision margin or a relevant policy counterfactual.

The next result shows that, under misspecification, the logit and product-specific logit estimators provide lower bounds on these average margins, while the random coefficient logit estimator does not.

Corollary 7. *Maintain the assumptions of Proposition 18. Further assume that demand is generated by a RUM in which consumers solve problem (12), and that market shares are small.*

(i) The product-specific logit estimator satisfies:

$$\hat{\mu}_j^{PS\ logit} \equiv \frac{1}{\hat{\alpha}_j} \xrightarrow{p} \frac{1}{\int_0^\infty \mathbb{E}\left[-\frac{\partial \log q_{jc}}{\partial p_j}(p_j, p_{-jc}) + O(s_{jc}) \,\middle|\, p_{jc}(0) \le p_j \le p_{jc}(1)\right] \omega_j^p(p_j)\, \mathrm{d}p_j}$$
$$\approx \frac{1}{\int_0^\infty \mathbb{E}\left[-\frac{\partial \log q_{jc}}{\partial p_j}(p_j, p_{-jc}) \,\middle|\, p_{jc}(0) \le p_j \le p_{jc}(1)\right] \omega_j^p(p_j)\, \mathrm{d}p_j} \le \bar{\mu}_j$$

(ii) The standard logit estimator satisfies:

$$\hat{\mu}^{logit} \equiv \frac{1}{\hat{\alpha}} \xrightarrow{p} \frac{1}{\sum_j \omega_j \int_0^\infty \mathbb{E}\left[-\frac{\partial \log q_{jc}}{\partial p_j}(p_j, p_{-jc}) + O(s_{jc}) \,\middle|\, p_{jc}(0) \le p_j \le p_{jc}(1)\right] \omega_j^p(p_j)\, \mathrm{d}p_j}$$
$$\approx \frac{1}{\sum_j \omega_j \int_0^\infty \mathbb{E}\left[-\frac{\partial \log q_{jc}}{\partial p_j}(p_j, p_{-jc}) \,\middle|\, p_{jc}(0) \le p_j \le p_{jc}(1)\right] \omega_j^p(p_j)\, \mathrm{d}p_j} \le \sum_j \omega_j \hat{\mu}_j^{PS\ logit}$$

(iii) In particular,

$$\hat{\mu}^{logit} \le \sum_j \omega_j \hat{\mu}_j^{PS\ logit} \le \bar{\mu},$$

Hence, both $\hat{\mu}^{logit}$ and $\sum_j \omega_j \hat{\mu}_j^{PS\ logit}$ provide lower bounds on weighted averages of true profit margins ($\bar{\mu}$). The product-specific logit yields a sharper bound by avoiding averaging across products.

Proof. Follows from Proposition 18 and Jensen's inequality. □

Remark. These bounds follow from Jensen's inequality. The logit and product-specific logit estimators recover convex combinations of semi-elasticities. The weighted average markup is a weighted average of inverse semi-elasticities, but the estimators instead estimate the inverse of a weighted average of semi-elasticities. Even when the "right" weights are attained, the fact that the profit margin is a convex function of the semi-elasticity means that the resulting estimate of average margin will be downward biased, and the corresponding estimate of average marginal cost will be upward biased. The bias is larger for the standard than the product-specific logit because the standard logit also averages across products, not just across markets and price levels.

In comparison, the random coefficient logit estimator extrapolates using affine combinations of elasticities, with weights that may be negative. This lack of convexity means the implied profit margins can lie outside the support of true margins and need not be biased in any particular direction. Consequently, no general lower bound analogous to Corollary 7 applies in that case.

8 Application to a pricing experiment

We now apply our theoretical results to a large-scale randomized pricing experiment, in which we find that the approximation error is negligible and that own-price responses are grossly at odds with the elasticities from the standard and random coefficient logit model. The estimates also suggest that the product-specific logit model provides sufficient flexibility in own-price responses across products to rationalize the experimental estimates.

8.1 Context and Data

We work with data from a retailer operating thousands of brick-and-mortar store locations across all 50 US states. The retailer sells both private label and national brand products with an assortment that includes food, consumables, and general merchandise, among other segments.

The retailer's pricing is decentralized across category managers. Each manager has broad discretion over the pricing strategy for their merchandise, subject to meeting quarterly and annual category-level financial targets and following a basic set of pricing rules.[11] There is no consensus approach across category managers on what estimates of price elasticities of demand to use in price setting. The retailer purchases own-price elasticity estimates for some products from third-party vendors. These estimates are based on transaction data collected by the vendor from several retailers across the country, but exact details of the methodology are not disclosed. Most pricing managers do not reference these estimates in setting prices, choosing instead to rely on their own judgment, past experience, and intuition.

The retailer groups unique items it sells (identified by UPC numbers) into "lines", sets of items that must have the same price within each store at every point in time. Examples of lines include all color variants of the same towel and all flavor variants of the same size and brand of yogurt. In the experiment, we assign prices at the line level.

Some lines are also organized into ordered collections called "ladders"; examples include products corresponding to different sizes of the same toothpaste. Higher products in a ladder must have higher prices. Additionally, for ladders that have an associated unit of measure, higher products in the ladder must have a lower price per unit than lower products in the ladder (e.g., a 6 oz tube of toothpaste must be less than twice as expensive as a 3 oz tube of the same toothpaste).

8.2 Experimental Design

Our instruments comes from an exogenous pricing test run by the retailer for nine weeks in late 2022. In the terminology of Harrison and List (2004), the test was a natural field experiment: an "experiment" in that it involved explicit randomization of prices, "field" in that it took place in the retail environment in which customers make their regular purchasing decisions, and "natural" in that customers were not aware that the price changes were part of an experiment.

[11] Examples of such rules include that larger-sized variants of a product must be priced above smaller-sized variants and that all products must be priced above their wholesale supplier cost.

For most products j in the test, three price point variants were constructed. First, a middle price variant z_j^2 was set, typically to the modal price of the product across stores prior to the start of the test. Next, low and high price variants z_j^1 and z_j^3 were set symmetrically (on a log scale) about the middle price. The difference between the middle and the low/high prices was chosen to be as large as possible subject to the following constraints:

- The difference between the middle and low/high prices can be no more than 0.15 log points.
- The difference between the middle and low/high prices must be at least \$0.05.
- Every price must be above the product's supplier cost and, if applicable, the supplier-negotiated minimum advertised price.
- Ladder constraints must be satisfied for any choice of price variants for each product in the ladder.

For some products, no amount of positive price variation could satisfy all of these constraints simultaneously. Such products were excluded from the test.

The test ran in 4,512 stores across 46 US states. These stores were grouped into 2,236 store clusters, designed to group together stores that customers frequently shop between. During the test, stores in the same cluster receive the same prices for every product, both to avoid negative customer experiences associated with seeing disparate prices across the retailer's stores and to limit the ability of customers to respond to price changes by shifting their demand between stores.

For each product, one-third of all store clusters carrying the product were randomly assigned to receive either the low, middle, or high price variant for the product.[12] Assignment was independent across products. Each store was sent a list of its assigned prices for products in the test to update its price stickers and scanner system. For products in the test, online shoppers saw prices corresponding to their selected store location. The retailer implemented operational locks to prevent other price changes on these products for the duration of the test.[13]

The design of the price variation—which is across markets within each product and independent across products—holds approximately fixed both the average price level of each store and the national average price level of each product. Because the average price level of each store is fixed, the experiment measures the intensive margin of demand (the choice of which product to buy in one's chosen store), not the extensive margin (the choice of which store to shop). Because the national average price level of each product is fixed, the experiment minimizes the probability of price responses from competing retailers, which typically price nationally (DellaVigna and Gentzkow, 2019).

[12] Randomization was stratified by geographic region and tercile of pre-period unit sales.

[13] Certain markdown/clearance sales were permitted to override the assigned prices, generating some non-compliance with assigned prices. Such sales were planned prior to the start of the test.

8.3 Choice of category and sample construction

The test included a total of 1,884 products across 23 categories collectively accounting for close to 2% of the retailer's total revenue. Since the discrete choice models we consider require that all products be substitutes with one another, we restrict focus to a single category from the field experiment: toothpaste. We chose toothpaste because it is the only category satisfying three desirable criteria:

- It is relatively straightforward to define and measure relevant product characteristics.
- It has a plausible "total market size," a required input for the discrete choice models we consider.
- It has close to complete sales coverage in the field experiment.

In our application, a "market" is a single retail store location over the course of the experiment. We estimate total market sizes as follows. First, we use the retailer's data to estimate the average number of distinct weekly households shopping at each store location. Next, we multiply this figure by PUMA estimates of individuals per household to estimate total population shopping at each store. Finally, we multiply by an estimate of per-capita toothpaste consumption, assuming an upper bound of three brushings per person per day and 0.5 grams of toothpaste per brushing (Creeth et al., 2013) to estimate total demand for toothpaste.

In our analysis, a product is a single toothpaste UPC. Our product characteristic space consists of indicators for the two largest toothpaste brands and an indicator for the presence of hydrogen peroxide (used in whitening toothpastes). We also include income as a demographic variable in some specifications of the random coefficient logit model. The distribution of household income is assumed to follow that of the Public Use Microdata Area (PUMA) in which the store is located.

Our sample consists of prices, shares, and characteristics of 190 products (UPCs) across 4484 markets (stores). The products are grouped into 64 lines while markets are grouped into 2235 clusters. The field experiment varies prices at the line $\times$ cluster level; all standard errors are clustered at this level. Table 2 provides summary statistics of the sample. The median toothpaste line is carried in 4242 stores and has a price of \$5.97. The first-stage F statistics from regressions of price on assigned price are high by conventional metrics. Table 3 shows that the experimental price assignments are uncorrelated with units sold in the pre-period.

8.4 Empirical specifications and instruments

We consider three empirical specifications, corresponding to the three special cases of (1) analyzed in the preceding sections.

The logit specification of the indirect utilities is

$$v_{ij} = \gamma_j - \alpha p_{jc} + \xi_{jc} + \varepsilon_{ij}$$

where γ_j is a product fixed effect and ξ_{jc} is an unobserved demand shifter.

Table 2: 10th, 50th, and 90th percentiles of line-level statistics (calculated across store-weeks) for toothpaste products.

	10th Percentile	50th Percentile	90th Percentile
Pre-test Period			
% Store-weeks with Sales	10.09	59.33	97.07
Avg. Units Sold	0.41	3.32	15.51
Avg. Price	2.47	4.95	10.17
Std. Dev. of Log Price	0.03	0.06	0.12
R^2 of Log Price on Store FEs	0.01	0.07	0.47
R^2 of Log Price on Week FEs	0.00	0.64	0.94
Test Period			
Number of Stores Carrying	2117.40	4242.50	4494.00
Median Price	2.87	5.97	10.50
Median Assigned Price	2.96	5.97	10.50
Std. Dev. of Log Price	0.04	0.12	0.15
Std. Dev. of Log Assigned Price	0.04	0.12	0.12
Corr. of Assigned Price and Pre-Period Units Sold	-0.04	0.00	0.04
Corr. of Assigned Price and Price	0.57	0.92	0.99
First-Stage F	2050.05	14810.52	270750.15

Table 3: Balance test for toothpaste products with three-point price variation.

	Pre-test Units Sold	Total Market Size (1000s oz.)
	(1)	(2)
Intercept	110.324	152.642
	(1.583)	(0.428)
Variant 2	-0.971	0.050
	(2.162)	(0.596)
Variant 3	-1.394	0.559
	(2.154)	(0.604)
Observations	224048	224048
R^2	0.000	0.000
F	0.216	0.529

Note: Each observation is a store-line pair. Clustered standard errors in parentheses.

The product-specific logit specification is

$$v_{ij} = \gamma_j - \alpha_j p_{jc} + \xi_{jc} + \varepsilon_{ij},$$

allowing the price sensitivity parameter to vary by product.

Finally, the random coefficient logit specification is

$$v_{ij} = \gamma_j - \alpha_i p_{jc} + \beta_i' x_j + \xi_{jc} + \varepsilon_{ij}, \qquad \begin{pmatrix}\alpha_i \\ \beta_i\end{pmatrix} = \begin{pmatrix}\mu_\alpha \\ 0\end{pmatrix} + \Lambda\nu_i + \Pi d_{ic},$$

where ν_i is a draw from a standard multivariate normal distribution and d_{ic} is a draw from the income distribution in market c. Income d_{ic} is standardized to have a mean of 0 and a standard deviation of 1 across markets. We always restrict Λ to be a nonnegative diagonal matrix. We also consider more restricted versions of the specification that restrict $\Pi = 0$ and that only include a random coefficient on price.

Let $\ell(j) \in \{1, \ldots, L\}$ denote the product line to which product j belongs. We leverage the pricing experiment to construct an L-dimensional instrument z_{jc} with[14]

$$z_{jc\ell} = \begin{cases} \text{assigned price for product line } \ell \text{ in market } c & \ell(j) = k \\ 0 & \text{else.} \end{cases}$$

In some specifications, we also include the following instruments to recover the parameters Λ and Π:

- Quadratic differentiation instruments based on the vector of product characteristics x_j (following Gandhi and Houde (2019)).
- The average income $\bar{d}_c$ in market c and its interaction with the product characteristics x_j (following Miller and Weinberg (2017)).

We do not use these instruments when estimating the standard logit model, the product-specific logit model, or the price-only random coefficient logit model. However, including the additional instruments in these specifications have virtually no impact on our estimates.

8.5 Linear IV estimates of own-price responses

As shown in Section 4, estimands of the discrete choice models can be written as combinations of product-specific IV estimands. Before estimating the models, we begin by studying these product-specific IV estimates. For each product j, we regress

$$\log(s_{jc}/s_{0c}) = \gamma_j - \alpha_j p_j + \xi_{jc}$$

using $(1, z_{jc\ell(j)})'$ as an instrument, obtaining estimates $\{\hat{\alpha}_j\}_{j=1}^J$. For ease of interpretation, we multiply these estimates by each product's average price level to obtain an approximate elasticity.

[14] Note that $z_{jc\ell}$ combines the price increase and price decrease for product line ℓ into a single instrument. This combination smooths over concerns that price elasticities may differ between increases and decreases due to left-digit/integer-notch effects (DellaVigna, 2009; Sokolova et al., 2020) when the price increases crosses an integer price and the decrease does not (or vice versa).

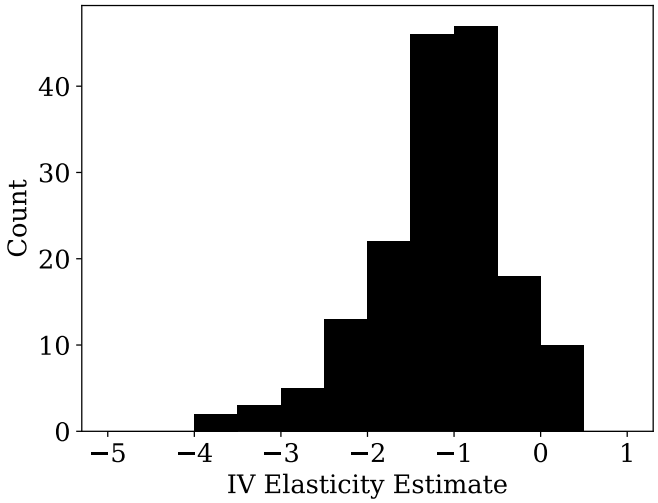


Figure 12: Distribution of product-specific linear IV elasticity estimates.

Figure 12 plots the marginal distribution of the elasticity estimates. 93.4% are negative, with 79.6% falling between −2 and 0. None of the positive point estimates are significantly different from 0 at the 10% level. The estimates are also fairly precise, with a median standard error of 0.22.

8.6 Standard and random coefficient logit estimates

We estimate the logit and random coefficient logit models using PyBLP (Conlon and Gortmaker, 2020). Table 4 presents the estimated parameters for an array of specifications.

Specification (1) is a standard logit model with no random coefficients and no demographic interactions. The price coefficient is negative and precisely estimated.

Specification (2) makes the coefficient on price a random variable. The mean price coefficient is more negative for random coefficient than standard logit, as is usually found in applications. The estimated standard deviation of the price coefficient is nearly half the mean, consistent with substantial dispersion in price sensitivity across the population.

Figure 13 plots the marginal distributions of own- and cross-price elasticities for specifications (1) and (2) in Table 4. Elasticities follow the typical patterns: the random coefficient logit model has more negative own-price elasticities and more dispersed cross-price elasticities.

Specification (3) adds random coefficients on non-price coefficients. The estimated mean and variance of the price random coefficient increase slightly relative to specification (2). The estimated variances of random coefficients on other characteristics are large and significant, suggesting that consumers tend to substitute within their preferred brand and that consumers who purchase peroxide toothpaste are more likely to substitute to other peroxide toothpastes.

Finally, specification (4) adds interactions between income and product characteristics. The interaction parameters Π are precisely estimated; their signs suggest that higher-income consumers are less price sensitive, favor Brand 1 less and Brand 2 more, and have less taste for peroxide compared to lower-income consumers.

Table 4: Logit and random coefficient logit parameter estimates.

	(1)	(2)		(3)		(4)		
	μ	μ	Λ	μ	Λ	μ	Λ	Π
Price ($/ounce)	-0.648	-1.331	0.645	-1.480	0.788	-1.533	0.740	0.293
	(0.012)	(0.038)	(0.019)	(0.057)	(0.031)	(0.063)	(0.039)	(0.009)
Brand 1	—	—	—	—	1.771	—	0.000	-0.119
					(0.591)		(4.657)	(0.014)
Brand 2	—	—	—	—	3.907	—	3.322	0.420
					(1.112)		(1.243)	(0.173)
Peroxide	—	—	—	—	2.577	—	3.359	-0.170
					(0.950)		(1.417)	(0.079)
Covariates								
Product fixed effects	✓	✓		✓		✓		
Instruments								
Experiment	✓	✓		✓		✓		
Differentiation				✓		✓		
Demographic						✓		

Note: Standard errors clustered by store cluster × product line in parentheses. μ parameters for non-price characteristics are absorbed by product fixed effects.
(1): Standard logit (no random coefficients or demographics). (2): Random coefficient on price only. (3): Random coefficients on price and non-price characteristics. (4): Random coefficients on price and non-price characteristics and income interactions.
Brand 1 and 2 are binary indicators for the two largest toothpaste manufacturers. Peroxide is a binary indicator for the presence of hydrogen peroxide, an ingredient often used in whitening toothpaste.
Experiment instruments are high-price indicators from the price experiment interacted with product fixed effects. Differentiation instruments are the Gandhi and Houde (2019) quadratic differentiation instruments based on the product characteristics. Demographic instruments are the average income in each market $\bar{d}_c$ and the average income interacted with product characteristics, following Miller and Weinberg (2017).

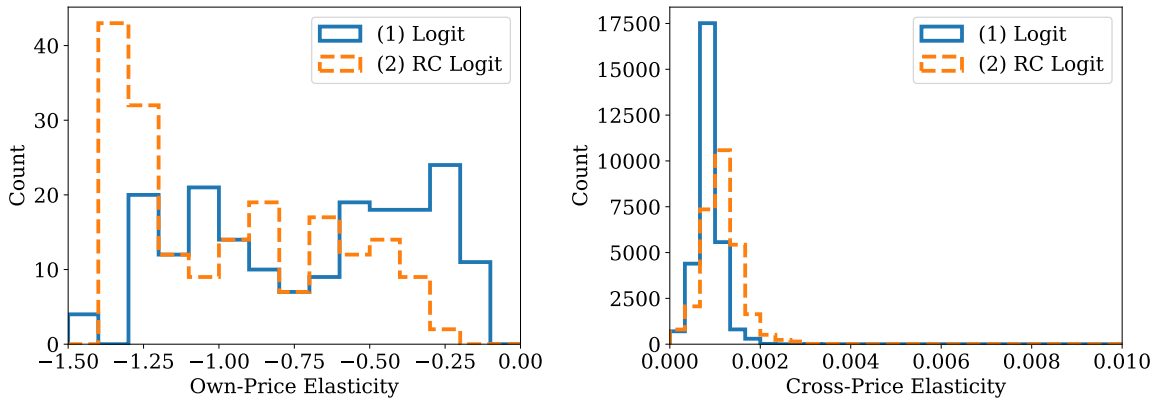


Figure 13: Marginal distributions of average own- and cross-price elasticities for standard and random coefficient logit models estimated on experiment data.

8.7 The anatomy of price sensitivity parameters

8.7.1 Standard logit

The top-left panel of Figure 14 illustrates the weighted averaging of product-specific IV estimates implicitly performed by the logit estimator. Points in green are the IV semi-elasticity estimates for each product j in the sample. The size of each dot is proportional to the weight given to each IV estimate by the logit estimator. The dashed line is the weighted average of these estimated semi-elasticities. Points in blue are semi-elasticities from the estimated logit model. The logit semi-elasticities lie almost exactly along the weighted average line, illustrating how the logit estimator collapses variation in the underlying data moments.

To underscore the extent to which the logit model collapses variation, we plot in the top right panel of Figure 14 the 95% confidence intervals of the semi-elasticity for each product against the logit $-\hat{\alpha}$ estimate. Only 31.7% of confidence intervals contain $-\hat{\alpha}$, suggesting that the single-parameter logit model cannot adequately capture the extent to which price sensitivities vary across products.

The middle panels of Figure 14 plot own-price elasticities against product average price levels (left) and IV elasticity estimates (right) for each product j. Logit own-price elasticities closely follow the linear pattern suggested by Proposition 2. The elasticities show little correlation with the underlying IV estimates.

Finally, the bottom panel of Figure 14 plots scaled cross-price elasticities against product price levels. The scaled cross-price elasticities are η_{1j}/s_j, the elasticity of the share of product 1 with respect to the price of product j divided by the share of product j. Consistent with Proposition 7, the scaled cross-price elasticities are linearly increasing in the price level of product j, and the slope of the relationship is close to that of the relationship between own-price elasticities and price levels.

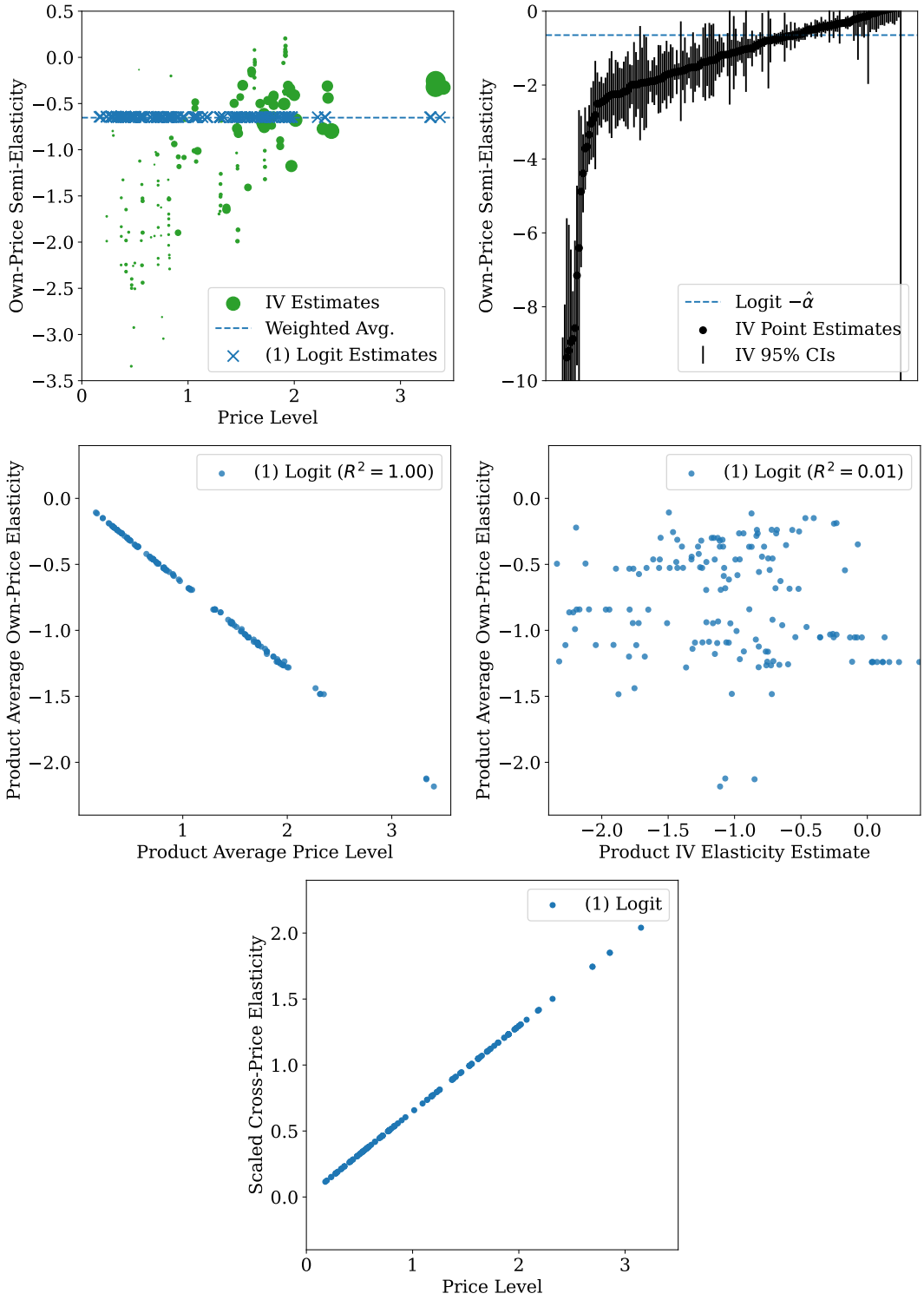


Figure 14: Top left: IV and logit semi-elasticity estimates against product price levels. Top right: Confidence intervals for product semi-elasticities against logit $-\hat{\alpha}$ estimate. Middle left: Average own-price elasticity against average price level for each product j. Middle right: Average own-price elasticity against IV elasticity estimate for each product j. Bottom: Cross-price elasticity against price level. R^2 values are from OLS regressions of own-price elasticity on price and price squared (middle left) and own-price elasticity on IV estimate (middle right).

8.7.2 Product-specific logit

We now shift attention to the product-specific logit model. The top panel of Figure 15 illustrates the identification of this model. Because the model is just-identified, it is able to exactly fit each IV moment.

The middle panels of Figure 15 shows that the product-specific logit model can perfectly fit variation in price elasticities across products. Unlike the other models, elasticities from the product-specific logit model show no clear pattern with respect to price levels. In contrast, the product-specific logit model predicts that products that are more complementary to outside products should be more elastic. To investigate this prediction, we look at IV estimates of each toothpaste's cross-price elasticities with mouthwash products. Figure 16 plots own-price elasticity estimates against average cross-price estimates, together with the OLS regression line. Own-price elasticities are positively correlated with average cross-price elasticities, consistent with the prediction of the product-specific logit model: toothpaste products that are more complementary to mouthwash are also more elastic.

8.7.3 Random coefficient logit

We finally turn to the random coefficient logit model. We focus here on specification (2) in Table 4 (random coefficient on price only), leaving discussion of specifications (3) and (4) to the next subsection. The top left panel of Figure 17 illustrates the weighted least squares regression of product-specific IV estimates onto price levels implicitly performed by the BLP estimator. Again, points in green are the IV semi-elasticity estimates for each product j in the sample. The dashed line is the weighted least squares regression line of these semi-elasticity estimates onto a constant and price levels. Points in orange are semi-elasticities from the estimated random coefficient logit model. These semi-elasticities lie almost exactly along the WLS regression line, illustrating how the random coefficient logit estimator collapses variation in the underlying data moments.

If the model were estimated using only a select subset of instruments, the estimated variance of the random coefficient on price would approximately equal a WLS regression run on the corresponding subset of semi-elasticity estimates. As Figure 17 suggests, such subset regressions could yield a wide range of estimates. We plot in the top right panel of Figure 17 the distribution of estimates that can be obtained by using every possible subset of two products.

The middle panels of Figure 17 plot own-price elasticities against product average price levels (left) and IV elasticity estimates (right) for each product j. Again, own-price elasticities closely follow the quadratic pattern suggested by Proposition 2 but do not correlate with the underlying IV estimates.

Finally, the bottom panel of Figure 17 plots scaled cross-price elasticities against product price levels. Consistent with Proposition 7, the scaled cross-price elasticities are quadratic in the price level of product j.

8.8 Comparing random coefficient logit specifications

We conclude by considering how the addition of non-price random coefficients and demographic interactions affects the own-price elasticities estimated by the random coefficient

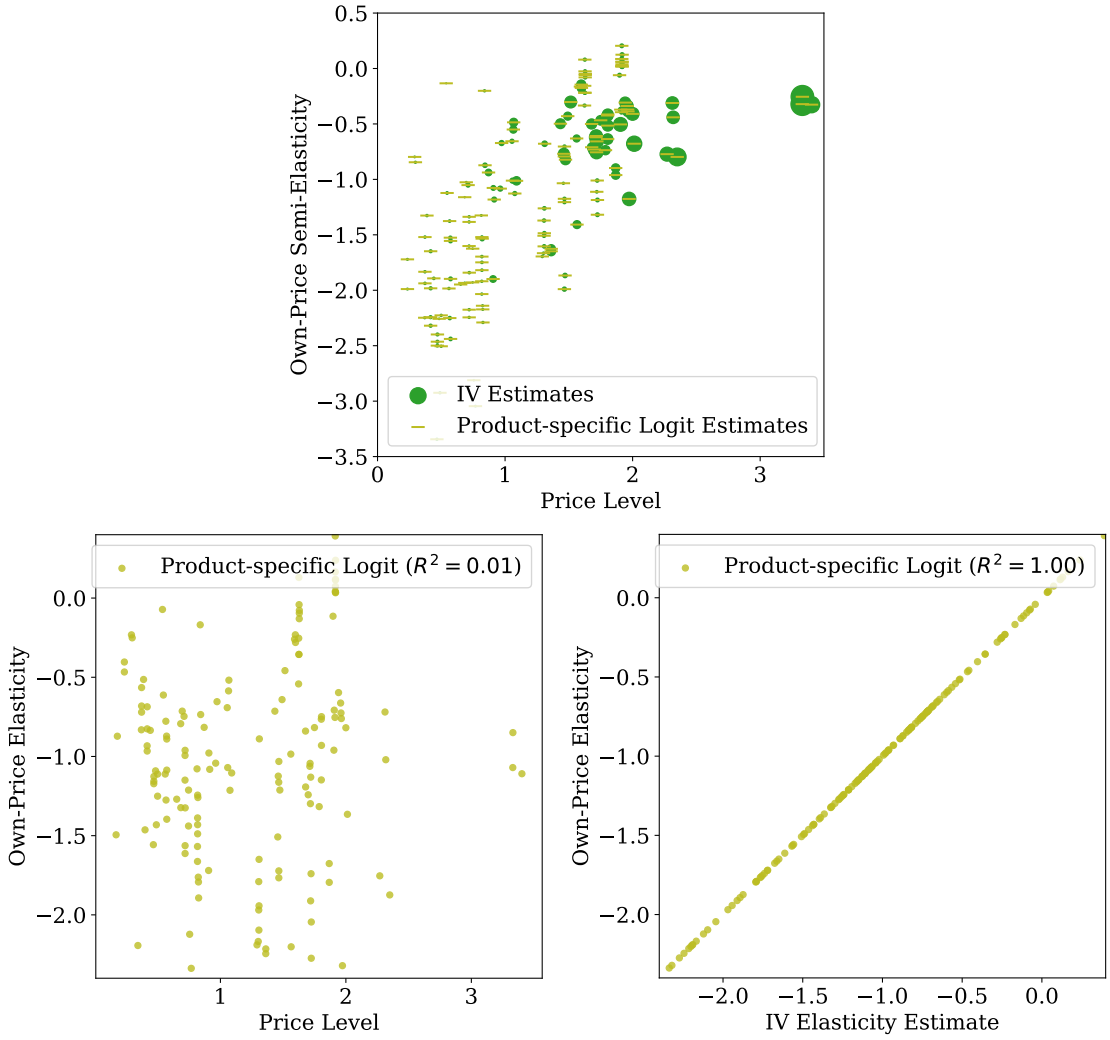


Figure 15: Top: IV and product-specific logit semi-elasticity estimates against product price levels. Bottom left: Average own-price elasticities against average price level for each product j. Bottom right: Average own-price elasticities against IV elasticity estimates for each product j.

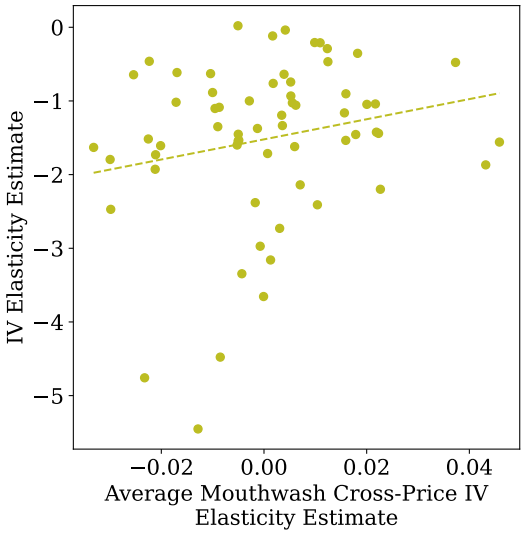


Figure 16: Scatter plot and OLS regression line of own-price IV elasticity estimates for each toothpaste product against average cross-price IV elasticity estimate for each toothpaste product with mouthwash products.

logit model.

8.8.1 Multiple random coefficients

We consider first specification (3) in Table 4, which adds random coefficients on non-price characteristics. The left panel of Figure 18 illustrates the weighted least squares regression of product-specific IV estimates onto price levels implicitly performed by the BLP estimator for this specification. Semi-elasticities again lie close to the WLS regression line, though deviations from the line are larger for this specification than for specification (2).

One might think that the model fits separate WLS regression lines for different groups of products with different characteristics, but this is not the case when the random coefficients on characteristics are uncorrelated with the random coefficient on price. We illustrate this point in the right panel of Figure 18. We estimate separate WLS regression lines for each product group. The regression lines differ significantly across the different product groups, yet the semi-elasticity estimates all approximately lie on a single line.

Turning to elasticities, the left panel of Figure 19 shows that the relationship between own-price elasticities and price levels remains quadratic within groups of products with the same characteristics. Differences in the level of the quadratic reflect differences in the accuracy of the small-share approximation.

The situation is different for cross-price elasticities. The right panel of Figure 19 plots the scaled cross-price elasticities of a single product j with respect to the prices of various products k, grouping by the brand of product k. Like the own-price elasticities, the scaled cross-price elasticities are quadratic in prices conditional on product characteristics. Unlike the own-price elasticities, however, the cross-price elasticities differ greatly across product characteristics: cross-price elasticities are much larger within brand than between brands. These patterns are consistent with the result of Proposition 7 that similarity in product

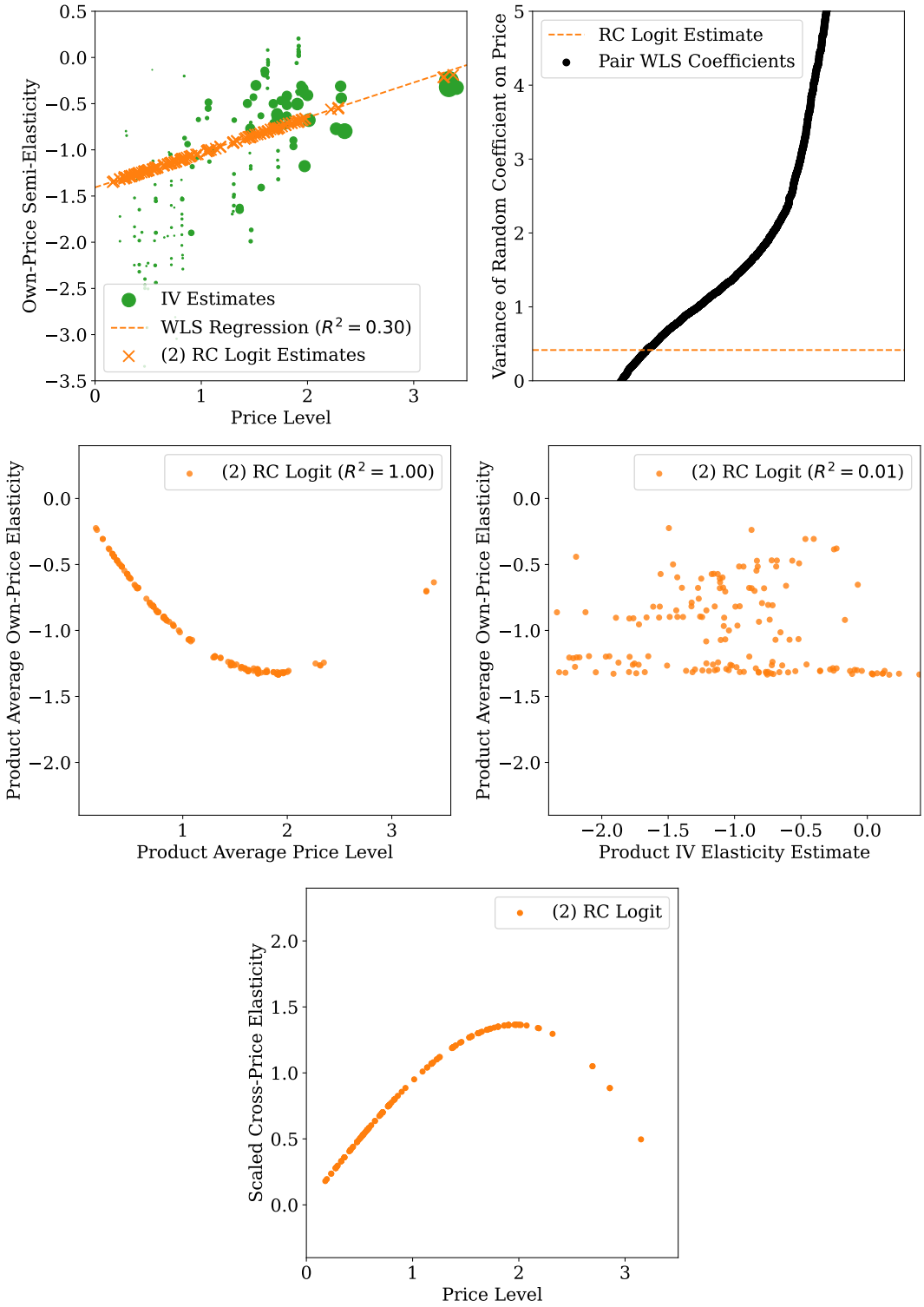


Figure 17: Top left: IV and random coefficient logit (specification (2)) semi-elasticity estimates against product price levels. Top right: Range of WLS coefficients obtainable from regressions using a subset of two products. Middle left: Average own-price elasticity against average price level for each product j. Middle right: Average own-price elasticity against IV elasticity estimate for each product j. Bottom: Scaled cross-price elasticity against price level for each product j. R^2 values are from OLS regressions of own-price elasticity on price and price squared (middle left) and own-price elasticity on IV estimate (middle right).

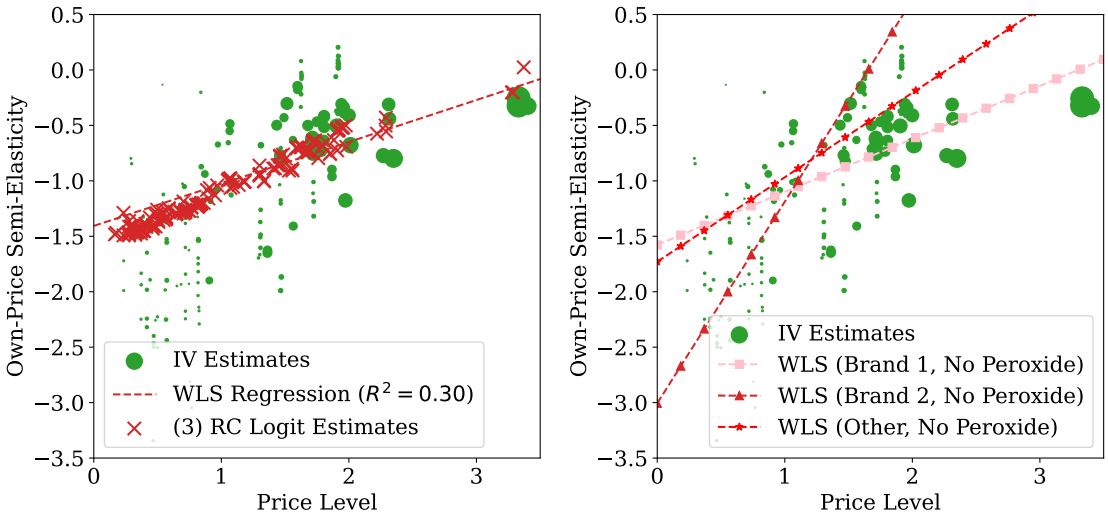


Figure 18: Left: IV and random coefficient logit (specification (3)) semi-elasticity estimates against product price levels. Right: WLS regression lines estimated separately on each product group.

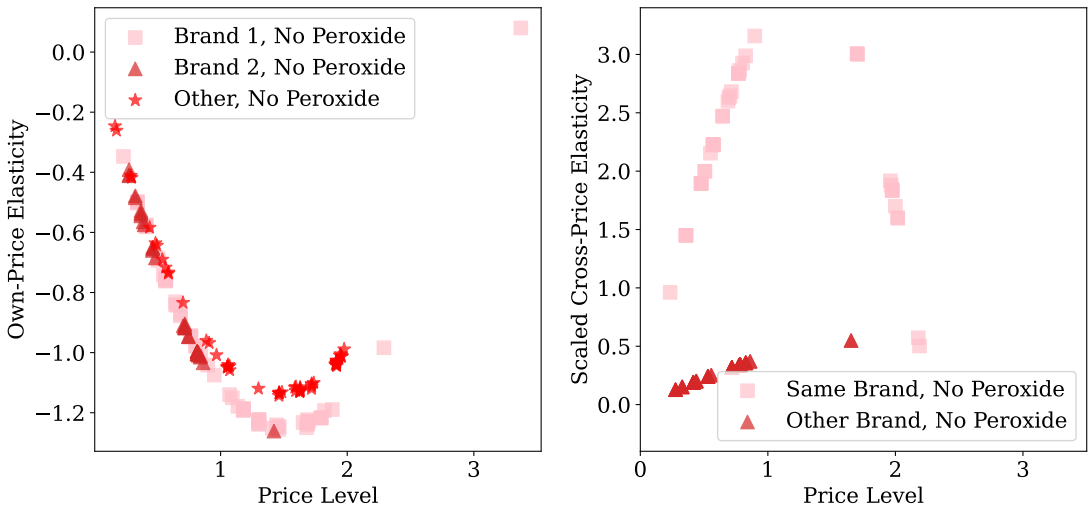


Figure 19: Left: Own-price elasticities from the random coefficient logit model (specification (3)), grouping products by characteristics. Right: Scaled cross-price elasticities for a single product with respect to prices of other products, grouping by brand of the other product.

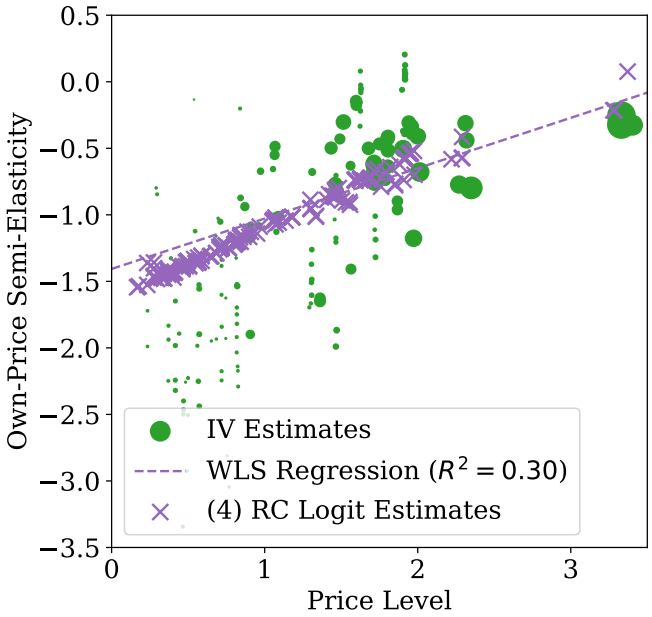


Figure 20: IV and random coefficient logit (specification (4)) semi-elasticity estimates against product price levels.

characteristics shifts up the quadratic relationship between cross-price elasticities and price levels.

8.8.2 Demographics

We consider next specification (4) in Table 4, which allows preferences to vary by income. Figure 20 illustrates the weighted least squares regression of product-specific IV estimates onto price levels implicitly performed by the BLP estimator for this specification. Once again, the approximate regression summarizes almost all of the variation in average own-price semi-elasticity estimates across products.

Turning to elasticities, the left panel of Figure 21 shows that the relationship between elasticities and price levels remains quadratic within each product group. In the random coefficient logit model with demographics, elasticities also vary across markets with the distribution of demographics. The right panel of Figure 21 compares own-price elasticities in a single low-income market to those in a medium-income market. Because the income interaction on price sensitivity is estimated to be positive, the quadratic is shifted down in the low-income market relative to the medium-income market.

9 Conclusion and takeaways for practitioners

We consider the identification and interpretation of own-price elasticities in a class of discrete choice demand models. Because nonparameteric estimation and inference of such models remains challenging with many products (Compiani, 2022), we focus on the parametric specifications that practitioners actually bring to data. Our results highlight that the common

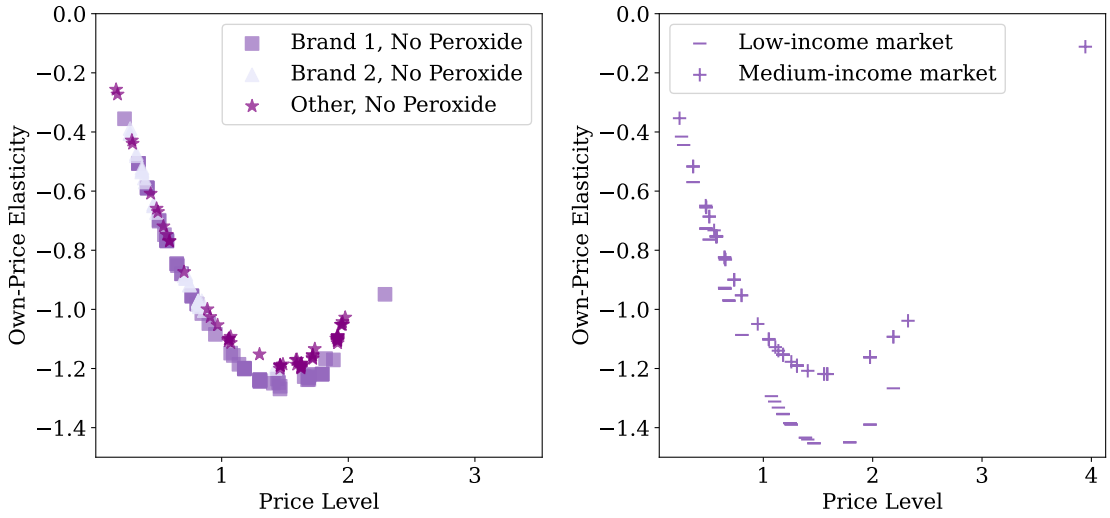


Figure 21: Left: Elasticities from the random coefficient logit model (specification (4)), grouping products by characteristics. Right: Elasticities for Brand 1, no peroxide products from the random coefficient logit model (specification (4)) in markets with different average income levels.

assumptions imposed in practice are not always innocuous.

The results in our paper suggest that the parametric specification of a demand model depends on two distinct but related issues: what theoretical restrictions to impose on consumer behavior and how to recover the parameters from the data. To intelligently make these choices, it is useful to consider a general (but empirically infeasible) version of the model, in which x_j is a vector of product indicators and α varies across both individuals i and products j:

$$v_{ij} = \beta_{ij} + \xi_{jc} + \alpha_{ij}(m_i - p_{jc}) + \varepsilon_{ij}. \tag{51}$$

With sufficiently rich joint distributions of the α_{ij} and β_{ij}, this model allows for own- and cross-price elasticities to vary flexibly across products and price levels. In practice, practitioners will necessarily restrict the model for identification and estimation. The question is how. We now discuss alternative restrictions of preferences within and between consumers, showing how each affects the interpretation and identification of demand elasticities.

The discussion is guided by data from the pricing experiment that generates two exogenous price changes for the same product. Figure 22 illustrates the discussion using data from four randomly drawn toothpaste products. For each product, the figure displays two points, corresponding to IV estimates $\hat{\rho}_j^{\log(q/q_0)}$ constructed using each of the two price changes as instruments.

The first common restriction is to impose a parametric assumption on the distribution of α_{ij} over i, which is often assumed to be normal or log-normal. The economic content of this restriction is that price sensitivity follows a particular, known distribution in the population. The consequence of this restriction is to pin down the particular functional relationship between own-price elasticities and price levels; for example, when α_{ij} follows a symmetric and thin-tailed distribution, price elasticities are approximately quadratic in

price levels for each product. The top left panel of Figure 22 illustrates this restriction: the own-price semi-elasticities are assumed to be linear in price for each product, with product-specific slopes and intercepts determined by fitting a line through the two IV estimands for that products.

The second common restriction is to assume that α_{ij} does not vary across products j. The economic content of this restriction is that differences in own-price elasticities between products are entirely due to heterogeneity in the price sensitivity across consumers who purchase the various products, not within-person heterogeneity in marginal rates of substitution across products. The consequence of this restriction is that own-price elasticities must share a common functional form in price across all products, up to a level shift. The top right panel of Figure 22 illustrates this restriction: the slope with respect to price is now constrained to be the same for all products and is determined by averaging, across products, the product-specific slopes implied by the IV estimands.

The third common restriction is to assume a factor structure $\beta_i' x_j$ for the random intercepts β_{ij}. The economic content of this restriction is that every individual's preferences for product j depend only on that product's observed characteristics x_j (and the i.i.d. shock ε_{ij}). A consequence is that own-price elasticities for any two products with the same observable characteristics must approximately follow the same (quadratic) function of price. The middle-left panel of Figure 22 illustrates this restriction: the intercepts of the elasticity–price relationship are constrained to be the same within groups of products that share the same observable characteristics; in this example, products 1 and 2 form one group, while products 3 and 4 form another.

The fourth common restriction is to assume that β_i is independent of α_i. The economic content of this restriction is that consumer taste for outside goods is independent of taste for inside goods. The consequence of this restriction is that own-price elasticities for all products must approximately follow the same (quadratic) function of price. The middle-right panel of Figure 22 illustrates this restriction: both the slope and intercept of the elasticity–price relationship are now constrained to be identical across products.

The fifth common restriction is to assume that α_i is constant across individuals. The economic content of this restriction is that all consumers share the same price sensitivity. In this case, own-price elasticities must collapse to a single constant value across all products and price levels. This restriction is illustrated in the bottom panel of Figure 22.

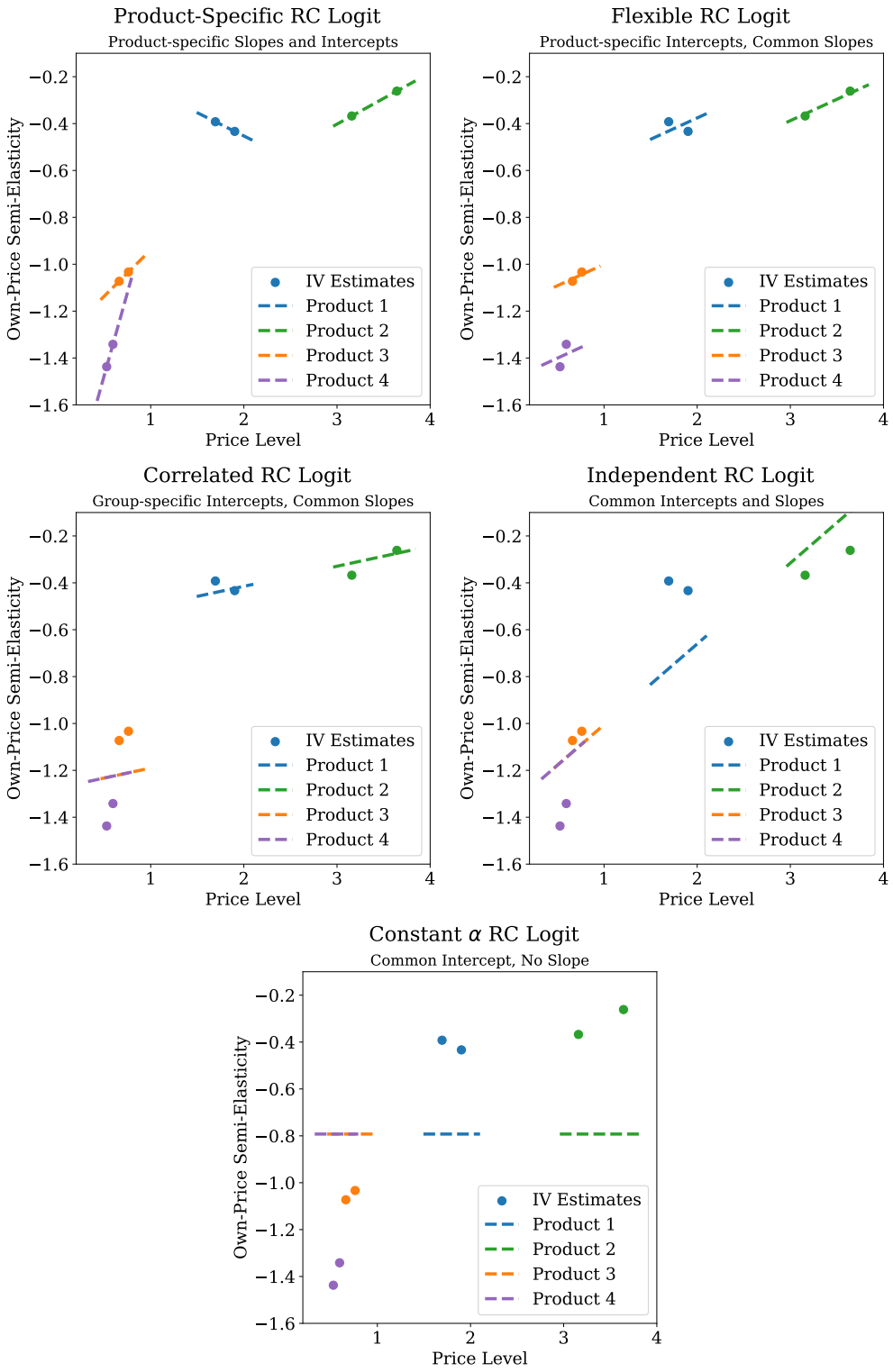


Figure 22: Implications of four common restrictions for semi-elasticities across products. Under restriction 1 (top left), own-price elasticities are linear in price for each product, but the slope and intercept of the linear functional form differs by product. Under restriction 2 (top right), the slope in price must be the same for all products. Under restriction 3 (bottom left), the intercept must be the same within groups of products that share the same characteristics (in this example, products 1 and 2 are one group and products 3 and 4 are another). Under restriction 4 (bottom right), the intercept must be the same for all products.

References

Allenby, G. M. and P. E. Rossi (1991). Quality perceptions and asymmetric switching between brands. *Marketing Science 10*(3), 185–204.

Anderson, S. P., A. de Palma, and J.-F. Thisse (1992). *Discrete Choice Theory of Product Differentiation.* The MIT Press.

Andrews, I., N. Barahona, M. Gentzkow, A. Rambachan, and J. M. Shapiro (2025, 03). Structural estimation under misspecification: Theory and implications for practice*. *The Quarterly Journal of Economics 140*(3), 1801–1855.

Angrist, J. D., K. Graddy, and G. W. Imbens (2000). The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish. *The Review of Economic Studies 67*, 499–527.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica 77*(4), 1229–1279.

Berry, S., A. Gandhi, and P. Haile (2013). Connected substitutes and invertibility of demand. *Econometrica 81*(5), 2087–2111.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile Prices in Market Equilibrium. *Econometrica 63*(4), 841–890.

Berry, S. T. (1994). Estimating Discrete-Choice Models of Product Differentiation. *The RAND Journal of Economics 25*(2), 242–262.

Berry, S. T. and P. A. Haile (2014). Identification in Differentiated Products Markets Using Market Level Data. *Econometrica 82*(5), 1749–1797.

Berry, S. T. and P. A. Haile (2024). Nonparametric Identification of Differentiated Products Demand Using Micro Data. *Econometrica 92*(4), 1135–1162.

Björnerstedt, J. and F. Verboven (2016, July). Does merger simulation work? evidence from the swedish analgesics market. *American Economic Journal: Applied Economics 8*(3), 125–64.

Blandhol, C., J. Bonney, M. Mogstad, and A. Torgovitsky (2025, January). When is TSLS *Actually* LATE? Working paper, Kenneth C. Griffin Department of Economics, University of Chicago. Pre-print, 14 January 2025.

Cardell, N. S. (1997). Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity. *Econometric Theory 13*(2), 185–213.

Compiani, G. (2022). Market counterfactuals and the specification of multiproduct demand: A nonparametric approach. *Quantitative Economics 13*(2), –.

Comtet, L. (1974). *Identities and Expansions*, pp. 127–175. Dordrecht: Springer Netherlands.

Conlon, C. and J. Gortmaker (2020). Best practices for differentiated products demand estimation with PyBLP. *The RAND Journal of Economics 51*(4), 1108–1161.

Creeth, J., M. L. Bosma, and K. Govier (2013). How much is a 'pea-sized amount'? A study of dentifrice dosing by parents in three countries. *International dental journal 63*, 25–30.

DellaVigna, S. (2009, June). Psychology and Economics: Evidence from the Field. *Journal of Economic Literature 47*(2), 315–72.

DellaVigna, S. and M. Gentzkow (2019, 06). Uniform Pricing in U.S. Retail Chains. *The Quarterly Journal of Economics 134*(4), 2011–2084.

Dubé, J.-P. (2019). Chapter 1 - Microeconometric models of consumer demand. In J.-P. Dubé and P. E. Rossi (Eds.), *Handbook of the Economics of Marketing, Volume 1*, Volume 1 of *Handbook of the Economics of Marketing*, pp. 1–68. North-Holland.

Fabinger, M. and E. G. Weyl (2018, Aug). Functional forms for tractable economic models and the cost structure of international trade. CIRJE F-Series CIRJE-F-1092, CIRJE, Faculty of Economics, University of Tokyo.

Gandhi, A. and J.-F. Houde (2019, October). Measuring Substitution Patterns in Differentiated-Products Industries. Technical report. NBER Working Paper No. 26375.

Gandhi, A. and A. Nevo (2021). Chapter 2 - Empirical models of demand and supply in differentiated products industries. In K. Ho, A. Hortaçsu, and A. Lizzeri (Eds.), *Handbook of Industrial Organization, Volume 4*, Volume 4 of *Handbook of Industrial Organization*, pp. 63–139. Elsevier.

Harrison, G. W. and J. A. List (2004, December). Field Experiments. *Journal of Economic Literature 42*(4), 1009–1055.

Imbens, G. W. and J. D. Angrist (1994). Identification and estimation of local average treatment effects. *Econometrica 62*(2), 467–475.

Johnson, W. P. (2002). The curious history of faà di bruno's formula. *The American Mathematical Monthly 109*(3), 217–234.

Knittel, C. R. and K. Metaxoglou (2014). Estimation of Random-Coefficient Demand Models: Two Empiricists' Perspective. *The Review of Economics and Statistics 96*(1), 34–59.

Krishnamurthi, L. and S. P. Raj (1988). A model of brand choice and purchase quantity price sensitivities. *Marketing Science 7*(1), 1–20.

Manski, C. F. (1977). The structure of random utility models. *Theory and Decision 8*(3), 229–254.

McFadden, D. and K. Train (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics 15*(5), 447–470.

Miller, N. H. and M. C. Weinberg (2017). Understanding the price effects of the millercoors joint venture. *Econometrica 85*(6), 1763–1791.

Miravete, E. J., K. Seim, and J. Thurk (2025). Elasticity and Curvature of Discrete Choice Demand Models. Technical report.

Mogstad, M., A. Torgovitsky, and C. R. Walters (2021, November). The causal interpretation of two-stage least squares with multiple instrumental variables. *American Economic Review 111*(11), 3663–98.

Nevo, A. (2000). A practitioner's guide to estimation of random-coefficients logit models of demand. *Journal of Economics & Management Strategy 9*(4), 513–548.

Reynaert, M. and F. Verboven (2014). Improving the performance of random coefficients demand models: The role of optimal instruments. *Journal of econometrics 179*(1), 83–98.

Salanié, B. and F. A. Wolak (2022). Fast, Detail-free, and Approximately Correct: Estimating Mixed Demand Systems. Technical report.

Sokolova, T., S. Seenivasan, and M. Thomas (2020). The left-digit bias: When and why are consumers penny wise and pound foolish? *Journal of Marketing Research 57*(4), 771–788.

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

A Proofs and derivations

A.1 Proof of Proposition 1

Proof. Normalize the price of the outside option to $p_{0c} = 1$. Then the indirect utilities generated by the maximization problem (12) are

$$v_{ij}(p_c, m_i) = U_i(m_i - p_{jc}, 0, \ldots, 1, \ldots, 0),$$
$$v_{i0}(p_c, m_i) = U_i(m_i, 0, \ldots, 0).$$

For (1) to hold, $v_{ij}(p_c, m_i)$ must be affine in $m_i - p_{jc}$ for each j. This is the case if and only if $U_i(q_0, q_1, \ldots, q_J)$ is affine in q_0 whenever $q_j = 1$ for some j and $q_k = 0$ for all $k \neq j$. Thus, for all j,

$$U_i(q_0, q_1, \ldots, q_J) = A_{ijc} + B_{ij} q_0,$$

where A_{ijc} collects the terms independent of q_0.

Matching this with (1) (after subtracting ε_{i0}) requires $A_{ijc} = \beta' x_{jc} + \xi_{jc} + \varepsilon_{ij} - \varepsilon_{i0}$ and $B_{ij} = \alpha_{ij}$ for $j > 0$, while for the outside option $A_{i0c} = 0$ and $B_{i0} = \alpha_{i0}$. Because the $q_j \in \{0, 1\}$ are mutually exclusive across $j > 0$, we can write the direct utility function as

$$U_i(q_0, q_1, \ldots, q_J) = \sum_{j=1}^{J} q_j(\beta' x_{jc} + \xi_{jc} + \varepsilon_{ij} - \varepsilon_{i0}) + \left(\sum_{j=1}^{J} (\alpha_{ij} - \alpha_{i0}) q_j \right) q_0 + \alpha_{i0} q_0,$$

which is exactly (13).

We claim that this utility specification generates product-specific random coefficient logit shares. Individual i chooses to purchase one unit of inside good j if and only if

$$\begin{aligned} &\beta' x_{jc} + \xi_{jc} + \varepsilon_{ij} - \varepsilon_{i0} + (\alpha_{ij} - \alpha_{i0})(m_i - p_{jc}) + \alpha_{i0}(m_i - p_{jc}) \\ &\quad \geq \max \left\{ \alpha_{i0} m_i, \max_{k=1,\ldots,J} \{\beta' x_{kc} \xi_{kc} + \varepsilon_{ik} - \varepsilon_{i0} + (\alpha_{ik} - \alpha_{i0})(m_i - p_{kc}) + \alpha_{i0}(m_i - p_{kc})\} \right\}. \end{aligned}$$

By adding ε_{i0} to all terms, we see this condition is equivalent to

$$\beta' x_{jc} + \xi_{jc} + \alpha_{ij}(m_i - p_{jc}) + \varepsilon_{ij} \geq \max \left\{ \alpha_{i0} m_i + \varepsilon_{i0}, \max_{k=1,\ldots,J} \{\beta' x_{kc} + \xi_{kc} + \alpha_{ik}(m_i - p_{kc}) + \varepsilon_{ik}\} \right\}.$$

Since the ε_{ij} are iid T1EV and independent of α_i and m_i, the share of individuals choosing to purchase one unit of inside good j is

$$s_{jc} = \int \frac{\exp(\beta' x_{jc} + \xi_{jc} + \alpha_{ij}(m_i - p_{jc}))}{\exp(\alpha_{i0} m_i) + \sum_{k=1}^{J} \exp(\beta' x_{kc} + \xi_{kc} + \alpha_{ik}(m_i - p_{kc}))} \, \mathrm{d}F_{m,\alpha}(m_i, \alpha_{i0}, \alpha_{i1}, \ldots, \alpha_{iJ}).$$

We claim that when income m_i and price sensitivities $(\alpha_{i0}, \ldots, \alpha_{iJ})$ are constant across individuals i, the demand system satisfies Slutsky symmetry. In this case, the demand system is

$$s_{jc} = \frac{\exp(\beta' x_{jc} + \xi_{jc} + \alpha_j(m - p_{jc}))}{\exp(\alpha_0 m) + \sum_{k=1}^{J} \exp(\beta' x_{kc} + \xi_{kc} + \alpha_k(m - p_{kc}))}.$$

We compute

$$\begin{aligned}
\frac{\partial s_k}{\partial p_j} &= \alpha_j s_j s_k \\
\frac{\partial s_k}{\partial m} &= \alpha_k(1-s_k)s_k - \sum_{\substack{j=0,\ldots,J \\ j\neq k}} \alpha_j s_j s_k \\
&= \alpha_k(1-s_k)s_k - \left(-\alpha_k s_k s_k + \sum_{j=0}^{J} \alpha_j s_j s_k\right) \\
&= (\alpha_k - \bar{\alpha})s_k
\end{aligned}$$

where $\bar{\alpha} \equiv \sum_{j=0}^{J} s_j \alpha_j$.

The (k,j) element of the Slutsky matrix is thus

$$\begin{aligned}
\frac{\partial s_k}{\partial p_j} + s_j \frac{\partial s_k}{\partial m} &= \alpha_j s_j s_k + (\alpha_k - \bar{\alpha}) s_j s_k \\
&= (\alpha_j + \alpha_k - \bar{\alpha}) s_j s_k.
\end{aligned}$$

This is equal to the (j,k) element of the Slutsky matrix, so Slutsky symmetry holds.

□

A.2 Proof of Proposition 2

The approximations for the logit and product-specific logit models are immediate; the approximation for the random coefficient logit model makes use of the following lemma:

Lemma 1. *In the random coefficient logit model,*

$$\eta_{jjc}(p_c)/p_{jc} = -\mathbb{E}[\alpha_i] + (\mathrm{Var}[\alpha_i](p_{jc} - \bar{p}(\mathbb{E}[\alpha_i])) - \mathrm{err}_1)\frac{1}{1+\mathrm{err}_2} + \mathrm{err}_3. \tag{52}$$

The errors are given by

$$\begin{aligned}
\mathrm{err}_1 &\equiv \frac{1}{2}\,\mathrm{Cov}\left[\alpha_i, \left[(p_{jc} - \bar{p}_c(t(\alpha_i)))^2 - (\overline{p_c^2}(t(\alpha_i)) - \bar{p}_c(t(\alpha_i))^2)\right]\frac{s_{jc}(t(\alpha_i))}{s_{jc}(\mathbb{E}[\alpha_i])}(\alpha_i - \mathbb{E}[\alpha_i])^2\right] \\
\mathrm{err}_2 &\equiv \frac{1}{2}\,\mathbb{E}\left[\left[(p_{jc} - \bar{p}_c(t(\alpha_i)))^2 - (\overline{p_c^2}(t(\alpha_i)) - \bar{p}_c(t(\alpha_i))^2)\right]\frac{s_{jc}(t(\alpha_i))}{s_{jc}(\mathbb{E}[\alpha_i])}(\alpha_i - \mathbb{E}[\alpha_i])^2\right] \\
\mathrm{err}_3 &\equiv \frac{\mathbb{E}[\alpha_i s_{jc}(\alpha_i)^2]}{\mathbb{E}[s_j(\alpha_i)]}
\end{aligned}$$

where t is a function such that $t(\alpha)$ is between α and $\mathbb{E}[\alpha_i]$ for all α,

$$s_{jc}(\alpha_i) \equiv \frac{\exp(\beta' x_{jc} - \alpha_i p_{jc} + \xi_{jc})}{1 + \sum_{k=1}^{J}\exp(\beta' x_{jc} - \alpha_i p_{kc} + \xi_{kc})},$$

$\bar{p}_c(\alpha_i) \equiv \sum_{j=1}^{J} p_{jc} s_{jc}(\alpha_i)$, *and* $\overline{p_c^2}(\alpha_i) \equiv \sum_{j=1}^{J} p_{jc}^2 s_{jc}(\alpha_i)$.

Proof. After factoring out the constant $-p_{jc}$, the numerator of (10) is simply the expectation of the product of α_i and $(1 - s_{jc}(\alpha_i))s_{jc}(\alpha_i)$:

$$\mathbb{E}[\alpha_i(1 - s_{jc}(\alpha_i))s_{jc}(\alpha_i)] = \mathrm{Cov}[\alpha_i, s_{jc}(\alpha_i)] + \mathbb{E}[\alpha_i]\,\mathbb{E}[s_{jc}(\alpha_i)] - \mathbb{E}[\alpha_i s_{jc}(\alpha_i)^2]$$

To approximate the covariance term, we take a second-order approximation to the nonlinear function $\alpha_i \mapsto s_{jc}(\alpha_i)$ around $\mathbb{E}[\alpha_i]$. A standard calculation using the chain rule yields

$$\begin{aligned} s'_{jc}(\alpha_i) &= (\bar{p}_c(\alpha_i) - p_{jc})s_{jc}(\alpha_i) \\ s''_{jc}(\alpha_i) &= \left[(p_{jc} - \bar{p}_c(\alpha_i))^2 - (\overline{p_c^2}(\alpha_i) - \bar{p}_c(\alpha_i)^2)\right] s_{jc}(\alpha_i) \end{aligned}$$

with $\bar{p}_c(\alpha_i) \equiv \sum_{k=1}^J p_k s_k(\alpha_i)$ the share-weighted average price and $\overline{p_c^2}(\alpha_i) \equiv \sum_{k=1}^J p_{kc}^2 s_{kc}(\alpha_i)$. Thus

$$\begin{aligned} &\mathrm{Cov}[\alpha_i, s_{jc}(\alpha_i)] \\ &\quad = \mathrm{Var}[\alpha_i] s_{jc}(\mathbb{E}[\alpha_i])(\bar{p}_c(\mathbb{E}[\alpha_i]) - p_{jc}) \\ &\qquad + \underbrace{\frac{1}{2}\,\mathrm{Cov}\left[\alpha_i, \left[(p_{jc} - \bar{p}_c(t(\alpha_i)))^2 - (\overline{p_c^2}(t(\alpha_i)) - \bar{p}_c(t(\alpha_i))^2)\right] s_{jc}(t(\alpha_i))(\alpha_i - \mathbb{E}[\alpha_i])^2\right]}_{\equiv \mathrm{err}_1 s_{jc}(\mathbb{E}[\alpha_i])} \end{aligned}$$

where $t(\alpha_i)$ is a number between α_i and $\mathbb{E}[\alpha_i]$ for each α_i.

The second-order approximation to $s_{jc}(\alpha_i)$ also gives

$$\begin{aligned} \mathbb{E}[s_{jc}(\alpha_i)] =& s_{jc}(\mathbb{E}[\alpha_i]) \\ &+ \frac{1}{2}\,\mathbb{E}\left[\left[(p_{jc} - \bar{p}_c(t(\alpha_i)))^2 - (\overline{p_c^2}(t(\alpha_i)) - \bar{p}_c(t(\alpha_i))^2)\right] s_{jc}(t(\alpha_i))(\alpha_i - \mathbb{E}[\alpha_i])^2\right] \end{aligned}$$

which we rearrange to

$$\begin{aligned} &\frac{s_{jc}(\mathbb{E}[\alpha_i])}{\mathbb{E}[s_{jc}(\alpha_i)]} \\ &= \frac{1}{1 + \underbrace{\frac{1}{2}\,\mathbb{E}\left[\left[(p_{jc} - \bar{p}_c(t(\alpha_i)))^2 - (\overline{p_c^2}(t(\alpha_i)) - \bar{p}_c(t(\alpha_i))^2)\right] \frac{s_{jc}(t(\alpha_i))}{s_{jc}(\mathbb{E}[\alpha_i])}(\alpha_i - \mathbb{E}[\alpha_i])^2\right]}_{\equiv \mathrm{err}_2}} \end{aligned}$$

Combining the above, we conclude

$$\begin{aligned} \eta_{jjc}(p_c)/p_{jc} &= -\frac{\mathbb{E}[\alpha_i(1 - s_{jc}(\alpha_i))s_{jc}(\alpha_i)]}{\mathbb{E}[s_{jc}(\alpha_i)]} \\ &= -\mathbb{E}[\alpha_i] + \frac{1}{1 + \mathrm{err}_2}\left(\mathrm{Var}[\alpha_i](p_{jc} - \bar{p}_c(\mathbb{E}[\alpha_i])) - \mathrm{err}_1\right) + \mathrm{err}_3 \end{aligned}$$

where $\mathrm{err}_3 \equiv \frac{\mathbb{E}[\alpha_i s_{jc}(\alpha_i)^2]}{\mathbb{E}[s_{jc}(\alpha_i)]}$. □

We next explain how equation (52) maps to the expression given in (16). First, multiply both sides of (52) by p_{jc} and rearrange terms to obtain

$$\begin{aligned}\eta_{jjc}(p_c) = &- \mathbb{E}[\alpha_i]p_{jc} + \mathrm{Var}[\alpha_i](p_{jc} - \bar{p}_c(\mathbb{E}[\alpha_i]))p_{jc} \\ &+ \underbrace{\left(\frac{1}{1+\mathrm{err}_2} - 1\right)\mathrm{Var}[\alpha_i](p_{jc} - \bar{p}_c(\mathbb{E}[\alpha_i]))p_{jc}}_{\equiv A} - \underbrace{\mathrm{err}_1 \frac{1}{1+\mathrm{err}_2} p_{jc}}_{\equiv B} + \underbrace{\mathrm{err}_3 p_{jc}}_{\equiv C}.\end{aligned}$$

To establish (16), we will show that C is $O(\frac{\mathbb{E}[\alpha_i s_{jc}(\alpha_i)^2]}{\mathbb{E}[s_{jc}(\alpha_i)]})$, A is $O(\mathrm{Var}[\alpha_i]^2)$, and B is $O(\mathbb{E}[|\alpha_i - \mathbb{E}[\alpha_i]|^3])$.

That C is $O(\frac{\mathbb{E}[\alpha_i s_{jc}(\alpha_i)^2]}{\mathbb{E}[s_{jc}(\alpha_i)]})$ follows immediately from the definition of $\mathrm{err}_3 = \frac{\mathbb{E}[\alpha_i s_{jc}(\alpha_i)^2]}{\mathbb{E}[s_{jc}(\alpha_i)]}$.

To analyze A, we note first that

$$\begin{aligned}|\mathrm{err}_2| &= \left|\frac{1}{2}\mathbb{E}\left[\left[(p_{jc} - \bar{p}_c(t(\alpha_i)))^2 - (\overline{p_c^2}(t(\alpha_i)) - \bar{p}_c(t(\alpha_i))^2)\right]\frac{s_{jc}(t(\alpha_i))}{s_{jc}(\mathbb{E}[\alpha_i])}(\alpha_i - \mathbb{E}[\alpha_i])^2\right]\right| \\ &\leq \mathrm{Var}[\alpha_i] \sup_{\alpha_i}\left|\left[(p_{jc} - \bar{p}_c(t(\alpha_i)))^2 - (\overline{p_c^2}(t(\alpha_i)) - \bar{p}_c(t(\alpha_i))^2)\right]\frac{s_{jc}(t(\alpha_i))}{s_{jc}(\mathbb{E}[\alpha_i])}\right| \\ &= O(\mathrm{Var}[\alpha_i]).\end{aligned}$$

Thus $\mathrm{err}_2 \to 0$ as $\mathrm{Var}[\alpha_i] \to 0$. In particular, $|\mathrm{err}_2| < 1$ for $\mathrm{Var}[\alpha_i]$ sufficiently small, so we can expand A as

$$\begin{aligned}\left(\frac{1}{1+\mathrm{err}_2} - 1\right)\mathrm{Var}[\alpha_i](p_{jc} - \bar{p}_c(\mathbb{E}[\alpha_i]))p_{jc} &= \left(\sum_{k=1}^{\infty}(-\mathrm{err}_2)^k\right)\mathrm{Var}[\alpha_i](p_{jc} - \bar{p}_c(\mathbb{E}[\alpha_i]))p_{jc} \\ &= \left(\sum_{k=1}^{\infty} O(\mathrm{Var}[\alpha_i]^k)\right)\mathrm{Var}[\alpha_i](p_{jc} - \bar{p}_c(\mathbb{E}[\alpha_i]))p_{jc} \\ &= O(\mathrm{Var}[\alpha_i])\,\mathrm{Var}[\alpha_i](p_{jc} - \bar{p}_c(\mathbb{E}[\alpha_i]))p_{jc} \\ &= O(\mathrm{Var}[\alpha_i]^2).\end{aligned}$$

Finally, to analyze B, we note that $\frac{1}{1+\mathrm{err}_2}p_{jc}$ is $O(1)$ by the same argument just applied in analyzing A while

$$\begin{aligned}|\mathrm{err}_1| &\equiv \left|\frac{1}{2}\mathrm{Cov}\left[\alpha_i, \left[(p_{jc} - \bar{p}_c(t(\alpha_i)))^2 - (\overline{p_c^2}(t(\alpha_i)) - \bar{p}_c(t(\alpha_i))^2)\right]\frac{s_{jc}(t(\alpha_i))}{s_{jc}(\mathbb{E}[\alpha_i])}(\alpha_i - \mathbb{E}[\alpha_i])^2\right]\right| \\ &= \left|\frac{1}{2}\mathbb{E}\left[(p_{jc} - \bar{p}_c(t(\alpha_i)))^2 - (\overline{p_c^2}(t(\alpha_i)) - \bar{p}_c(t(\alpha_i))^2)\frac{s_{jc}(t(\alpha_i))}{s_{jc}(\mathbb{E}[\alpha_i])}(\alpha_i - \mathbb{E}[\alpha_i])^3\right]\right| \\ &\leq \mathbb{E}[|\alpha_i - \mathbb{E}[\alpha_i]|^3] \sup_{\alpha_i}\left|\frac{1}{2}(p_{jc} - \bar{p}_c(t(\alpha_i)))^2 - (\overline{p_c^2}(t(\alpha_i)) - \bar{p}_c(t(\alpha_i))^2)\frac{s_{jc}(t(\alpha_i))}{s_{jc}(\mathbb{E}[\alpha_i])}\right| \\ &= O(\mathbb{E}[|\alpha_i - \mathbb{E}[\alpha_i]|^3]).\end{aligned}$$

Thus B is $O(|\alpha_i - \mathbb{E}[\alpha_i]|^3)$.

A.3 Proof of (17)

We derive a third-order analogue to Lemma 1; the derivation for the fourth-order analogue is similar.

Lemma 2. *In the random coefficient logit model,*

$$
\begin{aligned}
&\eta_{jjc}(p_c)/p_{jc} \\
&= -\frac{\mathbb{E}[\alpha_i(1-s_{jc}(\alpha_i))s_{jc}(\alpha_i)]}{\mathbb{E}[s_{jc}(\alpha_i)]} \\
&= -\mathbb{E}[\alpha_i] \\
&+ \frac{1}{1+\text{err}_2}\Bigg(\text{Var}[\alpha_i](p_{jc}-\bar{p}_c(\mathbb{E}[\alpha_i])) \\
&\quad - \frac{1}{2}\mathbb{E}[(\alpha_i-\mathbb{E}[\alpha_i])^3]\left[(p_{jc}-\bar{p}_c(\mathbb{E}[\alpha_i]))^2 - (\overline{p_c^2}(\mathbb{E}[\alpha_i]) - \bar{p}_c(\mathbb{E}[\alpha_i])^2)\right] \\
&\quad - \text{err}_1\Bigg) \\
&+ \text{err}_3
\end{aligned}
$$

where err_1, err_2, *and* err_3 *are defined in the proof below.*

Proof. The proof follows the same logic as that of Lemma 1, but using fourth-order Taylor expansions. We first compute

$$
s''_{jc}(\alpha_i) = \left[(p_{jc}-\bar{p}_c(\alpha_i))^2 - (\overline{p_c^2}(\alpha_i) - \bar{p}_c(\mathbb{E}[\alpha_i])^2)\right] s_{jc}(\alpha_i)
$$

$$
\begin{aligned}
\text{Cov}[\alpha_i, s_{jc}(\alpha_i)] =& \text{Var}[\alpha_i]s_{jc}(\mathbb{E}[\alpha_i])(\bar{p}_c(\mathbb{E}[\alpha_i]) - p_{jc}) \\
&+ \frac{1}{2}\mathbb{E}[(\alpha_i-\mathbb{E}[\alpha_i])^3]s_{jc}(\mathbb{E}[\alpha_i])\left[(p_{jc}-\bar{p}(\mathbb{E}[\alpha_i]))^2 - (\overline{p_c^2}(\mathbb{E}[\alpha_i]) - \bar{p}_c(\mathbb{E}[\alpha_i])^2)\right] \\
&+ \underbrace{\frac{1}{6}\text{Cov}\left[\alpha_i, s'''_{jc}(t(\alpha_i))(\alpha_i-\mathbb{E}[\alpha_i])^3\right]}_{\equiv \text{err}_1 s_{jc}(\mathbb{E}[\alpha_i])}
\end{aligned}
$$

where $t(\alpha_i)$ is a number between α_i and $\mathbb{E}[\alpha_i]$ for each α_i. We also have

$$
\begin{aligned}
\mathbb{E}[s_{jc}(\alpha_i)] =& s_{jc}(\mathbb{E}[\alpha_i]) \\
&+ \frac{1}{2}\text{Var}[\alpha_i]s_{jc}(\mathbb{E}[\alpha_i])\left[(p_{jc}-\bar{p}(\mathbb{E}[\alpha_i]))^2 - (\overline{p_c^2}(\mathbb{E}[\alpha_i]) - \bar{p}_c(\mathbb{E}[\alpha_i])^2)\right] \\
&+ \frac{1}{6}\mathbb{E}[s'''_{jc}(t(\alpha_i))(\alpha_i-\mathbb{E}[\alpha_i])^3]
\end{aligned}
$$

which we rearrange to

$$\frac{s_{jc}(\mathbb{E}[\alpha_i])}{\mathbb{E}[s_{jc}(\alpha_i)]}$$
$$= \frac{1}{1 + \underbrace{\frac{1}{2}\mathrm{Var}[\alpha_i]\left[(p_{jc} - \bar{p}_c(\mathbb{E}[\alpha_i]))^2 - (\overline{p_c^2}(\mathbb{E}[\alpha_i]) - \bar{p}_c(\mathbb{E}[\alpha_i])^2)\right] + \frac{1}{6}\frac{\mathbb{E}[s'''_{jc}(t(\alpha_i))(\alpha_i - \mathbb{E}[\alpha_i])^3]}{s_{jc}(\mathbb{E}[\alpha_i])}}_{\equiv \mathrm{err}_2}}$$

Combining the above, we conclude

$$\begin{aligned}
\eta_{jjc}(p_c)/p_{jc} &= -\frac{\mathbb{E}[\alpha_i(1 - s_{jc}(\alpha_i))s_{jc}(\alpha_i)]}{\mathbb{E}[s_{jc}(\alpha_i)]} \\
&= -\mathbb{E}[\alpha_i] \\
&\quad + \frac{1}{1 + \mathrm{err}_2}\Bigg(\mathrm{Var}[\alpha_i](p_{jc} - \bar{p}_c(\mathbb{E}[\alpha_i])) \\
&\qquad - \frac{1}{2}\mathbb{E}[(\alpha_i - \mathbb{E}[\alpha_i])^3]\left[(p_{jc} - \bar{p}_c(\mathbb{E}[\alpha_i]))^2 - (\overline{p_c^2}(\mathbb{E}[\alpha_i]) - \bar{p}_c(\mathbb{E}[\alpha_i])^2)\right] \\
&\qquad - \mathrm{err}_1\Bigg) \\
&\quad + \mathrm{err}_3
\end{aligned}$$

where $\mathrm{err}_3 \equiv \frac{\mathbb{E}[\alpha_i s_{jc}(\alpha_i)^2]}{\mathbb{E}[s_{jc}(\alpha_i)]}$. □

A.4 Proof of Proposition 3

By tracing the logic used in Lemmas 1 and 2, we can see that the mth term in the Mth order approximation to the own-price elasticity is

$$\frac{\mathbb{E}[(\alpha_i - \mathbb{E}[\alpha_i])^m]}{(m-1)!}\frac{s_{jc}^{(m-1)}(\mathbb{E}[\alpha_i])}{s_{jc}(\mathbb{E}[\alpha_i])}p_{jc}$$

where $s_{jc}^{(m-1)}(\mathbb{E}[\alpha_i])$ denotes the $(m-1)$th derivative of the function $s_{jc}(\cdot)$ evaluated at $\mathbb{E}[\alpha_i]$. To establish Proposition 3, it will suffice to show that

$$L_{m-1}(p_{jc} - \bar{p}_c; \bar{p}_c(\mathbb{E}[\alpha_i]), \ldots, \overline{p_c^{m-1}}(\mathbb{E}[\alpha_i])) \equiv \frac{s_{jc}^{(m-1)}(\mathbb{E}[\alpha_i])}{s_{jc}(\mathbb{E}[\alpha_i])}$$

is an $(m-1)$th order polynomial in $p_{jc} - \bar{p}_c$, the coefficients of which depend only on the first $m-1$ moments $\bar{p}_c(\mathbb{E}[\alpha_i]), \ldots, \overline{p_c^{m-1}}(\mathbb{E}[\alpha_i])$ of the distribution of prices (weighted by the purchase shares of individuals with price sensitivity $\mathbb{E}[\alpha_i]$).

Toward this end, we first define a random variable P taking values in the discrete set $\{0, p_{1c}, \ldots, p_{Jc}\}$ with probabilities $\{s_{0c}(\mathbb{E}[\alpha_i]), s_{1c}(\mathbb{E}[\alpha_i]), \ldots, s_{Jc}(\mathbb{E}[\alpha_i])\}$, respectively. Let

$$\begin{aligned} C(t) &\equiv \log(\mathbb{E}[\exp(tP)]) \\ &= \log \sum_{j=0}^{J} \frac{\exp(\beta' x_{jc} + \xi_{jc} - \mathbb{E}[\alpha_i] p_{jc})}{1 + \sum_{k=1}^{J} \exp(\beta' x_{kc} + \xi_{kc} - \mathbb{E}[\alpha_i] p_{kc})} \exp(t p_{jc}) \\ &= \log \frac{1 + \sum_{j=1}^{J} \exp(\beta' x_{jc} + \xi_{jc} - (\mathbb{E}[\alpha_i] - t) p_{jc})}{1 + \sum_{k=1}^{J} \exp(\beta' x_{kc} + \xi_{kc} - \mathbb{E}[\alpha_i] p_{kc})} \end{aligned}$$

denote the cumulant generating function of P.

We can rewrite $s_{jc}(\alpha_i)$ as an exponential of a function involving C:

$$\begin{aligned} s_{jc}(\alpha_i) &= \exp\left(\beta' x_{jc} + \xi_{jc} - \alpha_i p_{jc} - \log\left(1 + \sum_{k=1}^{J} \exp(\beta' x_{kc} + \xi_{kc} - \alpha_i p_{kc})\right)\right) \\ &= \exp\left(\beta' x_{jc} + \xi_{jc} - \alpha_i p_{jc} - C(\mathbb{E}[\alpha_i] - \alpha_i) - \log\left(1 + \sum_{k=1}^{J} \exp(\beta' x_{kc} + \xi_{kc} - \mathbb{E}[\alpha_i] p_{kc})\right)\right). \end{aligned}$$

Let $g(\alpha_i) \equiv \beta' x_{jc} + \xi_{jc} - \alpha_i p_{jc} - C(\mathbb{E}[\alpha_i] - \alpha_i) - \log\left(1 + \sum_{k=1}^{J} \exp(\beta' x_{kc} + \xi_{kc} - \mathbb{E}[\alpha_i] p_{kc})\right)$ denote the expression inside the exponential. By Faà di Bruno's formula, we have

$$\begin{aligned} s_{jc}^{(m-1)}(\alpha_i) &= B_{m-1}(g'(\alpha_i), g''(\alpha_i), \ldots, g^{(m-1)}(\alpha_i)) s_{jc}(\alpha_i) \\ &= B_{m-1}(-p_{jc} + C'(\mathbb{E}[\alpha_i] - \alpha_i), -C''(\mathbb{E}[\alpha_i] - \alpha_i), \ldots, (-1)^m C^{(m-1)}(\mathbb{E}[\alpha_i] - \alpha_i)) s_{jc}(\alpha_i) \end{aligned}$$

where B_{m-1} is the $(m-1)$th complete exponential Bell polynomial (Johnson, 2002; Comtet, 1974). At $\alpha_i = \mathbb{E}[\alpha_i]$, we have

$$s_{jc}^{(m-1)}(\mathbb{E}[\alpha_i]) = B_{m-1}(C'(0) - p_{jc}, -C''(0), \ldots, (-1)^m C^{(m-1)}(0)) s_{jc}(\mathbb{E}[\alpha_i]).$$

The ℓth derivative of C evaluated at 0 is the ℓth cumulant of P, κ_ℓ; in particular, $C'(0) = \bar{p}_c = \sum_{j=0}^{J} s_{jc}(\mathbb{E}[\alpha_i]) p_{jc}$ is the expectation of P and $C''(0) = \overline{p_c^2} - \bar{p}_c^2$ is the variance of P. We thus have

$$\frac{s_{jc}^{(m-1)}(\mathbb{E}[\alpha_i])}{s_{jc}(\mathbb{E}[\alpha_i])} = B_{m-1}(\bar{p}_c - p_{jc}, -\kappa_2, \ldots, (-1)^m \kappa_{m-1}).$$

The right-hand side is a polynomial in $\bar{p}_c - p_{jc}$ with coefficients depending on the first $(m-1)$ cumulants of the distribution of prices. Since the ℓth cumulant can be written in terms of the first ℓ moments, we are done.

Next, we further simplify the expression for $\frac{s_{jc}^{(m-1)}(\mathbb{E}[\alpha_i])}{s_{jc}(\mathbb{E}[\alpha_i])}$. First, using the homogeneity property of the Bell polynomial, we can rewrite the expression as

$$\begin{aligned} \frac{s_{jc}^{(m-1)}(\mathbb{E}[\alpha_i])}{s_{jc}(\mathbb{E}[\alpha_i])} &= \kappa_2^{(m-1)/2} B_{m-1}\left(\frac{\bar{p}_c - p_{jc}}{\kappa_2^{1/2}}, -1, \frac{\kappa_3}{\kappa_2^{3/2}}, \ldots, (-1)^m \frac{\kappa_{m-1}}{\kappa_2^{(m-1)/2}}\right) \\ &= (-1)^{m-1} \kappa_2^{(m-1)/2} B_{m-1}\left(\frac{p_{jc} - \bar{p}_c}{\kappa_2^{1/2}}, -1, -\frac{\kappa_3}{\kappa_2^{3/2}}, \ldots, -\frac{\kappa_{m-1}}{\kappa_2^{(m-1)/2}}\right) \end{aligned}$$

$\frac{p_{jc}-\bar{p}_c}{\kappa_2^{1/2}}$ is the number of standard deviations between p_{jc} and the mean price in the market. $\kappa_\ell/\kappa_2^{\ell/2}$ is the ℓth standardized cumulant of the price distribution.

To further interpret this expression, consider the special case where P is (approximately) normally distributed, that is, when the weighted distribution of *prices* in the market is (approximately) normal. In this case, $\kappa_3 = \cdots = \kappa_m = 0$, so

$$B_{m-1}\left(\frac{p_{jc}-\bar{p}_c}{\kappa_2^{1/2}}, -1, 0, \ldots, 0\right) = (m-1)! \sum_{\substack{r_1,r_2 \\ r_1+2r_2=m-1}} \frac{(-1)^{r_2}}{2^{r_2} r_1! r_2!} \left(\frac{p_{jc}-\bar{p}_c}{\kappa_2^{1/2}}\right)^{r_1}$$

A.5 Proof of Proposition 7

The proof follows from the following lemma:

Lemma 3. *In the random coefficient logit model,*

$$\begin{aligned}
&\eta_{kjc}(p_c)/p_{jc} \\
&= \mathbb{E}[\alpha_i]\,\mathbb{E}[s_{jc}(\alpha_i,\beta_i)]\Bigg[1 + \frac{1}{1+\text{err}_{2k}}\frac{1}{1+\text{err}_{2j}}\text{err}_3 \\
&\qquad + \frac{1}{1+\text{err}_{2k}}\frac{1}{1+\text{err}_{2j}}\,\text{Var}[\alpha_i](p_{jc}-\bar{p}_c(\mathbb{E}[\alpha_i],\mathbb{E}[\beta_i]))(p_{kc}-\bar{p}_c(\mathbb{E}[\alpha_i],\mathbb{E}[\beta_i])) \\
&\qquad + \frac{1}{1+\text{err}_{2k}}\frac{1}{1+\text{err}_{2j}}\sum_{r=1}^{R}\text{Var}[\beta_{ir}](x_{jrc}-\bar{x}_{rc}(\mathbb{E}[\alpha_i],\mathbb{E}[\beta_i]))(x_{krc}-\bar{x}_{rc}(\mathbb{E}[\alpha_i],\mathbb{E}[\beta_i]))\Bigg] \\
&\quad + \frac{1}{1+\text{err}_{2k}}\Bigg(\text{Var}[\alpha_i]s_{jc}(\mathbb{E}[\alpha_i],\mathbb{E}[\beta_i])(2\bar{p}_c(\mathbb{E}[\alpha_i],\mathbb{E}[\beta_i]) - p_{kc} - p_{jc}) + \text{err}_1\Bigg),
\end{aligned}$$

where the errors are defined below.

Proof. The numerator of the cross-price elasticity is the expectation of the product of α_i and $s_{kc}(\alpha_i,\beta_i)s_{jc}(\alpha_i,\beta_i)$:

$$\mathbb{E}[\alpha_i s_{kc}(\alpha_i,\beta_i)s_{jc}(\alpha_i,\beta_i)] = \text{Cov}[\alpha_i, s_{kc}(\alpha_i,\beta_i)s_{jc}(\alpha_i,\beta_i)] + \mathbb{E}[\alpha_i]\,\mathbb{E}[s_{kc}(\alpha_i,\beta_i)s_{jc}(\alpha_i,\beta_i)]$$

To approximate the covariance term, we take a second-order approximation to $s_{kc}(\alpha_i,\beta_i)s_{jc}(\alpha_i,\beta_i)$ around $(\mathbb{E}[\alpha_i],\mathbb{E}[\beta_i])$. We have

$$\begin{aligned}
&\text{Cov}[\alpha_i, s_{kc}(\alpha_i,\beta_i)s_{jc}(\alpha_i,\beta_i)] \\
&\quad = \text{Var}[\alpha_i](2\bar{p}_c(\mathbb{E}[\alpha_i]) - p_{kc} - p_{jc})s_{kc}(\mathbb{E}[\alpha_i],\mathbb{E}[\beta_i])s_{jc}(\mathbb{E}[\alpha_i],\mathbb{E}[\beta_i]) \\
&\qquad + \underbrace{\frac{1}{2}\,\text{Cov}\left[\alpha_i, [(\alpha_i,\beta_i) - \mathbb{E}[(\alpha_i,\beta_i)]]' H_{s_{kc}s_{jc}}(t(\alpha_i,\beta_i))[(\alpha_i,\beta_i) - \mathbb{E}[(\alpha_i,\beta_i)]]\right]}_{\equiv \text{err}_1 s_{kc}(\mathbb{E}[\alpha_i],\mathbb{E}[\beta_i])}
\end{aligned}$$

where $t(\alpha_i,\beta_i)$ lies on the line segment connecting (α_i,β_i) and $(\mathbb{E}[\alpha_i],\mathbb{E}[\beta_i])$ for each value of (α_i,β_i) and $H_{s_{kc}s_{jc}}(\alpha_i,\beta_i)$ is the Hessian of the function $(\alpha_i,\beta_i) \mapsto s_{kc}(\alpha_i,\beta_i)s_{jc}(\alpha_i,\beta_i)$.

The second-order approximation to $s_{kc}(\alpha_i, \beta_i)$ also gives

$$\begin{aligned}\mathbb{E}[s_{kc}(\alpha_i, \beta_i)] =& s_{kc}(\mathbb{E}[\alpha_i], \mathbb{E}[\beta_i]) \\ &+ \frac{1}{2}\mathbb{E}\left[[(\alpha_i, \beta_i) - \mathbb{E}[(\alpha_i, \beta_i)]]' H_{s_{kc}}(t(\alpha_i, \beta_i))[(\alpha_i, \beta_i) - \mathbb{E}[(\alpha_i, \beta_i)]]\right]\end{aligned}$$

which we rearrange to

$$\frac{s_{kc}(\mathbb{E}[\alpha_i], \mathbb{E}[\beta_i])}{\mathbb{E}[s_{kc}(\alpha_i, \beta_i)]} = \frac{1}{1 + \underbrace{\frac{1}{2}\frac{\mathbb{E}\left[[(\alpha_i, \beta_i) - \mathbb{E}[(\alpha_i, \beta_i)]]' H_{s_{kc}}(t(\alpha_i, \beta_i))[(\alpha_i, \beta_i) - \mathbb{E}[(\alpha_i, \beta_i)]]\right]}{s_k(\mathbb{E}[\alpha_i], \mathbb{E}[\beta_i])}}_{\equiv \mathrm{err}_{2k}}}.$$

To approximate the cross-product term, we use another Taylor expansion:

$$\begin{aligned}&\mathbb{E}[s_{kc}(\alpha_i, \beta_i) s_{jc}(\alpha_i, \beta_i)] \\ &= \mathbb{E}[s_{kc}(\alpha_i, \beta_i)]\,\mathbb{E}[s_{jc}(\alpha_i, \beta_i)] + \mathrm{Cov}[s_{kc}(\alpha_i, \beta_i), s_{jc}(\alpha_i, \beta_i)] \\ &= \mathbb{E}[s_{kc}(\alpha_i, \beta_i)]\,\mathbb{E}[s_{jc}(\alpha_i, \beta_i)]\Bigg[1 + \frac{1}{1 + \mathrm{err}_{2k}}\frac{1}{1 + \mathrm{err}_{2j}}\mathrm{err}_3 \\ &\quad + \frac{1}{1 + \mathrm{err}_{2k}}\frac{1}{1 + \mathrm{err}_{2j}}\mathrm{Var}[\alpha_i](p_j - \bar{p}_c(\mathbb{E}[\alpha_i], \mathbb{E}[\beta_i]))(p_{kc} - \bar{p}_c(\mathbb{E}[\alpha_i], \mathbb{E}[\beta_i])) \\ &\quad + \frac{1}{1 + \mathrm{err}_{2k}}\frac{1}{1 + \mathrm{err}_{2j}}\sum_{r=1}^{R}\mathrm{Var}[\beta_{ir}](x_{jrc} - \bar{x}_{rc}(\mathbb{E}[\alpha_i], \mathbb{E}[\beta_i]))(x_{krc} - \bar{x}_{rc}(\mathbb{E}[\alpha_i], \mathbb{E}[\beta_i]))\Bigg]\end{aligned}$$

where

$$\begin{aligned}\mathrm{err}_3 \equiv& \frac{1}{2}\mathrm{Cov}\left[s_{kc}(\alpha_i, \beta_i), [(\alpha_i, \beta_i) - \mathbb{E}[(\alpha_i, \beta_i)]]' D^2 s_{jc}(\mathbb{E}[\alpha_i], \mathbb{E}[\beta_i])[(\alpha_i, \beta_i) - \mathbb{E}[(\alpha_i, \beta_i)]]\right] \\ &+ \frac{1}{2}\mathrm{Cov}\left[s_{jc}(\alpha_i, \beta_i), [(\alpha_i, \beta_i) - \mathbb{E}[(\alpha_i, \beta_i)]]' D^2 s_{kc}(\mathbb{E}[\alpha_i], \mathbb{E}[\beta_i])[(\alpha_i, \beta_i) - \mathbb{E}[(\alpha_i, \beta_i)]]\right] \\ &- \frac{1}{4}\mathrm{Cov}\Bigg[[(\alpha_i, \beta_i) - \mathbb{E}[(\alpha_i, \beta_i)]]' D^2 s_{kc}(\mathbb{E}[\alpha_i], \mathbb{E}[\beta_i])[(\alpha_i, \beta_i) - \mathbb{E}[(\alpha_i, \beta_i)]], \\ &\qquad [(\alpha_i, \beta_i) - \mathbb{E}[(\alpha_i, \beta_i)]]' D^2 s_{jc}(\mathbb{E}[\alpha_i], \mathbb{E}[\beta_i])[(\alpha_i, \beta_i) - \mathbb{E}[(\alpha_i, \beta_i)]]\Bigg].\end{aligned}$$

Combining the above, we have

$$
\begin{aligned}
&\eta_{kjc}(p_c)/p_{jc} \\
&= \frac{\mathbb{E}[\alpha_i s_{kc}(\alpha_i,\beta_i) s_{jc}(\alpha_i,\beta_i)]}{\mathbb{E}[s_{kc}(\alpha_i,\beta_i)]} \\
&= \mathbb{E}[\alpha_i]\,\mathbb{E}[s_{jc}(\alpha_i,\beta_i)]\Bigg[1 + \frac{1}{1+\mathrm{err}_{2k}}\frac{1}{1+\mathrm{err}_{2j}}\mathrm{err}_3 \\
&\qquad + \frac{1}{1+\mathrm{err}_{2k}}\frac{1}{1+\mathrm{err}_{2j}}\mathrm{Var}[\alpha_i](p_{jc} - \bar{p}_c(\mathbb{E}[\alpha_i],\mathbb{E}[\beta_i]))(p_{kc} - \bar{p}_c(\mathbb{E}[\alpha_i],\mathbb{E}[\beta_i])) \\
&\qquad + \frac{1}{1+\mathrm{err}_{2k}}\frac{1}{1+\mathrm{err}_{2j}}\sum_{r=1}^{R}\mathrm{Var}[\beta_{ir}](x_{jrc} - \bar{x}_{rc}(\mathbb{E}[\alpha_i],\mathbb{E}[\beta_i]))(x_{krc} - \bar{x}_{rc}(\mathbb{E}[\alpha_i],\mathbb{E}[\beta_i]))\Bigg] \\
&\quad + \frac{1}{1+\mathrm{err}_{2k}}\Bigg(\mathrm{Var}[\alpha_i] s_{jc}(\mathbb{E}[\alpha_i],\mathbb{E}[\beta_i])(2\bar{p}_c(\mathbb{E}[\alpha_i],\mathbb{E}[\beta_i]) - p_{kc} - p_{jc}) + \mathrm{err}_1\Bigg).
\end{aligned}
$$

□

A.6 Lemma 4

Lemma 4.

$$
\frac{\partial^2 \sigma_j^{-1}}{\partial \lambda^2}(s_c;0) = -p_{jc}(p_{jc} - 2\bar{p}_c), \tag{53}
$$

where $\bar{p}_c \equiv \sum_{j=1}^{J} s_{jc} p_{jc}$ is the share-weighted average price in market c (counting the outside option with price 0).

Proof. **Calculating $\frac{\partial \sigma_j^{-1}}{\partial \lambda}(s_c;\lambda)$.** We first derive an expression for $\frac{\partial \sigma_j^{-1}}{\partial \lambda}(s_c;\lambda)$ with the implicit and inverse function theorems. Implicitly differentiating the identity $\sigma^{-1}(\sigma(\delta;\lambda);\lambda) = \delta$ with respect to λ yields

$$
\frac{\partial \sigma^{-1}}{\partial s}(\sigma(\delta;\lambda);\lambda)\frac{\partial \sigma}{\partial \lambda}(\delta;\lambda) + \frac{\partial \sigma^{-1}}{\partial \lambda}(\sigma(\delta;\lambda);\lambda) = 0. \tag{54}
$$

(Here $\frac{\partial \sigma^{-1}}{\partial s}(\sigma(\delta;\lambda);\lambda)$ is $J \times J$ and $\frac{\partial \sigma}{\partial \lambda}(\delta;\lambda)$ and $\frac{\partial \sigma^{-1}}{\partial \lambda}(\sigma(\delta;\lambda);\lambda)$ are both $J \times 1$.) The inverse function theorem gives $\frac{\partial \sigma^{-1}}{\partial s}(\sigma(\delta;\lambda);\lambda) = \left[\frac{\partial \sigma}{\partial \delta}(\delta;\lambda)\right]^{-1}$, so

$$
-\frac{\partial \sigma^{-1}}{\partial \lambda}(\sigma(\delta;\lambda);\lambda) = \left[\frac{\partial \sigma}{\partial \delta}(\delta;\lambda)\right]^{-1}\frac{\partial \sigma}{\partial \lambda}(\delta;\lambda). \tag{55}
$$

Direct calculation yields

$$
\frac{\partial \sigma_j}{\partial \lambda}(\delta_c;\lambda) = \int \nu_i (p_{jc} - \bar{p}_{ic})\sigma_j(\delta_c + \lambda \nu_i p_c;0)\,\mathrm{d}\Phi(\nu_i)
$$

where $\sigma_j(\delta_c + \lambda\nu_i p_c; 0) = \frac{\exp(\delta_{jc}+\lambda\nu_i p_{jc})}{1+\sum_{k=1}^J \exp(\delta_{kc}+\lambda\nu_i p_{kc})}$ and $\bar{p}_{ic} \equiv \sum_{k=1}^J p_{kc}\sigma_k(\delta_c + \lambda\nu_i p_c; 0)$. Since Φ has mean 0,

$$\frac{\partial \sigma_j}{\partial \lambda}(\delta_c; 0) = (p_{jc} - \bar{p}_c)\sigma_j(\delta_c; 0) \int \nu_i \, \mathrm{d}\Phi(\nu_i) = 0$$

and thus

$$\frac{\partial \sigma^{-1}}{\partial \lambda}(s; 0) = -\left[\frac{\partial \sigma}{\partial \delta}(\sigma^{-1}(s; 0); 0)\right]^{-1} \frac{\partial \sigma}{\partial \lambda}(\sigma^{-1}(s; 0); 0) = 0$$

for all s.

Second derivative. We next compute $\frac{\partial^2 \sigma_j^{-1}}{\partial \lambda^2}$. We implicitly differentiate (54) again with respect to λ and set $\lambda = 0$. All terms with a factor of $\frac{\partial \sigma_j}{\partial \lambda}$ vanish, leaving

$$\frac{\partial \sigma^{-1}}{\partial s}(\sigma(\delta; 0); 0)\frac{\partial^2 \sigma}{\partial \lambda^2}(\delta; 0) + \frac{\partial^2 \sigma^{-1}}{\partial \lambda^2}(\sigma(\delta; 0); 0) = 0.$$

By the inverse function theorem, $\frac{\partial \sigma^{-1}}{\partial s}(\sigma(\delta; 0); 0) = \left[\frac{\partial \sigma}{\partial \delta}(\delta; 0)\right]^{-1}$, so we have

$$-\frac{\partial^2 \sigma^{-1}}{\partial \lambda^2}(\sigma(\delta; 0); 0) = \left[\frac{\partial \sigma}{\partial \delta}(\delta; 0)\right]^{-1} \frac{\partial^2 \sigma}{\partial \lambda^2}(\delta; 0). \tag{56}$$

We first solve for $\left[\frac{\partial \sigma}{\partial \delta}(\delta; \lambda)\right]^{-1}$. Direct calculation yields

$$\begin{aligned}
\frac{\partial \sigma_j}{\partial \delta_j}(\delta_c; \lambda) &= \int (1 - \sigma_j(\delta_c + \lambda\nu_i p_c; 0))\sigma_j(\delta_c + \lambda\nu_i p_c; 0) \, \mathrm{d}\Phi(\nu_i), \\
\frac{\partial \sigma_j}{\partial \delta_k}(\delta_c; \lambda) &= \int -\sigma_j(\delta_c + \lambda\nu_i p_c; 0)\sigma_k(\delta_c + \lambda\nu_i p_c; 0) \, \mathrm{d}\Phi(\nu_i), \qquad k \neq j.
\end{aligned}$$

In matrix form, we have

$$\frac{\partial \sigma}{\partial \delta}(\delta; 0) = \mathrm{diag}(\sigma(\delta; 0)) - \sigma(\delta; 0)\sigma(\delta; 0)^T.$$

By the Sherman-Morrison formula,

$$\begin{aligned}
\left[\frac{\partial \sigma}{\partial \delta}\right]^{-1} &= \mathrm{diag}(1/\sigma_1, \ldots, 1/\sigma_J) + \frac{\mathrm{diag}(1/\sigma_1, \ldots, 1/\sigma_J)\sigma\sigma^T\mathrm{diag}(1/\sigma_1, \ldots, 1/\sigma_J)}{1 - \sigma^T\mathrm{diag}(1/\sigma_1, \ldots, 1/\sigma_J)\sigma} \\
&= \mathrm{diag}(1/\sigma_1, \ldots, 1/\sigma_J) + \frac{\mathbf{1}\mathbf{1}^T}{1 - \sum_{k=1}^J \sigma_k}
\end{aligned} \tag{57}$$

where $\mathbf{1}$ is a $J \times 1$ vector of ones.

Next, we calculate

$$\frac{\partial^2 \sigma_j}{\partial \lambda^2}(\delta_c; \lambda) = \int \nu_i^2[(p_{jc} - \bar{p}_{ic})^2 - (\overline{p^2}_{ic} - \bar{p}^2_{ic})]\sigma_j(\delta_c + \lambda\nu_i p_c; 0) \, \mathrm{d}\Phi(\nu_i)$$

and thus

$$\begin{aligned}\frac{\partial^2\sigma_j}{\partial\lambda^2}(\sigma^{-1}(s_c;0);0) &= [(p_{jc}-\bar{p}_c)^2-(\overline{p^2}_c-\bar{p}_c^2)]\sigma_j(\sigma^{-1}(s_c;0);0)\int\nu_i^2\,\mathrm{d}\Phi(\nu_i)\\ &= [(p_{jc}-\bar{p}_c)^2-(\overline{p^2}_c-\bar{p}_c^2)]s_{jc}\\ &= [p_{jc}^2-2\bar{p}_cp_{jc}+2\bar{p}_c^2-\overline{p^2}_c]s_{jc} \qquad (58)\end{aligned}$$

where $\overline{p_c^2}=\sum_{k=1}^J p_{kc}^2 s_{kc}$ and $\bar{p}_c=\sum_{k=1}^J p_{kc}s_{kc}$.

Putting it all together. Summing the expression for $\frac{\partial^2\sigma_j}{\partial\lambda^2}(\sigma^{-1}(s_c;0);0)$ over j yields

$$\sum_{j=1}^J[p_{jc}^2-2\bar{p}_cp_{jc}+2\bar{p}_c^2-\overline{p^2}_c]s_{jc}=\overline{p^2}_c-2\bar{p}_c^2+[2\bar{p}_c^2-\overline{p^2}_c]\left(\sum_{j=1}^J s_{jc}\right)=-[2\bar{p}_c^2-\overline{p_c^2}]s_{0c}. \quad (59)$$

Finally, plugging (57) and (58) into (56), setting $\lambda=0$ and $\delta=\sigma^{-1}(s_c;0)$, and using (59), we have

$$\begin{aligned}-\frac{\partial^2\sigma_j^{-1}}{\partial\lambda^2}(s_c;0) &= \frac{[p_{jc}^2-2\bar{p}_cp_{jc}+2\bar{p}_c^2-\overline{p^2}_c]s_{jc}}{s_{jc}}+\frac{-[2\bar{p}_c^2-\overline{p_c^2}]s_{0c}}{s_{0c}}\\ &= p_{jc}(p_{jc}-2\bar{p}_c).\end{aligned}$$

□

A.7 Lemma 5

Lemma 5. *Consider a sequence of DGPs, indexed by $n\in\mathbb{N}$, for which $\lambda^{(n)}\to 0$, where $\left(\beta_0^{(n)},\mu^{(n)},\lambda^{(n)}\right)'$ is the solution to the nonlinear moment condition (31) for the nth DGP (which is assumed to exist and to be unique). Suppose that there exist constants $\varepsilon>0, B>0$, and $N\in\mathbb{N}$ such that*

$$\frac{\|\,\mathbb{E}[\frac{\partial^3 g}{\partial\lambda^3}(0,0,\bar{\lambda};p_c,s_c,z_c)]\|_2}{\sigma_{\min}(D)}\le B$$

for all $\bar{\lambda}\in(-\varepsilon,\varepsilon)$ and $n\ge N$, where

$$\begin{aligned}g(\beta_0,\mu,\lambda;p_c,s_c,z_c) &\equiv J^{-1}\sum_{j=1}^J z_{jc}(\sigma_j^{-1}(s_c;\lambda)-\beta_0-\mu p_{jc})\\ D &\equiv \mathbb{E}\left(\frac{\partial g}{\partial\beta_0}(0,0,0;p_c,s_c,z_c)\quad \frac{\partial g}{\partial\beta_0}(0,0,0;p_c,s_c,z_c)\quad \frac{1}{2}\frac{\partial^2 g}{\partial\lambda^2}(0,0,0;p_c,s_c,z_c)\right)\end{aligned}$$

and $\sigma_{\min}(D)$ denotes the smallest singular value of D. Then

$$\left|\frac{(\tilde{\lambda}^{(n)})^2-(\lambda^{(n)})^2}{(\lambda^{(n)})^2}\right|=O(|\lambda^{(n)}|)\to 0$$

where $\left(\tilde{\beta}_0^{(n)},\tilde{\mu}^{(n)},\tilde{\lambda}^{(n)}\right)'$ is the solution to the linear moment condition (37) for the nth DGP.

Proof. Since a unique solution to the moment condition (31) exists by assumption, we assume (possibly by choosing a subset of instruments) that the parameters are just identified, i.e., that z_{jc} is 3×1.

Using Taylor's theorem, we expand

$$g(\beta_0, \mu, \lambda; p_c, s_c, z_c) = g(0,0,0; p_c, s_c, z_c) + \frac{\partial g}{\partial \beta_0}(\beta_0, \mu, 0; p_c, s_c, z_c)\beta_0 + \frac{\partial g}{\partial \mu}(\beta_0, \mu, 0; p_c, s_c, z_c)\mu + \frac{1}{2}\frac{\partial^2 g}{\partial \lambda^2}(0,0,0; p_c, s_c, z_c)\lambda^2 + \frac{1}{6}\frac{\partial^3 g}{\partial \lambda^3}(0,0,\bar{\lambda}; p_c, s_c, z_c)\lambda^3$$

for some $\bar{\lambda} \in [0, \lambda]$ (which may depend on λ and c). We can then write the moment conditions (31) and (37) as, respectively,

$$0 = \mathbb{E}\left[g(\beta_0^{(n)}, \mu^{(n)}, 0; p_c, s_c, z_c) + \frac{1}{2}\frac{\partial^2 g}{\partial \lambda^2}(0,0,0; p_c, s_c, z_c)(\lambda^{(n)})^2 + \frac{1}{6}\frac{\partial^3 g}{\partial \lambda^3}(0,0,\bar{\lambda}; p_c, s_c, z_c)(\lambda^{(n)})^3 \right]$$

$$0 = \mathbb{E}\left[g(\tilde{\beta}_0^{(n)}, \tilde{\mu}^{(n)}, 0; p_c, s_c, z_c) + \frac{1}{2}\frac{\partial^2 g}{\partial \lambda^2}(0,0,0; p_c, s_c, z_c)(\tilde{\lambda}^{(n)})^2 \right].$$

Taking the difference between these two equations and rearranging, we obtain

$$\begin{pmatrix} \beta_0^{(n)} \\ \mu^{(n)} \\ (\lambda^{(n)})^2 \end{pmatrix} - \begin{pmatrix} \tilde{\beta}_0^{(n)} \\ \tilde{\mu}^{(n)} \\ (\tilde{\lambda}^{(n)})^2 \end{pmatrix} = D^{-1}\,\mathbb{E}\left[\frac{1}{6}\frac{\partial^3 g}{\partial \lambda^3}(0,0,\bar{\lambda}; p_c, s_c, z_c)\right](\lambda^{(n)})^3$$

where

$$D \equiv \mathbb{E}\left(\frac{\partial g}{\partial \beta_0}(0,0,0; p_c, s_c, z_c) \quad \frac{\partial g}{\partial \beta_0}(0,0,0; p_c, s_c, z_c) \quad \frac{1}{2}\frac{\partial^2 g}{\partial \lambda^2}(0,0,0; p_c, s_c, z_c) \right).$$

Taking 2-norms of both sides, applying standard facts from linear algebra,[15] and dividing by $|\lambda^{(n)}|^2$, we obtain

$$\left| \frac{(\lambda^{(n)})^2 - (\tilde{\lambda}^{(n)})^2}{(\lambda^{(n)})^2} \right| \leq \frac{1}{6}\frac{\| \mathbb{E}[\frac{\partial^3 g}{\partial \lambda^3}(0,0,\bar{\lambda}; p_c, s_c, z_c)] \|_2}{\sigma_{\min}(D)} |\lambda^{(n)}|.$$

Under our assumptions, the factor multiplying $|\lambda^{(n)}|$ on the right-hand side of the inequality is eventually bounded as $n \to \infty$, and thus the right-hand side converges to 0. □

The calculations in the lemma also shed light on when the error in the epi-convergence result is likely to be small or large. First, the error will be smaller the smaller is

$$\mathbb{E}\left[\frac{\partial^3 g}{\partial \lambda^3}(0,0,\bar{\lambda}; p_c, s_c, z_c)\right] = \mathbb{E}\left[J^{-1} \sum_{j=1}^{J} z_{jc} \frac{\partial^3 \sigma_j^{-1}}{\partial \lambda^3}(s_c; \bar{\lambda}) \right].$$

[15] The facts are

1. $\|Ax\|_2 \leq \|A\|_2 \|x\|_2$
2. $\|A\|_2 = \sigma_{\max}(A)$
3. $\sigma_{\max}(A) = 1/\sigma_{\min}(A^{-1})$ if A is invertible.

Results in the appendix of Salanié and Wolak (2022) show that $\frac{\partial^3 \sigma_j^{-1}}{\partial \lambda^3}(s_c; \bar{\lambda})$ is a cubic function of prices. The error in the epi-convergence result will be smaller the lower is the correlation between the instruments z_{jc} and this particular cubic function of prices. Second, the error will be larger the smaller is the smallest singular value of D. This singular value will be small when the projections of p_{jc} and $\frac{1}{2}p_{jc}(p_{jc} - 2\bar{p}_c)$ onto the instruments z_{jc} are highly correlated; thus the epi-convergence result will be more accurate the more the instruments vary these two terms in a non-collinear way.

A.8 Proof of (35)

It suffices to show

$$\frac{\mathbb{E}[p_{jc}^2/2 \mid \dot{z}_{jc} = 1] - \mathbb{E}[p_{jc}^2/2 \mid \dot{z}_{jc} = 0]}{\mathbb{E}[p_{jc} \mid \dot{z}_{jc} = 1] - \mathbb{E}[p_{jc} \mid \dot{z}_{jc} = 0]} = \int_0^\infty p_j \omega_j^p(p_j)\, \mathrm{d}p_j.$$

The proof follows that of Lemma 5 in Angrist et al. (2000). Let $p_{jc}(z)$ denote the counterfactual price that would be realized in market c for product j if $\dot{z}_{jc} = z$. Since $\dot{z}_{jc}$ is randomized,

$$\mathbb{E}[p_{jc}^2/2 \mid \dot{z}_{jc} = 1] - \mathbb{E}[p_{jc}^2/2 \mid \dot{z}_{jc} = 0] = \mathbb{E}[p_{jc}(1)^2/2 - p_{jc}(0)^2/2].$$

By the fundamental theorem of calculus,

$$\begin{aligned}
\mathbb{E}[p_{jc}(1)^2/2 - p_{jc}(0)^2/2] &= \mathbb{E}\left[\int_{p_{jc}(0)}^{p_{jc}(1)} p_j\, \mathrm{d}p_j\right] \\
&= \mathbb{E}\left[\int_0^{p_{jc}(1)} p_j\, \mathrm{d}p_j\right] - \mathbb{E}\left[\int_0^{p_{jc}(0)} p_j\, \mathrm{d}p_j\right] \\
&= \mathbb{E}\left[\int_0^\infty p_j \mathbb{1}\{p_{jc}(1) \geq p_j\}\, \mathrm{d}p_j\right] - \mathbb{E}\left[\int_0^\infty p_j \mathbb{1}\{p_{jc}(0) \geq p_j\}\, \mathrm{d}p_j\right] \\
&= \int_0^\infty p_j \big(\mathbb{P}\{p_{jc}(1) \geq p_j\} - \mathbb{P}\{p_{jc}(0) \geq p_j\}\big)\, \mathrm{d}p_j.
\end{aligned}$$

A similar argument shows

$$\mathbb{E}[p_{jc} \mid \dot{z}_{jc} = 1] - \mathbb{E}[p_{jc} \mid \dot{z}_{jc} = 0] = \int_0^\infty \big(\mathbb{P}\{p_{jc}(1) \geq p_j\} - \mathbb{P}\{p_{jc}(0) \geq p_j\}\big)\, \mathrm{d}p_j,$$

whence the result follows.

A.9 Proof of Proposition 10

Applications of lemmas. In light of Lemmas 4 and 5, we focus on analyzing the IV parameter vector

$$\mathbb{E}\left[J^{-1}\sum_{j=1}^J z_{jc}\left(1, -p_j, \frac{1}{2}p_{jc}(p_{jc} - \bar{p}_c)\right)\right]^{-1} \mathbb{E}\left[J^{-1}\sum_{j=1}^J z_{jc} \log(q_{jc}/q_{0c})\right].$$

First-stage fitted values. We analyze the IV regression as two-stage least squares. We first solve for the first stage projections onto z_{jc}. To simplify notation, we assume $\mathbb{P}[\dot{z}_{jc} = 1] = \mathbb{P}[\dot{z}_{jc} = 0] = 1/2$. In this case, each z_{jck} takes the values $-1/2, 0$, and $1/2$, and the coefficient from a population OLS regression of p_{jc} on z_{jck} is

$$\begin{aligned}\frac{\mathbb{E}[J^{-1}\sum_{j=1}^{J} z_{jck}p_{jc}]}{\mathbb{E}[J^{-1}\sum_{j=1}^{J} z_{jck}^2]} &= \frac{\mathbb{E}[J^{-1}z_{kck}p_{kc}]}{\mathbb{E}[J^{-1}z_{kck}^2]} \\ &= \frac{(1/2)(1/2)\overline{p_k}(1) - (1/2)(1/2)\overline{p_k}(0)}{(1/2)^2} \\ &= \overline{p_k}(1) - \overline{p_k}(0)\end{aligned}$$

where the first equality used that $z_{jck} = 0$ if $j \neq k$ and we defined $\overline{p_k}(1) \equiv \mathbb{E}[p_{kc}|z_{kck} = 1/2]$ and $\overline{p_k}(0) \equiv \mathbb{E}[p_{kc}|z_{kck} = -1/2]$ in the second equality. Because the components of z_{jc} are mutually orthogonal, the fitted values from a projection of the endogenous regressor p_{jc} onto the instrument vector z_{jc} are

$$\begin{aligned}\hat{p}_{jc} &= \sum_{k=1}^{J} z_{jck}[\overline{p_k}(1) - \overline{p_k}(0)] \\ &= z_{jcj}[\overline{p_j}(1) - \overline{p_j}(0)] \\ &= \begin{cases}(1/2)[\overline{p_j}(1) - \overline{p_j}(0)] & \text{if } z_{jcj} = 1/2 \\ -(1/2)[\overline{p_j}(1) - \overline{p_j}(0)] & \text{if } z_{jcj} = -1/2.\end{cases}\end{aligned}$$

Analogously,

$$\begin{aligned}\hat{r}_{jc} &= \begin{cases}(1/2)[\overline{r_j}(1) - \overline{r_j}(0)] & \text{if } z_{jcj} = 1/2 \\ -(1/2)[\overline{r_j}(1) - \overline{r_j}(0)] & \text{if } z_{jcj} = -1/2,\end{cases} \\ \widehat{\log(q/q_0)}_{jc} &= \begin{cases}(1/2)[\overline{\log(q/q_0)_j}(1) - \overline{\log(q/q_0)_j}(0)] & \text{if } z_{jcj} = 1/2 \\ -(1/2)[\overline{\log(q/q_0)_j}(1) - \overline{\log(q/q_0)_j}(0)] & \text{if } z_{jcj} = -1/2,\end{cases}\end{aligned}$$

where $r_{jc} \equiv \frac{1}{2}p_{jc}(p_{jc} - 2\bar{p}_c)$.

Second stage. The second stage coefficients from a pooled regression of $\widehat{\log(q/q_0)}_{jc}$ on $\hat{p}_{jc}$ and $\hat{r}_{jc}$ are

$$\begin{aligned}\begin{pmatrix}-\mu \\ \lambda^2/2\end{pmatrix} &= \begin{pmatrix}\mathbb{E}[J^{-1}\sum_{j=1}^{J}\hat{p}_{jc}^2] & \mathbb{E}[J^{-1}\sum_{j=1}^{J}\hat{p}_{jc}\hat{r}_{jc}] \\ \mathbb{E}[J^{-1}\sum_{j=1}^{J}\hat{p}_{jc}\hat{r}_{jc}] & \mathbb{E}[J^{-1}\sum_{j=1}^{J}\hat{r}_{jc}^2]\end{pmatrix}^{-1}\begin{pmatrix}\mathbb{E}[J^{-1}\sum_{j=1}^{J}\hat{p}_{jc}\widehat{\log q}_{jc}] \\ \mathbb{E}[J^{-1}\sum_{j=1}^{J}\hat{r}_{jc}\widehat{\log q}_{jc}]\end{pmatrix} \\ &= \begin{pmatrix}\sum_{j=1}^{J}[\overline{p_j}(1) - \overline{p_j}(0)]^2 & \sum_{j=1}^{J}[\overline{p_j}(1) - \overline{p_j}(0)][\overline{r_j}(1) - \overline{r_j}(0)] \\ \sum_{j=1}^{J}[\overline{p_j}(1) - \overline{p_j}(0)][\overline{r_j}(1) - \overline{r_j}(0)] & \sum_{j=1}^{J}[\overline{r_j}(1) - \overline{r_j}(0)]^2\end{pmatrix}^{-1} \\ &\quad \cdot \begin{pmatrix}\sum_{j=1}^{J}[\overline{p_j}(1) - \overline{p_j}(0)][\overline{\log(q/q_0)_j}(1) - \overline{\log(q/q_0)_j}(0)] \\ \sum_{j=1}^{J}[\overline{r_j}(1) - \overline{r_j}(0)][\overline{\log(q/q_0)_j}(1) - \overline{\log(q/q_0)_j}(0)]\end{pmatrix}.\end{aligned} \tag{60}$$

These are the coefficients in a product-level OLS regression of $\Delta \log q_j \equiv [\overline{\log(q/q_0)_j}(1) - \overline{\log(q/q_0)_j}(0)]$ on $\Delta p_j \equiv [\overline{p_j}(1) - \overline{p_j}(0)]$ and $\Delta r_j \equiv [\overline{r_j}(1) - \overline{r_j}(0)]$ (and no constant). These coefficients exist if and only if neither Δp_j nor Δr_j is identically 0 and Δp_j and Δr_j are not perfectly correlated across products j.

WLS formulation. We have that the parameter vector $(-\mu, \lambda^2)'$ is equal to the vector of coefficients from the product-level OLS regression of

$$\Delta \log(q/q_0)_j \texttt{ on } \Delta p_j, \Delta r_j.$$

The coefficients from this regression are numerically identical to the coefficients from the weighted least-squares regression

$$\frac{\Delta \log(q/q_0)_j}{\Delta p_j} \texttt{ on } \frac{\Delta p_j}{\Delta p_j}, \frac{\Delta r_j}{\Delta p_j}$$

with weights $(\Delta p_j)^2$. $\Delta p_j / \Delta p_j = 1$, so we conclude that the parameters $(-\mathbb{E}[\alpha_i] - \bar{p}\lambda^2, \lambda^2)$ are given by a $(\Delta p_j)^2$-weighted least squares regression of the product-specific Wald estimands $\frac{\Delta \log(q/q_0)_j}{\Delta p_j} = \frac{\overline{\log(q/q_0)_j}(1) - \overline{\log(q/q_0)_j}(0)}{\overline{p_j}(1) - \overline{p_j}(0)}$ on a constant and $\frac{\Delta r_j}{\Delta p_j}$:

$$\frac{\Delta \log(q/q_0)_j}{\Delta p_j} \texttt{ on } 1, \frac{\Delta r_j}{\Delta p_j}.$$

A.10 Proof of Proposition 11

We show the following Lemma, which we use to linearize the moment condition (39). The rest of the argument then follows the proof of Proposition 10.

Lemma 6.

$$\frac{\partial \sigma^{-1}}{\partial \pi}(s_c, F_{dc}; 0) = -p_{jc}\bar{d}_c \tag{61}$$

where $\bar{d}_c \equiv \int d_{ic} \, \mathrm{d}F_{dc}(d_{ic})$.

Proof. The implicit and inverse function theorems give

$$-\frac{\partial \sigma^{-1}}{\partial \pi}(\sigma(\delta, F; \pi), F; \pi) = \left[\frac{\partial \sigma}{\partial \delta}(\delta, F; \pi)\right]^{-1} \frac{\partial \sigma}{\partial \pi}(\delta, F; \pi).$$

Direct calculation yields

$$\frac{\partial \sigma_j}{\partial \pi}(\delta_c, F_c; \pi) = \int d_{ic}(p_{jc} - \bar{p}_{ic}) \sigma_j(\delta_c + \pi d_{ic} p_c, F_c; 0) \, \mathrm{d}F(d_{ic}).$$

At $\pi = 0$,

$$\frac{\partial \sigma_j}{\partial \pi}(\delta_c, F_c; 0) = (p_{jc} - \bar{p}_c)\sigma_j(\delta_c, F_c; 0)\bar{d}_c$$

where $\bar{d}_c \equiv \int d_{ic}\,\mathrm{d}F_c(d_{ic})$. (57) continues to hold, so we have

$$\begin{aligned}
-\frac{\partial \sigma_j^{-1}}{\partial \pi}(s_c, F; 0) &= \left(\left[\frac{\partial \sigma}{\partial \delta}(\sigma^{-1}(s_c, F_c; 0), F; 0) \right]^{-1} \frac{\partial \sigma}{\partial \pi}(\sigma^{-1}(s_c, F_c; 0), F; 0) \right)_j \\
&= \frac{(p_{jc} - \bar{p}_c) s_{jc} \bar{d}_c}{s_{jc}} + \frac{s_{0c} \bar{p}_c \bar{d}_c}{s_{0c}} \\
&= p_{jc} \bar{d}_c.
\end{aligned}$$

□

A.11 Proof of Proposition 12

Approximate inversion of market shares. An approximation of the share inversion function for the model (20) follows easily from the results in Salanié and Wolak (2022). To simplify notation, we suppose β_i is R-dimensional, define $x_{j0c} \equiv -p_{jc}$, and let $\mu \equiv \mathbb{E}[(\alpha_i, \beta_i')']$ and $\Lambda \equiv \mathrm{Var}[(\alpha_i, \beta_i')']$. We have the following analogue to Lemma 4:

Lemma 7.

$$\begin{aligned}
&\sigma_j^{-1}(s_c; \Lambda) - \log(s_{jc}/s_{0c}) \\
&\qquad = -\frac{1}{2} \sum_{r=0}^{R} \sum_{s=0}^{R} x_{jrc}(x_{jsc} - 2\bar{x}_{sc}) \Lambda_{rs} + o(\|\Lambda\|) \\
&\qquad = -\frac{1}{2} \sum_{r=0}^{R} x_{jrc}(x_{jrc} - 2\bar{x}_{rc}) \Lambda_{rr} - \sum_{0 \le r < s \le R} [x_{jrc} x_{jsc} - x_{jrc} \bar{x}_{sc} - x_{jsc} \bar{x}_{rc}] \Lambda_{rs} + o(\|\Lambda\|).
\end{aligned}$$

Proof. Follows from Theorem 2 in Salanié and Wolak (2022). □

By this Lemma and a generalization of Lemma 5, the BLP estimator of (μ, Λ) can be approximated by the IV regression

$$\begin{aligned}
\log(s_{jc}/s_{0c}) = -\mu_0 p_{jc} + x_{jc}' \mu_{1:R} &+ \sum_{r=0}^{R} \frac{1}{2} x_{jrc}(x_{jrc} - 2\bar{x}_{rc}) \Lambda_{rr} \\
&+ \sum_{0 \le r < s \le R} [x_{jrc} x_{jsc} - x_{jrc} \bar{x}_{sc} - x_{jsc} \bar{x}_{rc}] \Lambda_{rs} + \xi_{jc}
\end{aligned}$$

using instruments $1, z_{jc}, z_{jc}^x$. In particular, the covariance $\mathrm{Cov}[\alpha_i, \beta_{ir}]$ is approximately the regression coefficient on the interaction between price and the corresponding product characteristic:

$$p_{jc} x_{jrc} - p_{jc} \bar{x}_{rc} - x_{jrc} \bar{p}_c = (p_{jc} - \bar{p}_c)(x_{jrc} - \bar{x}_{rc}) - \bar{p}_c \bar{x}_{rc}.$$

Under the block-diagonal assumption, the parameters μ_0 and Λ_{0r} $(0 \le r \le R)$ in the IV regression above are the same as those in the IV regression

$$\log(s_{jc}/s_{0c}) = -\mu_\alpha p_{jc} + \frac{1}{2} p_{jc}(p_{jc} - 2\bar{p}_c) \Lambda_{0,0} + \sum_{r=1}^{R} (p_{jc} x_{jrc} - p_{jc} \bar{x}_{rc} - x_{jrc} \bar{p}_c) \Lambda_{0,r} + \xi_{jc}.$$

From here, an argument similar to that employed in the proof of Proposition 10 shows that $(-\mu_0, \Lambda_{0,0}, \Lambda'_{0,1:R}) = (\gamma_1, \gamma_p, \gamma'_x)$, the coefficients in the ω_j-weighted least squares regression

$$\rho_j^{\log(q/q_0)} = \gamma_1 + \gamma_p \rho_j^{\frac{1}{2}p(p-2\bar{p})} + \sum_{r=1}^{R} \gamma_{x_r} \rho_j^{px_r - p\bar{x}_r - x_r\bar{p}} + u_j.$$

A.12 Proof of Proposition 13

Since z_{jc} is pooled orthonormal, the pooled projected values of p_{jc} and $\log(q_{jc}/q_{0c})$ on z_{jc} (after residualizing out the constant) are given by

$$\begin{aligned}
\hat{p}_{jc} &= \sum_{\ell=1}^{L} \omega_\ell (z_{jc\ell} - \bar{z}_\ell) \\
\widehat{\log(q_{jc}/q_{0c})} &= \sum_{\ell=1}^{L} \mathbb{E}\left[J^{-1} \sum_{j'=1}^{J} (z_{j'c'\ell} - \bar{z}_\ell) \log(q_{j'c'}/q_{0c'}) \right] (z_{jc\ell} - \bar{z}_\ell).
\end{aligned}$$

The 2SLS estimand $-\alpha$ is then

$$\begin{aligned}
-\alpha &= \frac{\mathbb{E}[J^{-1} \sum_{j=1}^{J} \hat{p}_{jc} \widehat{\log(q_{jc}/q_{0c})}]}{\mathbb{E}[J^{-1} \sum_{j=1}^{J} \hat{p}_{jc}^2]} \\
&= \frac{\sum_{\ell=1}^{L} \omega_\ell \, \mathbb{E}\left[J^{-1} \sum_{j'=1}^{J} (z_{j'c'\ell} - \bar{z}_\ell) \log(q_{j'c'}/q_{0c'}) \right]}{\sum_{\ell=1}^{L} \omega_\ell^2},
\end{aligned}$$

where the second line uses the orthonormality of z_{jc}. Substituting in the definition of $\rho_\ell^{\log(q/q_0)}$, we have

$$-\alpha = \sum_{\ell=1}^{L} \frac{\omega_\ell^2}{\sum_{\ell'=1}^{L} \omega_{\ell'}^2} \rho_\ell^{\log(q/q_0)}.$$

A.13 Proof of Proposition 14

After residualizing out the constant, the fitted values from the pooled first-stage regression of $\mathbb{1}[j=k]p_{jc}$ onto z_{jc} are

$$\begin{aligned}
(\widehat{\mathbb{1}[j'=k]p})_{j'c'} &= \sum_{\ell=1}^{L} \mathbb{E}\left[J^{-1} \sum_{j=1}^{J} (z_{jc\ell} - \bar{z}_\ell) \mathbb{1}[j=k] p_{jc} \right] (z_{j'c'\ell} - \bar{z}_\ell) \\
&= \sum_{\ell=1}^{L} \mathbb{E}\left[J^{-1} (z_{kc\ell} - \bar{z}_\ell) \right] (z_{j'c'\ell} - \bar{z}_\ell).
\end{aligned}$$

The second-stage regression is a pooled regression of $(\widehat{\log(q_j/q_0)})_{jc}$ on $(\widehat{\mathbb{1}[j=1]p})_{jc}, \ldots, (\widehat{\mathbb{1}[j=J]p})_{jc}$. Using the pooled orthonormality of z_{jc}, we have

$$\mathbb{E}\left[J^{-1} \sum_{j=1}^{J} (\widehat{\mathbb{1}[j=k]p})_{jc} (\widehat{\mathbb{1}[j=k']p})_{jc} \right] = \sum_{\ell=1}^{L} \mathbb{E}[J^{-1}(z_{kc\ell} - \bar{z}_\ell) p_{kc}] \, \mathbb{E}[J^{-1}(z_{k'c\ell} - \bar{z}_\ell) p_{k'c}]$$

and

$$\mathbb{E}\left[J^{-1}\sum_{j=1}^{J}(\widehat{\mathbb{1}[j=k]p})_{jc}(\widehat{\log(q_j/q_0)})_{jc}\right] = \sum_{\ell=1}^{L}\mathbb{E}[J^{-1}(z_{kc\ell}-\bar{z}_\ell)p_{kc}]\,\mathbb{E}\left[J^{-1}\sum_{j'=1}^{J}(z_{j'c\ell}-\bar{z}_\ell)\log(q_{j'c}/q_{0c})\right].$$

Together, these two results establish that the coefficients in the second-stage regression of $(\widehat{\log(q_j/q_0)})_{jc}$ on $(\widehat{\mathbb{1}[j=1]p})_{jc},\ldots,(\widehat{\mathbb{1}[j=J]p})_{jc}$ are identical to the coefficients in the following ℓ-level OLS regression:

$$\mathbb{E}\left[J^{-1}\sum_{j'=1}^{J}(z_{j'c\ell}-\bar{z}_\ell)\log(q_{j'c}/q_{0c})\right] = \sum_{k=1}^{J}\alpha_k\,\mathbb{E}[J^{-1}(z_{kc\ell}-\bar{z}_\ell)p_{kc}] + u_\ell.$$

The coefficients in this OLS regression are in turn identical to the coefficients in a ω_ℓ^2-weighted ℓ-level WLS regression of $\rho_\ell^{\log(q/q_0)}$ on $\frac{\mathbb{E}[J^{-1}(z_{1c\ell}-\bar{z}_\ell)p_{1c}]}{\mathbb{E}[J^{-1}\sum_{k'=1}^{J}(z_{k'c\ell}-\bar{z}_\ell)p_{k'c}]},\ldots,\frac{\mathbb{E}[J^{-1}(z_{Jc\ell}-\bar{z}_\ell)p_{Jc}]}{\mathbb{E}[J^{-1}\sum_{k'=1}^{J}(z_{k'c\ell}-\bar{z}_\ell)p_{k'c}]}$, as was to be shown.

A.14 Instruments for synthetic cereal data

To apply the results in Section 4 to the synthetic cereal DGPs in Section 5, we construct J product-specific price instruments $\hat{z}_{jc} = (\hat{z}_{jc1},\ldots,\hat{z}_{jcJ})'$ to use in estimation. For each product $j = 1,\ldots,J$, we regress

$$p_{jc} = \pi_{0j} + \pi_{1j}'z_{jc} + u_{jc}$$

where p_{jc} is the price and $z_{jc} = (z_{jc1},\ldots,z_{jcL})'$ is the vector of instruments provided in the original Nevo (2000) data. We then set

$$\hat{z}_{jck} = \begin{cases}\hat{\pi}_{1j}'\left(z_{jc} - C^{-1}\sum_{c'=1}^{C}z_{jc'}\right) & k = j\\ 0 & \text{else.}\end{cases}$$

Intuitively, the kth constructed instrument $\hat{z}_{jck}$ is the price of the kth product projected onto the original instruments. The validity of these instruments $\hat{z}_{jc} = (\hat{z}_{jc1},\ldots,\hat{z}_{jcJ})'$ follows immediately from the validity of the original instruments $z_{jc} = (z_{jc1},\ldots,z_{jcL})'$, since $\hat{z}_{jc}$ consists of linear combinations of the original instruments (interacted with product fixed effects). Moreover, these instruments are pooled orthogonal by construction. Thus the construction provides J product-specific price instruments that are similar to those considered in Section 4.

A.15 Proof of Proposition 18

Proof of Proposition 18, case (c). By Corollary 3,

$$\operatorname*{plim}_{C\to\infty}\hat{\eta}_{j'j'c'}(p_{c'}) = \left[\gamma_0 + \gamma_1\left(p_{j'c'} - \bar{p}_{c'}\right)\right]p_{j'c'} \;+\; O(|\lambda|^3) + O(s_{j'c'}),$$

where (γ_0, γ_1) are the ω_j-weighted least squares (WLS) coefficients from (33):

$$\rho_j^{\log(q/q_0)} \;=\; \gamma_0 + \gamma_1\,\tilde{p}_j + u_j, \qquad \sum_{j=1}^{J} \omega_j \begin{pmatrix} 1 \\ \tilde{p}_j \end{pmatrix} u_j = 0.$$

When market shares s_{jc} are small, the constructed regressor $\tilde{p}_j$ is approximately equal to $p_{jc} - \overline{p}_c$. We therefore work with $\tilde{p}_j \approx p_{jc} - \overline{p}_c$.

Define the ω-weighted mean and variance of the centered regressor,

$$\mu_\omega \;\equiv\; \sum_{j=1}^{J} \omega_j (p_{jc} - \overline{p}_c), \qquad \sigma_\omega^2 \;\equiv\; \sum_{j=1}^{J} \omega_j \big[(p_{jc} - \overline{p}_c) - \mu_\omega\big]^2,$$

using that $\sum_j \omega_j = 1$. The expressions for coefficients of a WLS regression imply

$$\gamma_1 = \frac{\sum_j \omega_j \big[(p_{jc} - \overline{p}_c) - \mu_\omega\big]\, \rho_j^{\log(q/q_0)}}{\sigma_\omega^2}, \qquad \gamma_0 = \sum_j \omega_j \rho_j^{\log(q/q_0)} - \gamma_1 \mu_\omega.$$

Therefore the WLS predictor at $p_{j'c'} - \overline{p}_{c'}$ is

$$\begin{aligned}
\gamma_0 + \gamma_1 (p_{j'c'} - \overline{p}_{c'}) &= \sum_{j=1}^{J} \omega_j \rho_j^{\log(q/q_0)} + \frac{p_{j'c'} - \overline{p}_{c'} - \mu_\omega}{\sigma_\omega^2} \sum_{j=1}^{J} \omega_j \big(p_{jc} - \overline{p}_c - \mu_\omega\big)\, \rho_j^{\log(q/q_0)} \\
&= \sum_{j=1}^{J} \omega_j \left(1 + \frac{(p_{j'c'} - \overline{p}_{c'} - \mu_\omega)\,\big(p_{jc} - \overline{p}_c - \mu_\omega\big)}{\sigma_\omega^2}\right) \rho_j^{\log(q/q_0)}, \\
&= \sum_{j=1}^{J} \Psi_{j'c'}(j, p_{j'c'}) \rho_j^{\log(q/q_0)},
\end{aligned}$$

where $\sum_j \Psi_{j'c'}(j, p_{j'c'}) = 1$.
Substituting the Angrist-Graddy–Imbens result and plugging into Corollary 3, we obtain the desired expression.

Sign of weights. Because $\Psi_{j'c'}(j, p_{j'c'})$ contains the covariance adjustment

$$\frac{(p_{j'c'} - \overline{p}_{c'} - \mu_\omega)\,(p_{jc} - \overline{p}_c - \mu_\omega)}{\sigma_\omega^2},$$

the weights can be negative whenever the price deviation of product j tends to have the opposite sign of the focal product's deviation. □

B An economic motivation of non-separability

Non-separability of the outside option can be motivated by allowing choices of inside product to vary in their complementarity with some outside option. For example, suppose the agent

makes a discrete choice of car $j = 1, \ldots, J$, a continuous choice of gallons of gasoline g, and a continuous choice of the remaining aggregate outside option q_0, with preferences

$$U_i(q, g) = \sum_j \tilde{\psi}_j q_j + \sum_j (\alpha_j - 1)\, q_j \log g + \log q_0. \tag{62}$$

Preferences are separable in the aggregate outside good q_0 but not between the car choice and gasoline consumption. Conditional on choosing car j at price p_j, the optimal choice of g (with p_0 and the gasoline price normalized to 1) is

$$g = (\alpha_j - 1)q_0 = \frac{\alpha_j - 1}{\alpha_j}(m - p_j).$$

Substituting this into the utility function in (62), and imposing the budget constraint $q_0 = m - \sum_j p_j q_j - g$, yields the indirect utility of purchasing good j:

$$v_{ij} = \psi_j + \alpha_j \log (m - p_j) \tag{63}$$

where $\psi_j \equiv \tilde{\psi}_j + \log \left[(\alpha_j - 1)^{(\alpha_j - 1)} \alpha_j^{-\alpha_j} \right]$. Thus the sensitivity to the price of j is given by:

$$\frac{\partial v_{ij}}{\partial p_j} = -\frac{\alpha_j}{m - p_j}$$

which is product-specific for two reasons. First, because α_j, or the degree of complementarity with gasoline, varies across inside goods. Second, because variation in p_j generates differences in price sensitivity via an income effect. While these income effects are negligible when prices are small relative to income, they may be important in settings such as the automobile market.

C Product-specific logit with variable income

We argue that the Section 4.4 results about the product-specific logit model in the constant-income case generalize to the variable income-case. Consider first the case where income varies between but not within markets. We have the following expression for the market share inversion:

$$\log(s_{jc}/s_{0c}) = \beta' x_{jc} + \alpha_j m_c - \alpha_j p_{jc} + \xi_{jc}.$$

This equation can be estimated using methods for interactive fixed effects (Bai, 2009) to flexibly control for $\alpha_j m_c$. If m_c is uncorrelated with the instruments z_{jc} for each product j, then the $\alpha_j m_c$ term can be ignored in estimation of α_j, and the result in Proposition 9 continues to apply.

We next consider the case where the within-market variance of income is nonzero. The market share inversion can not be computed in closed form in this case. When the within-market variance of income is small, however, an argument analogous to that developed in the following subsection leads to an approximate expression for the market share inversion:

$$\log(s_{jc}/s_{0c}) \approx \beta' x_{jc} + \alpha_j \mathbb{E}_{i\in c}[m_i] + \frac{1}{2}\alpha_j(\alpha_j - 2\bar{\alpha}_c)\operatorname{Var}_{i\in c}[m_i] - \alpha_j p_{jc} + \xi_{jc}$$

where $\bar{\alpha}_c \equiv \sum_{j=1}^{J} \alpha_j s_{jc}$. Again, this equation can be estimated using methods for interactive fixed effects to flexibly control for $\alpha_j \mathbb{E}_{i\in c}[m_i] + \frac{1}{2}\alpha_j(\alpha_j - 2\bar{\alpha}_c)\operatorname{Var}_{i\in c}[m_i]$. If $\mathbb{E}_{i\in c}[m_i]$, $\operatorname{Var}_{i\in c}[m_i]$, and $\bar{\alpha}_c \operatorname{Var}_{i\in c}[m_i]$ are all uncorrelated with z_{jc} for each product j, then these terms can be ignored in estimation of α_j, and the result in Proposition 9 continues to apply.

D Approximate interpretation and identification with $\log(m - p_j)$ utility

In this appendix we adapt the results in Sections 3 and 4 to the case where price and income enter utility via

$$v_{ijc} = \beta' x_{jc} + \xi_{jc} + \alpha_{ij} \log(m - p_{jc}) + \varepsilon_{ij},$$

To simplify exposition, we first consider a special case with common income m. In Section D.5, we derive the implications of allowing for heterogeneous incomes.

Similar to the main text, we consider three specifications: (i) a common coefficient α (i.e. logit), (ii) product-specific coefficients α_j (i.e. product-specific logit), (iii) random coefficients α_i (i.e. random coefficient logit).

Throughout, for each product j we define

$$L_j \equiv \log(m - p_j),$$

and for a given scalar α we denote the logit share of product k by $s_k(\alpha)$ and its share-weighted averages by

$$\bar{L}(\alpha) \equiv \sum_{k=1}^{J} L_k \, s_k(\alpha), \qquad \overline{L^2}(\alpha) \equiv \sum_{k=1}^{J} L_k^2 \, s_k(\alpha).$$

D.1 Small-share elasticities with $\log(m - p_j)$

We begin with a general small-share result (similar to Lemma 1) for the random coefficient logit model with $\log(m - p_j)$.

Lemma 8. *In the random coefficient logit model with indirect utility*

$$v_{ij} = \beta' x_{jc} + \xi_{jc} + \alpha_i \log(m - p_j) + \varepsilon_{ij}$$

and common income m, the own-price elasticity of demand for product j satisfies

$$\frac{\eta_{jj}(p)}{p_j} = -\frac{\mathbb{E}[\alpha_i]}{m - p_j} + \left(\frac{1}{m - p_j}\left(\operatorname{Var}[\alpha_i]\big(\bar{L}(\mathbb{E}[\alpha_i]) - L_j\big) + \operatorname{err}_1^L\right)\right)\frac{1}{1 + \operatorname{err}_2^L} + \operatorname{err}_3^L, \quad (64)$$

where

$$\mathrm{err}_1^L \equiv \frac{1}{2}\,\mathrm{Cov}\Big[\alpha_i,\ s_j''\big(t(\alpha_i)\big)\,\big(t(\alpha_i)-\mathbb{E}[\alpha_i]\big)^2\Big],$$

$$\mathrm{err}_2^L \equiv \frac{1}{2}\,\mathbb{E}\Big[s_j''\big(t(\alpha_i)\big)\,\big(t(\alpha_i)-\mathbb{E}[\alpha_i]\big)^2\Big]\Big/ s_j(\mathbb{E}[\alpha_i]),$$

$$\mathrm{err}_3^L \equiv \frac{1}{m-p_j}\,\frac{\mathbb{E}\big[\alpha_i\, s_j(\alpha_i)^2\big]}{\mathbb{E}\big[s_j(\alpha_i)\big]},$$

$t(\alpha_i)$ lies between α_i and $\mathbb{E}[\alpha_i]$ for each i, and

$$s_j''(\alpha) = \Big[(L_j-\bar{L}(\alpha))^2 - \big(\overline{L^2}(\alpha)-\bar{L}(\alpha)^2\big)\Big]\ s_j(\alpha).$$

Proof. For each i,

$$\frac{\partial v_{ij}}{\partial p_j} = -\frac{\alpha_i}{m-p_j},$$

so the individual logit derivative is

$$\frac{\partial s_j(\alpha_i)}{\partial p_j} = -\frac{\alpha_i}{m-p_j}\,s_j(\alpha_i)\big(1-s_j(\alpha_i)\big).$$

Aggregating over i,

$$\eta_{jj}(p) = -\,\frac{p_j}{\mathbb{E}[s_j(\alpha_i)]}\ \mathbb{E}\Big[\frac{\alpha_i}{m-p_j}\,s_j(\alpha_i)\big(1-s_j(\alpha_i)\big)\Big].$$

Factor out $1/(m-p_j)$ and use

$$\mathbb{E}[\alpha_i s_j(1-s_j)] = \mathrm{Cov}(\alpha_i, s_j) + \mathbb{E}[\alpha_i]\,\mathbb{E}[s_j] - \mathbb{E}[\alpha_i s_j^2]$$

to obtain

$$\frac{\eta_{jj}(p)}{p_j} = -\,\frac{\mathbb{E}[\alpha_i]}{m-p_j} \;-\; \frac{1}{m-p_j}\,\frac{\mathrm{Cov}(\alpha_i, s_j(\alpha_i))}{\mathbb{E}[s_j(\alpha_i)]} \;+\; \frac{1}{m-p_j}\,\frac{\mathbb{E}[\alpha_i s_j(\alpha_i)^2]}{\mathbb{E}[s_j(\alpha_i)]}.$$

To approximate the covariance term, apply a second-order Taylor expansion of $s_j(\alpha)$ around $\mathbb{E}[\alpha_i]$ (as in Appendix A.2):

$$\mathrm{Cov}(\alpha_i, s_j(\alpha_i)) = \mathrm{Var}[\alpha_i]\ s_j'(\mathbb{E}[\alpha_i]) + \mathrm{err}_1^L,$$

with

$$s_j'(\alpha) = (L_j - \bar{L}(\alpha))\,s_j(\alpha),$$

and err_1^L defined as in the statement via the mean-value form of the remainder using $s_j''(\cdot)$.

Similarly,

$$\mathbb{E}[s_j(\alpha_i)] = s_j(\mathbb{E}[\alpha_i])\big(1+\mathrm{err}_2^L\big),$$

with err_2^L as given, collecting the second-order remainder. Dividing,

$$\frac{\mathrm{Cov}(\alpha_i, s_j(\alpha_i))}{\mathbb{E}[s_j(\alpha_i)]} = \frac{\mathrm{Var}[\alpha_i]\,(L_j - \bar{L}(\mathbb{E}[\alpha_i]))\,s_j(\mathbb{E}[\alpha_i]) + \mathrm{err}_1^L}{s_j(\mathbb{E}[\alpha_i])\,(1+\mathrm{err}_2^L)} = \frac{\mathrm{Var}[\alpha_i]\,(L_j - \bar{L}(\mathbb{E}[\alpha_i])) + \mathrm{err}_1^L}{1+\mathrm{err}_2^L}.$$

Substituting back into the expression for $\eta_{jj}(p)/p_j$ yields (64). □

Lemma 8 also nests the simpler logit case and leads almost directly to the following small-share approximations.

Proposition 19 (Small-share approximations with $\log(m-p_j)$)**.** *Assume common income m.*

(a) In the logit model with indirect utility $v_{ij} = \beta' x_{jc} + \xi_{jc} + \alpha \log(m - p_j) + \varepsilon_{ij}$,

$$\begin{aligned}\eta_{jj}(p) &= -\frac{\alpha}{m-p_j} p_j (1 - s_j) \\ &= -\frac{\alpha}{m-p_j} p_j + O(s_j).\end{aligned} \tag{65}$$

(b) In the product-specific logit model with indirect utility $v_{ij} = \beta' x_{jc} + \xi_{jc} + \alpha_j \log(m - p_j) + \varepsilon_{ij}$,

$$\begin{aligned}\eta_{jj}(p) &= -\frac{\alpha_j}{m-p_j} p_j (1 - s_j) \\ &= -\frac{\alpha_j}{m-p_j} p_j + O(s_j).\end{aligned} \tag{66}$$

(c) In the random coefficient logit model with indirect utility $v_{ij} = \beta' x_{jc} + \xi_{jc} + \alpha_i \log(m - p_j) + \varepsilon_{ij}$,

$$\begin{aligned}\eta_{jj}(p) &= -\frac{\mathbb{E}[\alpha_i]}{m-p_j} p_j \;+\; \left(\frac{p_j}{m-p_j}\Big(\operatorname{Var}[\alpha_i]\big(\bar{L}(\mathbb{E}[\alpha_i]) - L_j\big) + \operatorname{err}_1^L\Big)\right) \frac{1}{1+\operatorname{err}_2^L} \;+\; \operatorname{err}_3^L \\ &= -\frac{\mathbb{E}[\alpha_i]}{m-p_j} p_j + \left(\frac{p_j}{m-p_j}\Big(\operatorname{Var}[\alpha_i]\bar{L}(\mathbb{E}[\alpha_i]) - L_j\Big)\right)(1 + O(\operatorname{Var}[\alpha_i])) \\ &\quad + O(\mathbb{E}[|\alpha_i - \mathbb{E}[\alpha_i]|^3]) + O\left(\frac{\mathbb{E}[\alpha_i s_j(\alpha_i)^2]}{\mathbb{E}[s_j(\alpha_i)]}\right).\end{aligned} \tag{67}$$

Proof. Parts (a) and (b) follow directly from the exact logit formulas and the expansion $1 - s_j = 1 + O(s_j)$ in the small share approximation.

Part (c) follows from Lemma 8 by multiplying (64) by p_j and collecting terms. The $O(\operatorname{Var}[\alpha_i])$ term reflects truncation of the Taylor expansion at second order; the $O(\mathbb{E}[|\alpha_i - \mathbb{E}[\alpha_i]|^3])$ term collects cubic (and higher) central moments of α_i; and the last term is the small-share remainder arising from $\mathbb{E}[\alpha_i s_j(\alpha_i)^2]/\mathbb{E}[s_j(\alpha_i)]$. □

Proposition 19 is the $\log(m-p_j)$ analogue of the small-share elasticity approximations in Section 3. The remainder of this appendix shows how the identification results in Section 4 translate to the $\log(m-p_j)$ specification, using these approximations.

D.2 Logit with $\alpha \log(m - p_j)$

Consider

$$v_{ijc} = \beta_0 + \xi_{jc} + \alpha \log(m - p_{jc}) + \varepsilon_{ij}.$$

Then

$$\log \frac{s_{jc}}{s_{0c}} = \beta_0 + \alpha L_{jc} + \xi_{jc},$$

so the share inversion is

$$\sigma_j^{-1}(s_c) = \log \frac{s_{jc}}{s_{0c}}.$$

The analogue of the moment condition (24) is

$$\mathbb{E}\left[J^{-1}\sum_{j=1}^{J}\begin{pmatrix}1\\ z_{jc}\end{pmatrix}\left(\log\frac{s_{jc}}{s_{0c}} - \beta_0 - \alpha L_{jc}\right)\right] = 0,$$

and the corresponding 2SLS estimator is obtained by regressing $\log(s_{jc}/s_{0c})$ on $(1, -L_{jc})$ instrumented by z_{jc}.

Define the product-specific IV estimand with respect to L_{jc}:

$$\rho_j^{\log(q/q_0),L} \equiv \frac{\mathrm{Cov}[\dot{z}_{jc}, \log(q_{jc}/q_{0c})]}{\mathrm{Cov}[\dot{z}_{jc}, L_{jc}]},$$

and weights

$$\omega_j^L \equiv \frac{\mathrm{Cov}[\dot{z}_{jc}, L_{jc}]^2}{\sum_{k=1}^{J}\mathrm{Cov}[\dot{z}_{kc}, L_{kc}]^2}, \qquad \sum_{j=1}^{J}\omega_j^L = 1.$$

Proposition 20 (Logit with $\log(m-p)$)**.** *Let $\hat{\alpha}$ be the 2SLS estimator from the above moment condition. Then*

$$\operatorname*{plim}_{C\to\infty}(-\hat{\alpha}) = \sum_{j=1}^{J}\omega_j^L\,\rho_j^{\log(q/q_0),L}.$$

Proof. Identical to Proposition 8, replacing p_{jc} by L_{jc}: the 2SLS estimand is a positively weighted average of product-level Wald estimands in L_{jc}. □

Using (65), we obtain:

Corollary 8. *Let $\hat{\eta}_{j'j'c'}$ denote the own-price elasticity of product j' in market c' from the logit model with $\log(m-p_j)$ estimated as above. Then*

$$\operatorname*{plim}_{C\to\infty}\hat{\eta}_{j'j'c'} = \left(\sum_{j=1}^{J}\omega_j^L\,\rho_j^{\log(q/q_0),L}\right)\frac{p_{j'c'}}{m-p_{j'c'}} + O(s_{j'c'}).$$

D.3 Product-specific logit with $\alpha_j \log(m-p_j)$

Now consider

$$v_{ijc} = \beta_0 + \xi_{jc} + \alpha_j \log(m - p_{jc}) + \varepsilon_{ij},$$

with common m, so

$$\log \frac{s_{jc}}{s_{0c}} = \beta_0 + \alpha_j L_{jc} + \xi_{jc}.$$

The share inversion remains $\log(s_{jc}/s_{0c})$, and the analogue of (28) is the fully interacted IV regression of $\log(s_{jc}/s_{0c})$ on product fixed effects and L_{jc} using z_{jc} as instruments.

Proposition 21 (Product-specific logit with $\log(m-p)$)**.** *Let $(\hat{\alpha}_1, \ldots, \hat{\alpha}_J)'$ be the 2SLS estimators from this interacted regression. Then*

$$\operatorname*{plim}_{C\to\infty}(-\hat{\alpha}_j) = \rho_j^{\log(q/q_0),L}.$$

Equivalently, the limit of $(-\hat{\alpha}_j)$ equals the coefficient on the product-j dummy in a regression of $\rho_j^{\log(q/q_0),L}$ onto a full set of product indicators.

Proof. Identical to Proposition 9 with p_{jc} replaced by L_{jc}. □

Using (66), we obtain:

Corollary 9. *Let $\hat{\eta}_{jjc}$ denote the own-price elasticity from the product-specific logit model with $\log(m-p_j)$. Then*

$$\operatorname*{plim}_{C\to\infty}\hat{\eta}_{jjc} = \rho_j^{\log(q/q_0),L}\,\frac{p_{jc}}{m-p_{jc}} + O(s_{jc}).$$

D.4 Random coefficient logit with $\alpha_i \log(m-p_j)$

We now adapt the random coefficient logit analysis in Section 4.5. Assume

$$\alpha_i = \mu + \lambda\nu_i, \qquad \mathbb{E}[\nu_i] = 0, \quad \mathbb{E}[\nu_i^2] = 1,$$

and

$$v_{ijc} = \beta_0 + \xi_{jc} + \alpha_i L_{jc} + \varepsilon_{ij}.$$

Let $\delta_{jc} = \beta_0 + \xi_{jc}$. Define

$$\sigma_j(\delta_c;\lambda) \equiv \int \frac{\exp\left(\delta_{jc} + (\mu+\lambda\nu_i)L_{jc}\right)}{1+\sum_{k=1}^{J}\exp\left(\delta_{kc}+(\mu+\lambda\nu_i)L_{kc}\right)}\,\mathrm{d}\Phi(\nu_i),$$

so that $s_{jc} = \sigma_j(\delta_c;\lambda)$ at the true parameters. Let $\sigma^{-1}(\cdot;\lambda)$ denote the inverse in δ.

The moment condition becomes

$$\mathbb{E}\left[J^{-1}\sum_{j=1}^{J}\begin{pmatrix}1\\ z_{jc}\end{pmatrix}\left(\sigma_j^{-1}(s_c;\lambda) - \beta_0 - \mu L_{jc}\right)\right] = 0. \tag{68}$$

Second-order expansion of the share inversion

Lemma 9. *For the model above,*

$$\frac{\partial^2\sigma_j^{-1}}{\partial\lambda^2}(s_c;0) = -L_{jc}\big(L_{jc} - 2\bar{L}_c\big),$$

where $\bar{L}_c = \sum_{k=0}^{J} s_{kc}L_{kc}$ is the share-weighted average of L_{kc}, including the outside option. Consequently,

$$\sigma_j^{-1}(s_c;\lambda) = \log\frac{s_{jc}}{s_{0c}} - \frac{1}{2}L_{jc}\big(L_{jc} - 2\bar{L}_c\big)\lambda^2 + O(\lambda^3). \tag{69}$$

Proof. Same steps as Lemma 4, replacing p_{jc} by L_{jc}. □

Substituting (69) into (68) suggests the linearized moment condition

$$\mathbb{E}\left[J^{-1}\sum_{j=1}^{J} z_{jc}\left(\log\frac{s_{jc}}{s_{0c}} - \beta_0 - \mu L_{jc} - \frac{1}{2}L_{jc}(L_{jc} - 2\bar{L}_c)\lambda^2\right)\right] = 0. \tag{70}$$

Under conditions analogous to Lemma 5, the solution to (70) approximates the solution to (68) with relative error $O(|\lambda|^3)$ along sequences with $\lambda \to 0$.

WLS characterization

Define product-level Wald estimands based on L_{jc}:

$$\rho_j^{\log(q/q_0),L} \equiv \frac{\mathbb{E}[\log(q_{jc}/q_{0c}) \mid \dot{z}_{jc} = 1] - \mathbb{E}[\log(q_{jc}/q_{0c}) \mid \dot{z}_{jc} = 0]}{\mathbb{E}[L_{jc} \mid \dot{z}_{jc} = 1] - \mathbb{E}[L_{jc} \mid \dot{z}_{jc} = 0]},$$

and

$$\tilde{L}_j \equiv \frac{\mathbb{E}[\frac{1}{2}L_{jc}(L_{jc} - 2\bar{L}_c) \mid \dot{z}_{jc} = 1] - \mathbb{E}[\frac{1}{2}L_{jc}(L_{jc} - 2\bar{L}_c) \mid \dot{z}_{jc} = 0]}{\mathbb{E}[L_{jc} \mid \dot{z}_{jc} = 1] - \mathbb{E}[L_{jc} \mid \dot{z}_{jc} = 0]}.$$

Let

$$\omega_j^L \equiv \frac{\mathrm{Cov}[\dot{z}_{jc}, L_{jc}]^2}{\sum_{k=1}^{J}\mathrm{Cov}[\dot{z}_{kc}, L_{kc}]^2}.$$

Then the 2SLS coefficients solving (70) coincide with the WLS regression coefficients (γ_0, γ_1) from

$$\rho_j^{\log(q/q_0),L} = \gamma_0 + \gamma_1\tilde{L}_j + u_j, \qquad \sum_{j=1}^{J}\omega_j^L\begin{pmatrix}1\\ \tilde{L}_j\end{pmatrix}u_j = 0. \tag{71}$$

Proposition 22 (Random coefficient logit with $\log(m-p)$: WLS characterization)**.** *Let $(\hat{\mu}, \hat{\lambda})$ solve the sample analogue of (68) with instruments z_{jc} as in Section 4.2. Under regularity conditions analogous to Lemma 5,*

$$\operatorname*{plim}_{C\to\infty}(-\hat{\mu}) = \gamma_0 + O(|\lambda|^3),$$
$$\operatorname*{plim}_{C\to\infty}\hat{\lambda}^2 = \gamma_1 + O(|\lambda|^3),$$

where (γ_0, γ_1) are defined by the weighted regression (71).

Elasticity approximation

For the random coefficient model with $\alpha_i = \mu + \lambda\nu_i$, combining Proposition 19(c) with Proposition 22 yields:

Corollary 10. *Let $\hat{\eta}_{j'j'c'}$ denote the own-price elasticity of product j' in market c' from the random coefficient logit model with $\log(m - p_j)$ estimated via (68). Then*

$$\operatorname*{plim}_{C\to\infty}\hat{\eta}_{j'j'c'} = \left[\gamma_0 + \gamma_1\left(L_{j'c'} - \bar{L}_{c'}\right)\right]\frac{p_{j'c'}}{m - p_{j'c'}} + O(|\lambda|^3) + O(s_{j'c'}).$$

Thus, in the $\log(m-p_j)$ specification, the structure of the results in Section 4 is preserved: own-price semi-elasticities are approximately affine in the transformed regressor L_{jc} (and, in the random coefficient case, in $L_{jc} - \bar{L}_c$), and the parameters of this relationship are determined by WLS regressions on product-level Wald estimands, with an additional scaling factor $p_{jc}/(m - p_{jc})$ that reflects the derivative of $\log(m - p_{jc})$ with respect to p_{jc}.

D.5 Heterogeneous income and $\log(m_i - p_j)$

In the preceding subsections, we focused on the case of common income m. In practice, income is heterogeneous: m_i varies across consumers, and its distribution F_{mc} may differ across markets c. This section shows how heterogeneous income enters the logit and random coefficient logit models when utility depends on $\log(m_i - p_{jc})$, and why the main interpretation of our results is essentially unchanged when income is heterogeneous.

Setup

Consider

$$v_{ijc} = \beta' x_{jc} + \xi_{jc} + \alpha_i \log(m_i - p_{jc}) + \varepsilon_{ijc}, \tag{72}$$

with (α_i, m_i) drawn from a joint distribution with marginal F_{mc} for m_i in market c. Let $s_j(\alpha, m)$ denote the individual choice probability for type (α, m), and

$$s_{jc} = \mathbb{E}[s_j(\alpha_i, m_i)]$$

the market share with expectation taken over (α_i, m_i).

Differentiating as usual for logit gives

$$\eta_{jjc}(p) = \frac{p_{jc}}{s_{jc}} \, \mathbb{E}\left[\frac{\partial s_j(\alpha_i, m_i)}{\partial p_{jc}}\right] = -\frac{p_{jc}}{s_{jc}} \, \mathbb{E}\left[\frac{\alpha_i}{m_i - p_{jc}} \, s_j(\alpha_i, m_i)\big(1 - s_j(\alpha_i, m_i)\big)\right]. \tag{73}$$

This nests the standard logit (i.e. $\alpha_i \equiv \alpha$), and the random coefficient logit (i.e. α_i random) with or without dependence between α_i and m_i.

D.5.1 Logit with heterogeneous income

For the logit model with a common price sensitivity $\alpha_i \equiv \alpha$, (73) simplifies to

$$\frac{\eta_{jjc}(p)}{p_{jc}} = -\alpha \, \frac{\mathbb{E}\left[\dfrac{1}{m_i - p_{jc}} \, s_j(m_i)(1 - s_j(m_i))\right]}{\mathbb{E}[s_j(m_i)]}, \tag{74}$$

where $s_j(m_i) = s_j(\alpha, m_i)$.

Under the small-share approximation, $s_j(m_i)$ is small for all m_i, so $s_j(m_i)(1 - s_j(m_i)) = s_j(m_i) + O(s_j(m_i)^2)$ and

$$\frac{\eta_{jjc}(p)}{p_{jc}} = -\alpha \, \frac{\mathbb{E}\left[\dfrac{1}{m_i - p_{jc}} \, s_j(m_i)\right]}{\mathbb{E}[s_j(m_i)]} + O(s_{jc}) = -\alpha \, \mathbb{E}\left[\frac{1}{m_i - p_{jc}} \,\middle|\, \text{choose } j\right] + O(s_{jc}). \tag{75}$$

Hence,

$$\eta_{jjc}(p) \approx -\alpha\, p_{jc}\, \mathbb{E}\left[\frac{1}{m_i - p_{jc}} \,\middle|\, \text{choose } j\right]. \tag{76}$$

Because F_{mc} is observed, the conditional expectation $\mathbb{E}[1/(m_i - p_{jc}) \mid \text{choose } j]$ is a model-implied function of prices and the known income distribution. Thus, even with heterogeneous income, the cross-product pattern of elasticities in the logit model is pinned down by a single scalar α and the observed income distribution F_{mc}. No additional arbitrary product-level flexibility is introduced.

D.5.2 Random coefficients with heterogeneous income

Next, consider the random coefficient version of (72), in which price sensitivity depends linearly on income:

$$\alpha_i = \mu + \pi m_i + \lambda \nu_i, \tag{77}$$

where ν_i has mean zero and unit variance, independent of (m_i, z_{jc}). The indirect utility is

$$v_{ijc} = \beta' x_{jc} + \xi_{jc} + \big(\mu + \pi m_i + \lambda \nu_i\big) \log(m_i - p_{jc}) + \varepsilon_{ijc}.$$

This specification parallels the demographic extension in Section 4.5.2, with $d_i = m_i$.

Case 1: $\pi = 0$ (price sensitivity independent of income). When $\pi = 0$, $\alpha_i = \mu + \lambda\nu_i$ is independent of m_i. Conditional on any fixed m, Lemma 8 applies, yielding the second-order expansion

$$\frac{\eta_{jj}(p \,|\, m)}{p_j} = -\,\mathbb{E}[\alpha_i]/(m - p_j) \;+\; \frac{\mathrm{Var}[\alpha_i]\big(\bar{L}(\mathbb{E}[\alpha_i]) - L_j\big)}{m - p_j} \;+\; \text{higher-order terms},$$

where $L_j = \log(m - p_j)$ and $\bar{L}(\alpha) = \sum_k L_k\, s_k(\alpha)$. Integrating this expression over the income distribution F_{mc} gives

$$\frac{\eta_{jjc}(p)}{p_{jc}} = -\mu_c^w \;+\; \mathcal{Q}_c^w(j) \;+\; \text{higher-order terms}, \tag{78}$$

where

$$\mu_c^w \equiv \int \frac{\mathbb{E}[\alpha_i]}{m - p_{jc}}\, \omega_j(m)\, \mathrm{d}F_{mc}(m),$$

$$\mathcal{Q}_c^w(j) \equiv \int \frac{\mathrm{Var}[\alpha_i]\big(\bar{L}(\mathbb{E}[\alpha_i]) - L_j(m)\big)}{m - p_{jc}}\, \omega_j(m)\, \mathrm{d}F_{mc}(m),$$

and $\omega_j(m) \propto s_j(m; \mathbb{E}[\alpha_i])$ denotes the share weight for income level m. That is, μ_c^w is the share-weighted average over incomes of the baseline term $\mathbb{E}[\alpha_i]/(m - p_{jc})$ from Lemma 8, and $\mathcal{Q}_c^w(j)$ similarly averages the variance correction term. Both quantities depend only on (μ, λ), prices, and the observed income distribution F_{mc}. Hence, heterogeneous income simply rescales the constant-income approximation by taking share-weighted expectations over the income distribution—it does not introduce any additional functional flexibility in the mapping from prices to elasticities.

Case 2: $\pi \neq 0$ (price sensitivity depends on income). To isolate the role of systematic income dependence, we set $\lambda = 0$ in this case; allowing $\lambda \neq 0$ would add the same quadratic λ^2-type terms as in Proposition 22 without changing the basic identification logic. Hence,

$$\alpha_i = \mu + \pi m_i.$$

Individual utility is then

$$v_{ijc} = \beta' x_{jc} + \xi_{jc} + (\mu + \pi m_i)\log(m_i - p_{jc}) + \varepsilon_{ijc}.$$

When $\pi \neq 0$, the distribution of α_i varies with income m_i, and the mean utility $\sigma_j^{-1}(s_c; \mu, \pi, F_{mc})$ depends on both the parameters (μ, π) and the observed income distribution F_{mc}. A first-order expansion in π around $\pi = 0$ yields

$$\sigma_j^{-1}(s_c; \mu, \pi, F_{mc}) = \log \frac{s_{jc}}{s_{0c}} - \mu\, T_{jc} - \pi\, \kappa_{jc} + O(\pi^2), \tag{79}$$

where T_{jc} and κ_{jc} are known functions of prices and F_{mc}, constructed respectively from conditional averages of $\log(m_i - p_{jc})$ and $m_i \log(m_i - p_{jc})$ under the baseline model with $\pi = 0$.

Intuitively, the first term T_{jc} captures the average response of shares to changes in the common price sensitivity μ, while the second term κ_{jc} captures how differences in income m_i re-weight that response when α_i increases linearly with income.

Product-level WLS characterization. Define the product-level IV estimands and weights:

$$\rho_j^{\log(q/q_0)} \equiv \frac{\mathrm{Cov}[\dot{z}_{jc}, \log(q_{jc}/q_{0c})]}{\mathrm{Cov}[\dot{z}_{jc}, p_{jc}]}, \qquad \omega_j \equiv \frac{\mathrm{Cov}[\dot{z}_{jc}, p_{jc}]^2}{\sum_k \mathrm{Cov}[\dot{z}_{kc}, p_{kc}]^2}.$$

Let $\tilde{\kappa}_j$ denote the ω_j-weighted transform of κ_{jc}, defined in analogy with $\tilde{d}_j$ in Proposition 11:

$$\tilde{\kappa}_j \equiv \rho_j^{p\kappa} = \frac{\mathrm{Cov}[\dot{z}_{jc}, \kappa_{jc}]}{\mathrm{Cov}[\dot{z}_{jc}, p_{jc}]}.$$

Proposition 23. *Let $(\hat{\mu}, \hat{\pi})$ be the estimators solving the sample analogue of the moment conditions associated with (79). Then, for sufficiently small π and under the same regularity conditions as Proposition 11, there exist coefficients (γ_0, γ_1) such that*

$$\rho_j^{\log(q/q_0)} = \gamma_0 + \gamma_1 \tilde{\kappa}_j + u_j, \qquad \sum_{j=1}^{J} \omega_j \begin{pmatrix} 1 \\ \tilde{\kappa}_j \end{pmatrix} u_j = 0, \tag{80}$$

$$\operatorname*{plim}_{C\to\infty}(-\hat{\mu}) = \gamma_0 + O(\pi^2),$$
$$\operatorname*{plim}_{C\to\infty} \hat{\pi} = \gamma_1 + O(\pi^2).$$

Equation (80) mirrors the structure of Propositions 10 and 11: (μ, π) are identified by a weighted least-squares regression of product-level IV estimands $\rho_j^{\log(q/q_0)}$ on the known income-based transform $\tilde{\kappa}_j$, which is fully determined by the observed income distribution F_{mc} and prices $\{p_{jc}\}_j$.

Implications. Under the small-share and small-π approximations, the own-price elasticity satisfies

$$\eta_{jjc}(p) \approx -\mu\, p_{jc}\, \mathbb{E}\left[\frac{1}{m_i - p_{jc}} \,\middle|\, \text{choose } j\right] \;+\; \pi\, \Psi_{jc} \;+\; O(\pi^2) + O(s_{jc}),$$

where Ψ_{jc} is a known linear functional of $\mathbb{E}[m_i \log(m_i - p_{jc}) \mid \text{choose } j]$. Thus, across products, $\eta_{jjc}(p)$ is approximately an affine function of a small number of price–income transforms derived from F_{mc}. Even when price sensitivity depends on income, the pattern of elasticities cannot be chosen freely: it is pinned down by (μ, π) and the observed income distribution.

E Nested logit model

Suppose the J products are partitioned into G groups, and let $g(j)$ denote the group to which product j belongs. The nested logit model is a random utility model with indirect utilities

$$v_{ij} = \beta' x_j + \alpha(m_i - p_j) + \sum_{g=1}^{G} \nu_{ig} \mathbb{1}[g(j) = g] + \lambda \varepsilon_{ij}$$

where the ν_{ig} are independent of each other and of ε_{ij} and follow the $C(\lambda)$ distribution (Cardell, 1997). Dividing by λ and applying Proposition 5, we obtain the own-price elasticity approximation

$$\eta_{jjc}(p_c) \approx -\frac{\alpha}{\lambda} p_{jc}. \tag{81}$$

Own-price elasticities are linear in price levels, up to approximation error. Applying Proposition 7, we obtain the cross-price elasticity approximation

$$\eta_{kjc}(p_c) \approx \frac{\alpha}{\lambda} p_{jc} s_{jc} \left[1 + \sum_{g=1}^{G} \frac{\mathrm{Var}[\nu_{ig}]}{\lambda^2} (\mathbb{1}[g(j) = g] - s_{gc})(\mathbb{1}[g(k) = g] - s_{gc})\right]$$

where $s_{gc} \equiv \sum_{j \in g} s_{jc}$. Compared to the standard logit model, cross-price elasticities will be larger between products in the same nest; the difference in cross-price elasticities of the standard and nested logit models will be greater the smaller is λ. Fixing $g(j)$ and $g(k)$, the cross-price elasticity will be linearly increasing in the price of j and in the share of j.