Price Elasticities in Logit Demand Models*

Ariel Goldszmidt† Lancelot Henry de Frahan‡ John A. List§

Magne Mogstad¶ Ian Muir‖ Alexander Torgovitsky**

July 21, 2026

Abstract

We study the own-price elasticities produced by mixed (random coefficient) logit demand models. A well-known criticism of the standard (constant coefficient) logit model is that its own-price elasticities are stylized: when inside good shares are small, they are close to linear functions of own price alone. We show that mixed logit models inherit a version of this property—their own-price elasticities are close to *quadratic* functions of own price alone—unless they allow for correlation between the random coefficients on price and non-price characteristics. We survey the empirical literature and document that explicitly estimating such a correlation is rare in practice. Revisiting four applications, two of which are widely used for testing demand methods, we find that all four produce own-price elasticities that look quadratic in own price. We apply our results to study toothpaste demand using a large-scale natural field experiment. We find that it is hard to reconcile linear instrumental variable estimates, which we interpret through the lens of a nonparametric demand model, with comparable estimates from mixed logit models.

*We are grateful to Linqi Zhang for sharing the list of papers used in her survey on empirical applications of discrete choice demand models. Matt Backus, Kevin Chen, JP Dubé, Liran Einav, Ben Handel, Phil Haile, Sukjin Han, Alessandro Iaria, Julien Monardo, Olivia Natan, Áureo de Paula, and Jesse Shapiro provided helpful comments. Sheyan Lalmohammed provided outstanding research assistance.

†Kenneth C. Griffin Department of Economics, University of Chicago.

‡Kenneth C. Griffin Department of Economics, University of Chicago.

§Kenneth C. Griffin Department of Economics, University of Chicago; NBER.

¶Kenneth C. Griffin Department of Economics, University of Chicago; Statistics Norway; NBER.

‖Walmart.

**Kenneth C. Griffin Department of Economics, University of Chicago.

1 Introduction

Discrete choice demand models are central to empirical work in industrial organization and related fields. While some recent work has explored the possibility of using flexible nonparametric models (for example, Berry and Haile, 2014; Compiani, 2022; Tebaldi et al., 2023), logit-based models remain by far the most widely-used due to their ability to handle large numbers of products. Since the pioneering work of Berry (1994) and Berry et al. (1995, "BLP"), it has become common to use mixed (random coefficient) logit models to circumvent the well-known negative implications of the standard (constant coefficient) logit model for cross-price elasticities.

In this paper, we focus on the *own*-price elasticities produced by mixed logit demand models. Standard logit models are known to produce stylized own-price elasticities that are well-approximated by linear functions of own price when inside good market shares are small (Gandhi and Nevo, 2021, Section 3.2.1). We provide theoretical and empirical results which show that commonly-used specifications of the mixed logit lead to own-price elasticities that are *quadratic* functions of own price when type-specific choice probabilities for inside goods are small. Escaping this property requires incorporating correlation between the random coefficients on price and non-price characteristics, something which we document is rarely explicitly done in practice. The implication is that while common specifications of mixed logit models may help create more flexible cross-price elasticities, they still lead to own-price elasticities that are nearly as stylized as in the standard logit model.

The reason that the standard logit model has stylized own-price elasticities is that all product characteristics approximately cancel out of the semi-elasticity for goods with small market shares. Our first result shows that a similar phenomenon continues to happen in the mixed logit model for the characteristics with constant coefficients: semi-elasticities in the mixed logit model only depend on the characteristics with random coefficients if type-specific choice probabilities are small. This result implies that a researcher who uses a specification with only a random coefficient on price has—the same as in a standard logit—used a model whose own-price elasticities are well-approximated by a function that only depends on own price.

Specifications with random coefficients on both price and non-price characteristics are more commonly-used, but often with the assumption that these random coefficients are independent. Our second result shows that the structure of the logit model implies that this independence assumption produces an approximate factorization that again leads own-price elasticities to be well-approximated by functions of own prices alone. Correlated random coefficients or large variances of non-price random coefficients are needed to overturn this implication. While estimating variance matrices that explicitly allow for correlations is often statistically and computationally difficult, incorporating demographics can also induce correlations, potentially at lower implementation cost. Our second result implies that incorporating demographics can have the unexpected benefit of increasing the flexibility of own-price elasticities.

Our third result shows that there exists a quadratic approximation of own-price elasticities that fits the actual elasticities remarkably well. This quadratic approximation works well with both independent and correlated random coefficients. In the independent case, it implies that own-price elasticities are approximately quadratic functions of own prices alone.

The three results are linked together by showing that the mixed logit model is an exponentially-tilted adjustment to the standard logit model. This representation implies that if type-specific choice probabilities are small, then the semi-elasticities from the mixed logit are approximately the average coefficient on price under this exponentially-tilted distribution. The distribution approximately factors when price and non-price random coefficients are independent, limiting the role of the non-price characteristics. The distribution is also very smooth due to properties of exponential family distributions, making it well-approximated by a quadratic function.

Throughout our analysis, we illustrate our theoretical results with extensive empirical evidence. This evidence is based on applications using US automobile data (Berry et al., 1995), the synthetic cereal data used by Nevo (2000), modeled after Nevo (2001), which was packaged by Conlon and Gortmaker (2020), an application to the demand for automobiles in China by Li (2018), and a simulation calibrated to the French automobile data used by D'Haultfœuille et al. (2019). All four applications strongly support our theoretical results. We use modifications of the synthetic cereal data to explore settings in which the approximation suggested by our theory performs well, as well as ones in which it starts to break down.

Taken together, our theoretical and empirical results show that own-price elasticities in commonly-used mixed logit specifications tend to be stylized, typically being well-represented as quadratic functions of own prices. This has important implications for using own-price elasticities in downstream analyses, such as for estimating marginal costs, markups, pass-through, and diversion ratios. Stylized own-price elasticities make these downstream quantities stylized as well. For example, our results imply that commonly-used mixed logit models tend to produce markups that follow a U-shape, so that markups are high for both low-priced and high-priced products.

Cross-price elasticities produced by the mixed logit model behave differently. We extend our analysis to show that cross-price elasticities share a similar structure to own-price elasticities, but differ in a way that introduces a first-order role for both price and non-price characteristics of the two products under comparison, even with independent random coefficients. While this supports the usual intuition, our findings overall suggest that the cross-price flexibility of the mixed logit model may not be the first-order issue when considering the many types of counterfactuals that also depend on own-price effects.

We apply our results to study toothpaste demand using data from a natural field experiment that resulted when a large national US retailer randomly assigned prices in over four thousand stores. This setting is particularly well-suited for our purposes because its large size enables flexible linear instrumental variable (IV) estimation with high statistical precision using credible, randomly-assigned prices as instruments. This power also enables estimation of mixed logit models with correlated random coefficients. We assess whether mixed logit specifications can reproduce the weighted averages of arc semi-elasticities that are estimated by linear IV methods under weak nonparametric assumptions. We find that mixed logit models in which the quadratic own-price elasticity relationship is strong cannot reproduce these objects. More surprisingly, we find that this remains true even in more flexible mixed logit specifications with correlated random coefficients. This finding suggests that even when using correlated random coefficients, the mixed logit model is unable to capture the substantial heterogeneity across products with similar prices.

The organization of this paper is as follows. In Section 2, we introduce notation and definitions for the class of logit models we study. We also report the results of our survey of the specifications used in the empirical literature and document claims that mixed logit models are believed to solve the restrictive properties of standard logit models. Section 3 contains our main theoretical and empirical results about own-price elasticities. Section 4 contains the extension to cross-price elasticities. Section 5 contains our application to toothpaste demand using data from a large national US retailer. Section 6 concludes with implications of our results for empirical researchers, along with recommendations for practice. Proofs are in Appendix SA.2.

2 Logit Demand Models

This section contains a definition of the mixed logit model and a discussion of how it is applied in practice and described in theory. In Section 2.1, we lay out the notation we use for the mixed logit model. In Section 2.2, we report the results of a survey of the specifications of logit models used in empirical work. In Section 2.3, we discuss some examples of how the benefits of the mixed logit model are often described.

2.1 Model

Berry (1994) studied a family of discrete choice demand models with indirect utility

$$\alpha_i p_{jt} + \beta_i' r_{jt} + \gamma' c_{jt} + \xi_{jt} + \varepsilon_{ijt}, \tag{1}$$

where i indexes an individual in market $t = 1, \ldots, T$ and $j = 0, 1, \ldots, J$ indexes a product. The researcher observes the price p_{jt} and non-price characteristics (r_{jt}, c_{jt}) associated with product j in market t. They do not observe ξ_{jt}, which collects unobserved characteristics of the product and market, including market-level demand shocks, nor do they observe the idiosyncratic preference shock ε_{ijt}. Option $j = 0$ is an outside option that is normalized to have indirect utility ε_{i0t} for all i and t.

The observable characteristics of product j are (p_{jt}, r_{jt}, c_{jt}), which we have broken up into three components: (i) price p_{jt}, which is the characteristic for which we want to consider elasticities; (ii) characteristics r_{jt} that have random coefficients β_i which can vary with i; and (iii) characteristics c_{jt} that have non-random (constant) coefficients γ. The distinction between r_{jt} and c_{jt} will be important ahead. None of α_i, β_i, or γ are observed by the researcher.

The family of logit models that we consider comes from combining (1) with the assumption that ε_{ijt} follows a type I extreme value distribution, independently across individuals, products, and all variables in the model. This implies the type-specific choice probabilities

$$\pi_{jt}(\alpha_i, \beta_i) \equiv \frac{\exp\left(\alpha_i p_{jt} + \beta_i' r_{jt} + \gamma' c_{jt} + \xi_{jt}\right)}{\sum_{k=0}^{J} \exp\left(\alpha_i p_{kt} + \beta_i' r_{kt} + \gamma' c_{kt} + \xi_{kt}\right)}. \tag{2}$$

Integrating the type-specific choice probabilities over the distribution of preference parameters (types) creates market-level shares

$$s_{jt} \equiv \int \pi_{jt}(\alpha, \beta)\, dF_t(\alpha, \beta) \equiv \mathbb{E}[\pi_{jt}(\alpha, \beta)], \tag{3}$$

where F_t is the distribution of (α, β) in market t and we use $\mathbb{E}$ to represent expectation over (α, β) while suppressing the superfluous i subscripts. We focus on market-level demand applications in which the researcher observes s_{jt}, but not the choices or characteristics of individual consumers.

Constructing the market-level shares (3) requires taking a stance on the distribution F_t of the preference parameters (α, β). The *standard logit* model assumes that α and β are both constant (deterministic) parameters within a market, in which case there is no distinction between characteristics r_{jt} and c_{jt}. In this case the market-level shares (3) are the same as the type-specific choice probabilities (2). The *mixed logit* or *random coefficient logit* model allows α and β to be random variables, usually with the assumption that F_t belongs to a parametric class of distributions, such as multivariate normal. This model was proposed by Berry (1994), but is often associated with Berry et al. (1995, "BLP"), who implemented it while accounting for the ξ_{jt} unobservables.

Our analysis of the flexibility of own-price elasticities is centered on within market comparisons across products. Whether or how F_t varies across t is not relevant for this analysis. However, demographic interactions between product characteristics are a common source of cross t variation, and how these interactions are implemented does have important implications for the within-market structure of F_t. We return to this point in Section 3.4.

2.2 A survey of empirical work

Table 1 reports the results of a survey on the use of logit models of demand. The survey covers a collection of 29 papers assembled by Zhang (2026), which were published between 2018 and 2026 in *Econometrica*, *Journal of Political Economy*, *American Economic Review*, *The Review of Economic Studies*, or *The Quarterly Journal of Economics*. We drop four papers: three that do not include price as a product characteristic and one that uses a non-standard two-stage model. The remaining 25 papers are classified according to whether and how they use random coefficients for their main specification and at least once in any specification in the main text or appendices. Appendix SA.1 contains more details on the sample and classification rules.

The first row of Table 1 shows that logit (or nested logit) models without random coefficients are used for the main results in a minority of papers, although they are still frequently used in comparison or supplemental specifications. Random coefficients are used for the main results in 80% of papers. Over 70% of these use a random coefficient on price. We found four papers whose *only* random coefficient is on price, a specification which we show ahead is particularly stylized.

Since Berry et al. (1999), a common way to include income in a specification is to make the price coefficient $\alpha_i = \bar{\alpha}/\textsc{inc}_i$, where $\textsc{inc}_i$ is drawn from a market-level distribution of income.[1] For the purposes of our analysis, this is like having a random coefficient, just that the shape of the distribution is set to external data up to the estimated constant coefficient parameter $\bar{\alpha}$. Seven of the papers in our survey use a similar specification to interact the same demographics (or joint draws of different

[1] This specification is intended to approximate one that replaces price with $\log(\textsc{inc}_i - p_j)$ as in Berry et al. (1995). Conlon and Gortmaker (2020, footnote 98) discuss how the log specification can lead to "a host of problems."

Table 1: Survey of mixed logit specifications used in empirical work

Random coefficients	Any use #	Any use %	Main results #	Main results %	Products per market
None	13	52.0	5	20.0	2.2M^{t}, 8.6^{a}, 6^{m}, 2^{m}, n.r.
Not on price	3	12.0	2	8.0	$\sim 400^{\text{a}}$, 85^{m}
Price only	5	20.0	4	16.0	571^{m}, 10^{m}, $\sim 4^{\text{a}}$, 2^{m}
Price and independent	7	28.0	6	24.0	$\sim 400^{\text{a}}$, 94^{m}, 73^{m}, 64^{m}, 37^{t}, 29^{a}
Correlated only through demographics	7	28.0	7	28.0	690^{t}, 314^{m}, 312^{t}, 202^{m}, 39^{m}, 22^{m}, 5.2^{a}
Correlated through unobservables	1	4.0	1	4.0	9^{m}
Total	—	—	25	100.0	

Notes: See Appendix SA.1 for details. The survey starts with the same set of papers used by Zhang (2026). We are grateful to Linqi Zhang for sharing these papers with us. The superscripts in number of products indicate how we inferred them. If available, we report m as the per-market maximum number of products. If not available, we report a for the per-market mean number of products, t for the total number of products across markets. In one case we found no usable figure, recorded as n.r.

demographics) with both the price and non-price random coefficients. This induces a within-market correlation between the price and non-price random coefficients. We label these specifications “correlated only through demographics” because they do not contain a free parameter whose role is to produce correlation; the correlation is a byproduct of the estimated demographic loadings. As we show in Section 3.4, this correlation has unexpected benefits in making own-price elasticities more flexible.[2] The label “correlated through unobservables” means a freely estimated off-diagonal element in the covariance matrix of the random coefficients. This is much rarer, perhaps because it is notoriously difficult to estimate. The only such paper in our survey is Crawford et al. (2019), which had two characteristics (price and quality) with a correlated bivariate normal specification for their random coefficients.

The own-price elasticity approximations we derive in Section 3 rely on type-specific choice probabilities for the inside goods being small in a stochastic sense. The last column of Table 1 shows the number of products for each paper, grouped by the specification used for their main results. While the number of products is not always possible to infer from the paper and can vary across markets, the general pattern is clear: most applications use more than twenty products, and many use more than that. With this many products, inside good shares and type-specific choice probabilities tend to be small, so that these approximations tend to be accurate. Indeed, it is exactly this ability to scale in the number of products that makes logit-based models so attractive for discrete choice demand estimation (Hausman and Wise, 1978; Train, 2009).

[2] Of the seven papers, only Bourreau et al. (2021) discusses the correlation induced by demographics.

2.3 Discussions of the mixed logit model

Discussions of mixed logit models in textbooks and handbook chapters often describe them as solving the stylized properties of standard logit models. While our results show that this is a reasonable statement for cross-price elasticities, they show that it requires several caveats for own-price elasticities. The discussions we have seen are, at best, silent about the distinction.

For example, in describing the stylized own-price elasticities of the standard logit, Gandhi and Nevo (2021, pp. 78–79) write

> *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.*

They go on to explain that for the mixed logit model

> *Therefore, own price elasticities are not driven solely by functional form, but by the heterogeneity in the price sensitivity across consumers who purchase the various products.*

Our results ahead show that in the most common mixed logit specifications, the own-price elasticities are approximately quadratic in own price; the shape is largely driven by functional form.

Textbooks often contain less nuanced discussions that focus on cross-price elasticities to the exclusion of own-price elasticities. See, for example, Train (2009, Section 6.4) or Hortaçsu and Joo (2023, Section 3.2.3.2). This may be a result of the emphasis placed on cross-prices in the influential papers by Berry (1994) and Berry et al. (1995). Whatever the cause, the effect has been a lack of attention to the properties of own-price elasticities in the mixed logit model. The closest scrutiny we have found in print is from Björnerstedt and Verboven (2016, pg. 138):

> *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.*

The results we present in Section 3 provide justification of their suspicion.

3 Own-Price Elasticities

The own-price elasticity for good j in market t can be computed from (2)–(3) to be

$$\epsilon^{\star}_{jt}(\boldsymbol{p}_t, \boldsymbol{r}_t, \boldsymbol{c}_t, \boldsymbol{\xi}_t) \equiv \frac{\partial s_{jt}}{\partial p_{jt}} \frac{p_{jt}}{s_{jt}} = \frac{\mathbb{E}[\alpha p_{jt}\,(1-\pi_{jt}(\alpha,\beta))\,\pi_{jt}(\alpha,\beta)]}{\mathbb{E}[\pi_{jt}(\alpha,\beta)]} \equiv p_{jt}\eta^{\star}_{jt}(\boldsymbol{p}_t, \boldsymbol{r}_t, \boldsymbol{c}_t, \boldsymbol{\xi}_t), \tag{4}$$

where $\boldsymbol{p}_t, \boldsymbol{r}_t, \boldsymbol{c}_t$, and $\boldsymbol{\xi}_t$ are vectors that collect the characteristics of all products in the market—which enter through the choice probabilities $\pi_{jt}(\alpha,\beta)$—and $\eta^{\star}_{jt}$ is the corresponding own-price semi-elasticity function. In this section, we review the properties of (4) for the standard logit model before deriving new approximations to (4) for the mixed logit model. The results that we derive concern elasticities within any given market t, rather than across markets, so we suppress the t subscript for brevity. Our derivations focus on the semi-elasticity $\eta^{\star}_j$. Implications for the elasticities $\epsilon^{\star}_j$ follow directly from multiplying by price.[3]

3.1 Standard logit

When α and β are non-random $\pi_j(\alpha,\beta) = s_j$ is also non-random and (4) reduces to

$$\epsilon^{\star}_j(\boldsymbol{p}, \boldsymbol{r}, \boldsymbol{c}, \boldsymbol{\xi}) \equiv p_j\eta^{\star}_j(\boldsymbol{p}, \boldsymbol{r}, \boldsymbol{c}, \boldsymbol{\xi}) = p_j\,(\alpha - \alpha s_j) \approx p_j\alpha \qquad \text{if } s_j \approx 0. \tag{5}$$

Gandhi and Nevo (2021, pp. 88–90) highlight the approximation in (5), which holds when the share of good j is small, so that $s_j \approx 0$. They argue that (5) reveals an unattractive property of the logit model. Their argument has two components. First, if α has the expected negative sign, then (5) implies that higher-priced products are more elastic, an implication which may or may not be true depending on the context. Second, (5) is only a function of the price of good j. It does not depend on other characteristics of good j nor on the characteristics of other goods k in the market.

The second property is particularly unattractive for demand analysis with differentiated goods. Suppose that the market for breakfast cereal includes three products: Post Raisin Bran, Kellogg's Raisin Bran, and Lucky Charms, all with small market shares so that (5) is an accurate approximation. Suppose that all three products sell for the same price. Then (5) implies that Post Raisin Bran has the same own-price elasticity as Lucky Charms, even though it has a close competitor in Kellogg's Raisin Bran with other similar observed non-price characteristics.

[3] The following results can also be applied to elasticities with respect to characteristics other than price. We focus on price for concreteness and because price elasticities play a special role in the downstream purposes for which discrete choice demand models are often used (see Section 3.6).

3.2 The anatomy of mixed logit elasticities

It is more difficult to interpret (4) in a mixed logit model because α and β are random. However, the semi-elasticity in (4) has a structure that is quite similar to (5):

$$\eta_j^\star(\boldsymbol{p}, \boldsymbol{r}, \boldsymbol{c}, \boldsymbol{\xi}) = \mathbb{E}\left[\alpha \underbrace{\frac{\pi_j(\alpha, \beta)}{\mathbb{E}[\pi_j(\alpha, \beta)]}}_{\text{weight}}\right] - \mathbb{E}\left[\alpha \pi_j(\alpha, \beta) \underbrace{\frac{\pi_j(\alpha, \beta)}{\mathbb{E}[\pi_j(\alpha, \beta)]}}_{\text{weight}}\right]. \tag{6}$$

The second term of (6) will tend to be small when the type-specific choice probability of good j is small in a stochastic sense. This implies that own-price elasticities in the mixed logit will be driven by the first term of (6). This first term is the analogue of the constant α in (5), but now it is random and weighted by the type-specific choice probability of good j.

It turns out that this weighting only depends on the characteristics of good j that have random coefficients. The next proposition formalizes these observations.

Proposition 1. Own-price semi-elasticities in the mixed logit model satisfy

$$\eta_j^\star(\boldsymbol{p}, \boldsymbol{r}, \boldsymbol{c}, \boldsymbol{\xi}) = \eta(p_j, r_j) + \iota_j^{\text{SCP}} \qquad \text{where} \qquad \iota_j^{\text{SCP}} \equiv -\,\mathbb{E}\left[\alpha \frac{\pi_j(\alpha, \beta)^2}{\mathbb{E}[\pi_j(\alpha, \beta)]}\right],$$

$\eta(p_j, r_j)$ is a function that only depends on (p_j, r_j) for any given market, and ι_j^{SCP} is a term that is small when the choice probabilities $\pi_j(\alpha, \beta)$ tend to be small:

$$|\iota_j^{\text{SCP}}| \leq \mathbb{E}[\alpha^2]^{1/2} \left(\frac{\mathbb{E}[\pi_j(\alpha, \beta)^4]}{\mathbb{E}[\pi_j(\alpha, \beta)]^2}\right)^{1/2}.$$

Proposition 1 shows that if type-specific choice probabilities for good j are small in a stochastic sense, then the own-price semi-elasticity of good j is primarily determined by the function $\eta(p_j, r_j)$. The quantity ι_j^{SCP}—the acronym stands for "small choice probabilities"—generalizes the small share remainder αs_j from the standard logit approximation (5). It is small when the distribution of the type-specific choice probabilities $\pi_j(\alpha, \beta)$ is concentrated towards zero. In the following, we refer to the approximations that ignore ι_j^{SCP} as *small-π approximations.*

For a small-π approximation to work well, the market share of good j needs to be small, as in a standard logit. However, in a mixed logit model this is not sufficient, because it is possible for the market share $s_j \equiv \mathbb{E}[\pi_j(\alpha, \beta)]$ of good j to be small even while $\mathbb{E}[\alpha \pi_j(\alpha, \beta)^2]$ is large. This would happen if there is a large enough mass of certain types (realizations of α, β) who are strongly attracted to good j. In this case,

ι_j^{SCP} could be large for a good with small market share.[4] We don't find any empirical evidence of this occurring in the applications we consider ahead, however. In these applications, the small-π approximation tends to work well when market shares s_j are small.

The function η is a weighted average of α that has a known form discussed more ahead and is straightforward to compute from parameters and data. It depends on the market; if we were including t subscripts explicitly it would be η_t. However, within a market, it depends solely on the characteristics of product j that have random coefficients: η is not a function of the characteristics of other products, nor is it a function of the constant-coefficient characteristics c_j and ξ_j for product j. If two products j and k with small values of ι_j^{SCP} and ι_k^{SCP} have $p_j = p_k$ and $r_j = r_k$, then their own-price elasticities will be similar.[5]

We quantify the small-π approximation in Proposition 1 through the relative error:

$$\text{ERR}_j(\eta) \equiv \frac{|\eta(p_j, r_j) - \eta_j^\star(\boldsymbol{p}, \boldsymbol{r}, \boldsymbol{c}, \boldsymbol{\xi})|}{|\eta_j^\star(\boldsymbol{p}, \boldsymbol{r}, \boldsymbol{c}, \boldsymbol{\xi})|} \times 100\%. \tag{7}$$

Note that the relative error is the same for both the semi-elasticity and implied elasticity, because it remains the same after multiplying by p_j. Table 2 reports the distribution of the relative error pooled across markets and products for several mixed logit specifications in four datasets: the classic US automobile data from Berry et al. (1995), Chinese automobile data from Li (2018), simulated French automobile data from D'Haultfœuille et al. (2019), and the widely-used synthetic cereal data (Nevo, 2000; Conlon and Gortmaker, 2020), which is modeled after the application in Nevo (2001). Appendix SA.3 contains details on the data and specifications for each application. The subrows for synthetic cereal are adjustments to the baseline mixed logit estimates from Conlon and Gortmaker (2020), which we use to illustrate features of the theory developed ahead.

The small-π approximation in Proposition 1 is reflected in the columns headed $\text{ERR}(\eta)$, which summarize the percentage relative error from ignoring the small-π term ι_j^{SCP}. The 90th percentile is just 2.9% in Berry et al. (1995) data and no more than 15% across all other applications, while the median is just 0.7% in Berry et al. (1995) and similarly small in the baseline specifications for most other applications. This

[4] We thank Phil Haile for a useful discussion on this point.

[5] Researchers sometimes place random coefficients on characteristics that are the same for all inside products but are zero for the outside product (for example, Berry et al., 1995; Li, 2018). In our notation, these characteristics would be part of r_j, but would be the same for all inside products, so do not contribute to variation in own-price elasticities in the small-π approximation.

Table 2: Relative error of semi-elasticity approximations

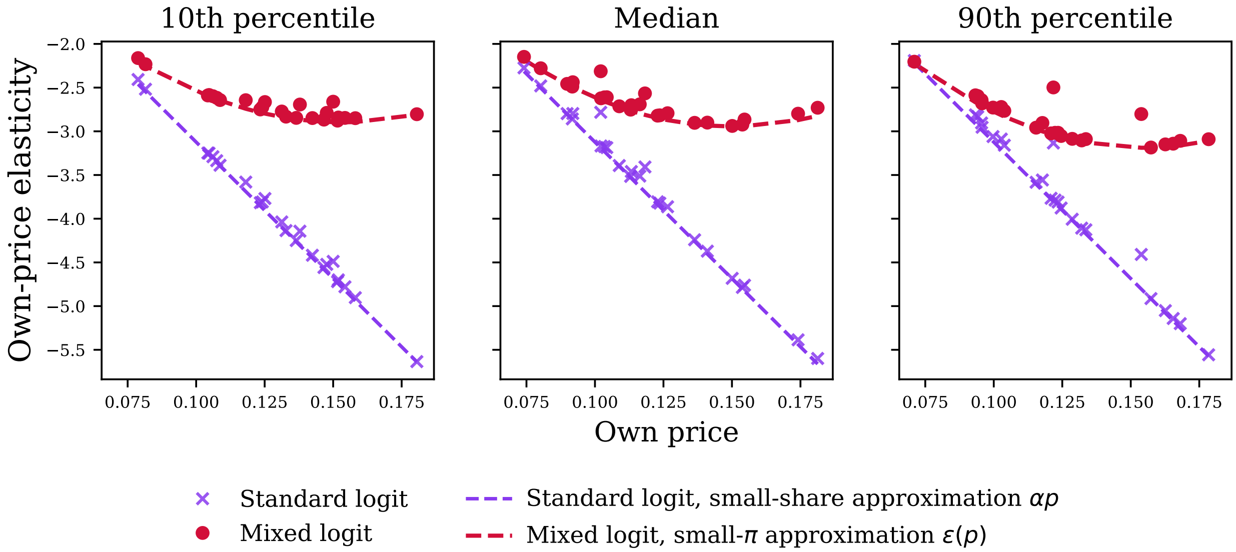
	ERR(η)			ERR($\bar{\eta}$)			ERR(η_{Q})			ERR($\bar{\eta}_{\mathrm{Q}}$)		
Percentile:	10	50	90	10	50	90	10	50	90	10	50	90
US autos	0.1	0.7	2.9	0.3	2.1	11.6	0.3	2.7	16.5	1.9	7.7	24.9
China autos	0.0	0.0	0.3	0.0	0.1	0.7	0.0	0.3	1.6	0.0	0.2	1.5
France autos	0.1	0.7	2.3	0.1	0.7	2.3	0.1	0.7	2.3	0.1	0.7	2.3
Synthetic cereal	0.4	2.2	11.3	1.6	6.4	14.7	0.7	3.0	11.6	1.4	6.7	15.2
Price only	0.3	1.3	5.7	0.3	1.3	5.7	0.2	1.2	5.5	0.2	1.2	5.5
Independent	0.5	2.6	11.4	1.1	4.9	11.5	0.2	1.9	11.7	1.1	5.1	11.9
Independent, lognormal	0.5	2.8	11.8	0.5	3.1	10.5	0.5	2.9	12.3	0.6	3.7	11.5
Independent, $\mathbb{V}[\alpha] \times 4$	0.4	2.6	19.4	3.4	14.3	61.9	0.3	2.1	21.5	3.7	14.4	53.6
Independent, $\mathbb{V}[\alpha] \times 4$*, lognormal*	0.5	2.9	12.7	0.7	4.4	12.1	0.5	3.0	12.9	0.7	5.2	13.4
Independent, $\mathbb{V}[\alpha] \div 4$	0.4	2.5	11.2	0.5	2.4	10.2	0.2	2.3	11.1	0.5	2.6	10.4
Independent, $\mathbb{V}[\beta] \times 4$	0.7	4.5	23.1	1.6	8.0	24.0	0.3	2.8	23.6	1.6	7.4	22.7
Independent, $\mathbb{V}[\beta] \div 4$	0.3	1.7	7.0	0.6	2.4	6.4	0.2	1.5	7.1	0.5	2.5	6.8
Price and mushy, correlation .2	0.3	1.6	7.3	0.2	1.0	6.6	0.3	1.5	6.9	0.2	1.6	7.5
Price and mushy, correlation .8	0.3	1.5	6.9	4.7	9.8	22.0	0.2	1.3	6.5	3.8	6.7	23.6
Independent, demographics	0.5	2.7	11.2	1.9	10.9	29.1	0.5	4.2	15.0	1.8	10.1	27.4

Notes: The relative error is defined in (7). See Appendix SA.3 for details on each dataset and the specification of the mixed logit. Subrows for synthetic cereal show modifications to the baseline specification discussed in the main text. The percentiles are pooled across elasticities for all products and markets. Table SA.4 shows market-level percentiles for average within-market relative error. France autos and synthetic cereal "price only" only have random coefficients on price, so the second column is mechanically equal to the first and the fourth column is mechanically equal to the third. The first and third columns of France autos are only equal up to the precision shown in the table.

suggests that all of these applications are comfortably in the small-π regime described by Proposition 1. Table SA.4 takes the pooled statistics in Table 2 and first averages the relative error across goods within a market. The resulting figures show the same pattern, suggesting that when the small-π approximation breaks down, it does so only for certain goods within a market, rather than for the entire market.

One immediate implication of Proposition 1 is that if price is the only characteristic with a random coefficient, then the own-price elasticities produced by the mixed logit model will depend only on p_j, the same as in the standard logit model. An example of a paper in our survey that uses this type of specification is D'Haultfœuille et al. (2019), who study the French automobile market. The third row of Table 2 is computed from their simulated data generating process, which they calibrated to their actual data. It shows that the 90th percentile of error in their setting is only 2.3%, meaning that own-price elasticities are almost perfectly captured by $\epsilon(p_j, r_j) \equiv p_j \eta(p_j, r_j)$, which in their setting is just a function of p_j, because that is the only component with a random coefficient. Their specification falls victim to the same critique that Gandhi and Nevo (2021) level at the own-price elasticities of the standard logit model.

Figure 1: Own-price elasticities when only price has a random coefficient



Notes: This specification corresponds to the "Price only" row in Table 2 for the synthetic cereal data, in which the mixed logit only has a normally distributed random coefficient on price. The figure shows the elasticities and small-π approximations for markets with the 10th, 50th, and 90th percentile of average (across goods) relative error.

Figure 1 illustrates these points in the synthetic cereal data modified to only have a random price coefficient, using three markets chosen by their average relative error. Each dot is the own-price elasticity $\epsilon_j^\star(\boldsymbol{p}, \boldsymbol{c}, \boldsymbol{\xi})$ produced by the mixed logit model, while the line is the small-π approximation $\epsilon(p_j) \equiv p_j \eta(p_j)$ from Proposition 1. The crosses are own-price elasticities from the comparable standard logit model together with their small-share approximation (5). For all three markets, the standard logit and mixed logit produce own-price elasticities that are much different. Yet, in all three cases, the own-price elasticities are quite nearly a function of own price, consistent with both approximations. The few products that deviate from the approximation are those with unusually large market shares; the two outliers in the 90th percentile market have market shares of 16% and 10%. Both the logit and mixed logit approximations work less well for these two products, even while they continue to work well for the rest of the products, which have smaller market shares.

3.3 Independent random coefficients

Many applications of mixed logit models also include non-price characteristics with random coefficients. As we saw from Table 1, it is common in empirical work to assume that all of these coefficients β are independent of the random price coefficient,

α. The next proposition shows that if this is done, then own-price elasticities are still closely approximated by functions of price alone.

Proposition 2. Let $\bar{r} \equiv J^{-1} \sum_{j=1}^{J} r_j$ be the average characteristics among the inside products. Let $\bar{\eta}(p_j) \equiv \eta(p_j, \bar{r})$. Let $r_j(\tau) \equiv \bar{r} + \tau(r_j - \bar{r})$ for $\tau \in [0,1]$. Suppose that (α, β) has density $f(a,b)$. Then

$$\eta(p_j, r_j) = \bar{\eta}(p_j) + \iota_j^{\text{DEP}} \quad \text{where} \quad \iota_j^{\text{DEP}} \equiv \int_0^1 (r_j - \bar{r})' \, \mathbb{C}^h_{p_j, r_j(\tau)}[\alpha, \beta] \, d\tau.$$

Here, $\mathbb{C}^h_{p,r}[\alpha, \beta]$ denotes the covariance between (α, β) when they are distributed according to the density

$$h_{p,r}(a,b) \;\propto\; \overbrace{\pi_0(a,b)}^{\text{outside choice probability}} \underbrace{\exp(pa + r'b) f(a,b)}_{\text{exponential tilting}}, \tag{8}$$

where we use the "proportional to" symbol $\propto$ as a shorthand to avoid writing out the normalizing constant of the density.

Proposition 2 shows that the dependence of $\eta(p_j, r_j)$ on r_j is tied to the covariance between α and β.[6] This is not the direct covariance of (α, β) according to their density f, but a covariance computed under the density h defined in (8), which is tilted around the contribution of the characteristics (p, r) with random coefficients, then weighted by the choice probability of the outside option. Suppose that this choice probability were close to one ($\pi_0(a,b) \approx 1$), so that only the exponential tilting remained. If α and β are independent, so that $f(a,b) = f_\alpha(a) f_\beta(b)$, then

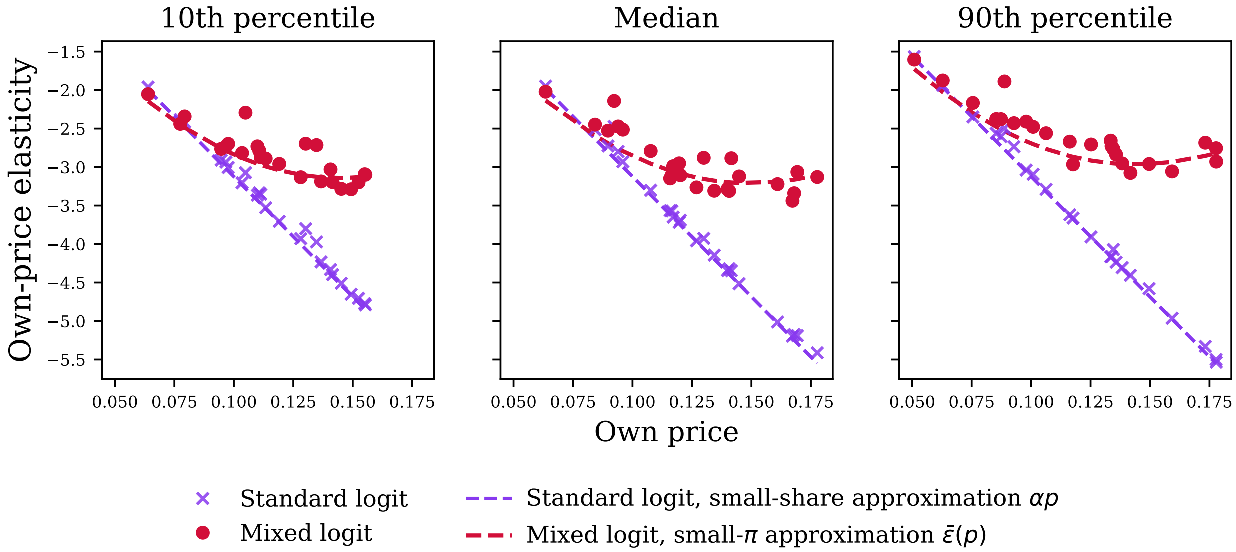
$$h_{p,r}(a,b) \propto_{\text{approx.}} \exp(pa) f_\alpha(a) \times \exp(r'b) f_\beta(b), \tag{9}$$

so that α and β would be independent under the tilted density as well. In this case, $\mathbb{C}^h_{p_j, r_j(\tau)}[\alpha, \beta]$ would be zero for all τ and Proposition 2 would imply that $\eta(p_j, r_j) = \bar{\eta}(p_j)$ is only a function of p_j.

The choice probability of the outside option is not one—$\pi_0(a,b) < 1$ for some or all (a,b)—so the preceding argument does not hold exactly. However, the intuition it

[6] Choosing $\bar{r}$ as the point of evaluation for $\bar{\eta}(p_j) \equiv \eta(p_j, \bar{r})$ is a transparent way to choose a characteristic vector that is central among the inside products. However, the conclusion of Proposition 2 holds for any product-invariant choice $r^\star$. Proposition 3 ahead shows that making $r^\star$ centrally located, such as by setting $r^\star = \bar{r}$, will tend to make ι_j^{DEP} smaller in magnitude. We have investigated ways to choose $r^\star$ optimally, but found that this provides little improvement in our empirical examples, so we stick with the simpler evaluation choice $\bar{r}$.

Figure 2: Own-price elasticities with independent random coefficients



Notes: This specification corresponds to the "Independent" row in Table 2 for the synthetic cereal data, which is the same as the baseline synthetic cereal except that the random coefficients have been made independent. See Figure 1 for further notes.

carries suggests that ι_j^{DEP} should be small in magnitude when α and β are independent. This intuition is widely supported by the relative error results in Table 2 for the columns titled $\text{ERR}(\bar{\eta})$. All of the baseline applications except for synthetic cereal have independent random coefficients. While the error is in most cases larger than just the small-π approximation error $\text{ERR}(\eta)$, it is still generally quite small, and in many cases not appreciably larger than the small-π approximation.

For example, the median market-price combination in the US automobile data has a relative error of only 2.1%. For the synthetic cereal data with independent random coefficients (second subrow), the median relative error increases from 2.6% for just the small-π approximation that removes ι_j^{SCP} to 4.9% when also removing ι_j^{DEP}. Figure 2 illustrates this specification for three synthetic cereal markets, showing that own-price elasticities are still tightly tracked by the function $\bar{\epsilon}(p)$ that the intuition around (8) suggests should be a good approximation. The entries in Table 2 show that the errors that do arise are primarily from the small-π approximation, not the independence characterized by Proposition 2.

The next proposition develops some theoretical bounds on the magnitude of ι_j^{DEP}. This helps provide intuition about the forces that tend to make $\eta(p_j, r_j) \approx \bar{\eta}(p_j)$ a high-quality approximation when α and β are independent.

Proposition 3. Suppose that α and β are independent with density $f(a,b) = f_\alpha(a) f_\beta(b)$. Suppose that f_β is normally distributed. If f_α is also normal, then

$$|\iota_j^{\text{DEP}}| \leq \mathbb{V}[\alpha] \left((r_j - \bar{r})' \, \mathbb{V}[\beta](r_j - \bar{r})\right)^{1/2} \sup_{(a,b)} \left(\mathfrak{c}(a,b)' \, \mathbb{V}[\beta]\mathfrak{c}(a,b)\right)^{1/2}, \tag{10}$$

where $\mathfrak{c}(a,b)$ denotes the probability-weighted covariance vector between p_j and r_j:

$$\mathfrak{c}(a,b) \equiv \left(\sum_{j=0}^{J} \pi_j(a,b) p_j r_j\right) - \left(\sum_{j=0}^{J} \pi_j(a,b) p_j\right)\left(\sum_{j=0}^{J} \pi_j(a,b) r_j\right).$$

If $-\alpha$ is log-normal, so that $\log(-\alpha)$ is normal, then

$$|\iota_j^{\text{DEP}}| \leq \frac{\mathbb{V}[\alpha] \left((r_j - \bar{r})' \, \mathbb{V}[\beta](r_j - \bar{r})\right)^{1/2} \sup_{(a,b)} \left(\mathfrak{c}(a,b)' \, \mathbb{V}[\beta]\mathfrak{c}(a,b)\right)^{1/2}}{\inf_{\tau \in [0,1]} \, \mathbb{E}[\pi_0(\alpha,\beta) \times \exp(\alpha p_j + r_j(\tau)'\beta) / \, \mathbb{E}[\exp(\alpha p_j + r_j(\tau)'\beta)]]}, \tag{11}$$

where $r_j(\tau) \equiv \bar{r} + \tau(r_j - \bar{r})$, as in Proposition 2.

The bounds in Proposition 3 illustrate some factors that lead ι_j^{DEP} to be small in magnitude. The bounds are tighter when the probability-weighted covariance between prices and other characteristics is small at every configuration of the choice probabilities $\{\pi_j(a,b)\}_{j=0}^{J}$ produced by values of α and β. In particular, this term will tend to be small if the correlation between prices and other observed characteristics is small. The bounds in Proposition 3 will also be tighter for products with non-price characteristics $r_j \approx \bar{r}$ near the average characteristic. Both factors get inflated by the variances of α and β, which deteriorate the bound. In the case that $-\alpha$ has a log-normal distribution, the bound derived in Proposition 3 has an additional denominator reflecting a weighted average of the choice probability of the outside option.

The bounds given in Proposition 3 are derived from a Cauchy-Schwarz inequality followed by a supremum that together can be quite loose in practice. The bounds do not need to be used in practice because the exact value of the error ι_j^{DEP} can be computed directly. However, Proposition 3 is useful in providing intuition on sufficient conditions for ι_j^{DEP} to be small, even if these sufficient conditions are not necessary. The primary factor is the variance of the random coefficients: larger variances mean that smaller deviations from $\pi_0(\alpha,\beta) \approx 1$ lead to substantive dependence between α and β through (8).

None of the variances in the applications used in Table 2 are small from an empirical standpoint (see Appendix SA.3). With the exception of the synthetic cereal subrows, they are all estimates that we have taken directly from the authors who used them; we did not choose these parameter values. The synthetic cereal data shows that multiplying the random coefficient variances by a factor of four leads to a degraded approximation, consistent with Proposition 3. In the case of α, this degradation is large only when the distribution is normal, which leads to price-loving (positive α) individuals who explode the contribution of $\exp(p_j\alpha)$ in (8). Changing the distribution to log-normal restores the approximation even with a larger variance. For β, a four-fold increase in variance would imply a setting in which consumers are mostly heterogeneous in their taste for mushy and sugar cereals, rather than price. Even so, comparing the $\text{ERR}(\eta)$ and $\text{ERR}(\bar{\eta})$ columns in Table 2 shows that most of the error would come from degradation of the small-π error ι_j^{SCP}, rather than from the independence error ι_j^{DEP}.

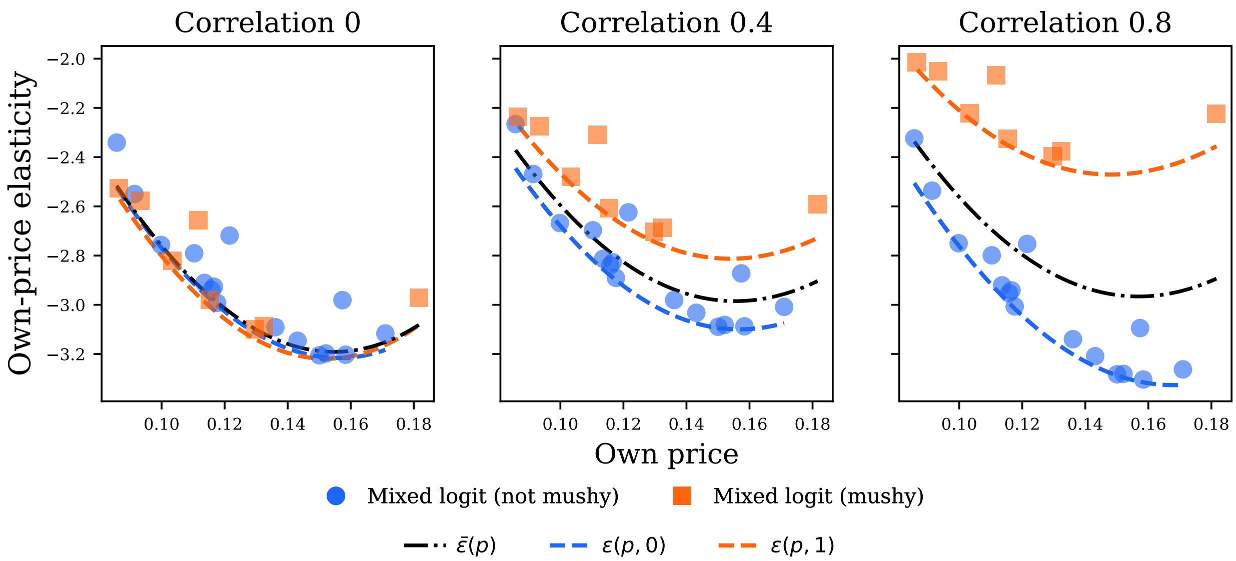
Taken together, Propositions 2 and 3 show that the negative implications of Proposition 1 for specifications that only have a random coefficient on price also apply to specifications with independent random coefficients. As we documented in Table 1, this constitutes a large share of mixed logit specifications used in practice. The implication is that many specifications of the mixed logit model used in practice are still subject to the same critique that Gandhi and Nevo (2021) level at the standard logit model: own-price elasticities are effectively functions of own prices.

3.4 Correlated random coefficients

The negative implications of Proposition 2 can be overturned by allowing for correlation between α and β. Proposition 1 still applies, showing that two goods with the same price $p_j = p_k$ can have different elasticities to the extent that they differ in other characteristics $r_j \neq r_k$ that have random coefficients. The synthetic cereal data is the only baseline specification in Table 2 that allows for correlation between the random coefficients; it uses the same estimates found by Conlon and Gortmaker (2020) in their `pyblp` tutorial. Comparing the $\text{ERR}(\eta)$ and $\text{ERR}(\bar{\eta})$ columns shows that the amount of correlation found in their estimates only provides a small departure from the conclusion that own-price elasticities are well-characterized by functions of own prices alone.

Figure 3 shows a stylized modification of the synthetic cereal specification in which the only two random coefficients are on price and mushy (a binary indicator for cereal texture). We adjust the correlation between these two coefficients across the three panels. When there is no correlation, a single function of own price fits all elasticities well, consistent with Propositions 2 and 3. This fit starts to degrade with a correlation of .4 and breaks down entirely when the correlation is .8. While a *single* function fits poorly in the latter case, separate functions for mushy and non-mushy continue to fit well, consistent with Proposition 1. For comparison, the correlation between

Figure 3: Own-price elasticities with correlated random coefficients



Notes: The baseline synthetic cereal specification is modified by keeping random coefficients only on price and mushy, then adjusting their correlation. We show results for the market that has the median average relative error across products when the correlation is zero.

the price and mushy coefficients in the baseline specification estimated by Conlon and Gortmaker (2020) is only .09. Allowing for correlation between random coefficients can only break the functional relationship between own-price elasticities and price if this correlation is also found to be strong in the data.

Conlon and Gortmaker (2020) also consider a specification with demographics, in which off-diagonals of the variance-covariance matrix of (α, β) are not directly estimated, but correlation is nonetheless induced through the demographics. The last row of Table 2 shows that this specification exhibits a more substantial departure from the conclusion that own-price elasticities are well-characterized by functions of own prices alone. This seems to be an unexpected beneficial byproduct of including demographics. The usual motivation for demographics is as an observable form of randomness without considering their role in inducing correlation across random coefficients (for example, Gandhi and Nevo, 2021, pg. 76).

3.5 Quadratic approximation

In addition to being well-approximated by functions of price alone, the elasticities shown in Figures 1–3 also tend to look like upward-opening quadratic functions. Table 3 quantifies this phenomenon by reporting the distribution of R-squareds from the

Table 3: The distribution of R^2 across markets from OLS fits

	$\text{R}^2(\epsilon^\star)$			$\text{R}^2(\bar{\epsilon})$			$\text{R}^2(\eta^\star)$			$\text{R}^2(\bar{\eta})$		
Percentile:	90	50	10	90	50	10	90	50	10	90	50	10
US autos	0.970	0.930	0.839	0.981	0.952	0.916	0.997	0.996	0.990	0.999	0.998	0.995
China autos	0.999	0.997	0.994	0.999	0.998	0.995	1.000	1.000	1.000	1.000	1.000	1.000
France autos	1.000	0.998	0.993	1.000	1.000	1.000	0.996	0.981	0.906	1.000	1.000	1.000
Synthetic cereal	0.776	0.595	0.414	1.000	1.000	0.999	0.878	0.809	0.741	1.000	1.000	1.000
Price only	0.967	0.921	0.777	1.000	1.000	1.000	0.991	0.979	0.941	1.000	1.000	1.000
Independent	0.840	0.683	0.545	1.000	1.000	0.999	0.915	0.859	0.790	1.000	1.000	1.000
Independent, lognormal	0.962	0.912	0.817	1.000	1.000	1.000	0.870	0.771	0.617	1.000	1.000	0.999
Independent, $\mathbb{V}[\alpha] \times 4$	0.911	0.816	0.682	1.000	1.000	0.998	0.950	0.918	0.869	1.000	1.000	0.999
Independent, $\mathbb{V}[\alpha] \times 4$*, lognormal*	0.838	0.646	0.403	1.000	0.999	0.997	0.935	0.886	0.819	1.000	1.000	0.999
Independent, $\mathbb{V}[\alpha] \div 4$	0.974	0.951	0.890	1.000	1.000	1.000	0.658	0.422	0.223	1.000	1.000	1.000
Independent, $\mathbb{V}[\beta] \times 4$	0.740	0.540	0.362	1.000	1.000	1.000	0.748	0.648	0.449	1.000	1.000	1.000
Independent, $\mathbb{V}[\beta] \div 4$	0.937	0.864	0.756	1.000	1.000	1.000	0.978	0.953	0.905	1.000	1.000	1.000
Price and mushy, correlation .2	0.946	0.881	0.730	1.000	1.000	1.000	0.981	0.958	0.900	1.000	1.000	1.000
Price and mushy, correlation .8	0.618	0.461	0.308	1.000	1.000	1.000	0.670	0.572	0.444	1.000	1.000	1.000
Independent, demographics	0.857	0.665	0.338	1.000	1.000	0.999	0.282	0.119	0.031	1.000	1.000	0.999

Notes: The first two columns report the 90th, 50th, and 10th percentile across markets of the product-level R^2 *from regression (12). The second two columns report these statistics for regressions where the regressand is* $\eta^\star_j(\boldsymbol{p}, \boldsymbol{r}, \boldsymbol{c}, \boldsymbol{\xi})$ *or* $\bar{\eta}(p_j)$ *and the regressors are* $1/p_j$*, 1, and* p_j*.*

following market-level ordinary least squares (OLS) fits:

$$\min_{b_0, b_1, b_2} \sum_{j=1}^{J} \left(\text{LHS}_j - b_0 - b_1 p_j - b_2 p_j^2\right)^2 \quad \text{for } \text{LHS}_j = \bar{\epsilon}(p_j) \text{ or } \text{LHS}_j = \epsilon^\star_j(\boldsymbol{p}, \boldsymbol{r}, \boldsymbol{c}, \boldsymbol{\xi}). \tag{12}$$

When taking LHS_j to be the actual mixed logit elasticity, the median R-squared is .93 in the US auto examples and even higher in the China and France auto examples. In the synthetic cereal specifications, the median is always above .60 for specifications in which the price-only relative error $\text{ERR}(\bar{\eta})$ of Table 2 is small. Taking $\text{LHS}_j = p_j \bar{\eta}(p_j)$ to be our price-only small-π approximation—which strips out the small-π and dependence terms ι_j^{SCP} and ι_j^{DEP}—leads to R-squareds that reflect almost perfect fit in nearly all cases. In this section, we provide some theory that explains this phenomenon.[7]

For simplicity, we start by considering the case in which price is the only characteristic with a random coefficient, so that r_j is null. As part of proving Proposition 1, we showed that

$$\eta(p_j) = \mathbb{E}^h_{p_j}[\alpha], \tag{13}$$

[7]We describe this phenomenon as "quadratic in own price," which presumes that price enters (1) linearly, as it does in the vast majority of applications. If price instead enters through a transformation, the same arguments apply with the transformed price playing the role of price, so own-price elasticities remain stylized functions of own price alone, although not necessarily quadratic ones. One example is the China auto application (Li, 2018), which uses log price; see Appendix SA.3.3 for more detail.

where $\mathbb{E}^h_{p_j}$ denotes expectation with respect to the density h_{p_j} defined in (8), which we previously discussed in the context of Proposition 2. These densities h_p form an exponential family indexed by p, a property that appears to be intimately related to the structure of the logit model (see Appendix SA.2.2).

The density $h_p(a)$ differs from the original density $f(a)$ of α in two ways: through the exponential tilting by $\exp(ap)$ and through the weighting by the outside option $\pi_0(a)$. Suppose that $\pi_0(a) \approx 1$—the same thought experiment we considered in (9)—so that the weighting can be ignored. If α were normally distributed under f, then exponentially tilting it by $\exp(ap)$ would keep its distribution normal but shift its mean to $\mathbb{E}^h_p[\alpha] = \mathbb{E}[\alpha] + \mathbb{V}[\alpha]p$. The implication from (13) is that $\eta(p)$ would be linear in p. If $|\iota^{\mathrm{SCP}}_j|$ is small, we would conclude from Proposition 1 that

$$\epsilon^\star_j(\boldsymbol{p}, \boldsymbol{c}, \boldsymbol{\xi}) = p_j \eta^\star_j(\boldsymbol{p}, \boldsymbol{c}, \boldsymbol{\xi}) \approx p_j \eta(p_j) = \mathbb{E}[\alpha] p_j + \mathbb{V}[\alpha] p_j^2, \tag{14}$$

so that own-price elasticities are quadratic in p_j. The shape of the quadratic is upward-opening because $\mathbb{V}[\alpha] > 0$.

The clean expression in (14) depends on the simplifying assumptions that α is normal and $\pi_0(a) \approx 1$. But the conclusion really rests on (13) together with the more general fact that h_p is an exponential family, which is a consequence of the logit structure. A well-known property of exponential families is that their cumulants are tied together as being higher-order derivatives of a single function called the log partition function. In our context, this property implies that

$$\eta(p) = \mathbb{E}^h_p[\alpha], \qquad \frac{d}{dp}\eta(p) = \mathbb{V}^h_p[\alpha], \qquad \frac{d^2}{dp^2}\eta(p) = \mathbb{E}^h_p[(\alpha - \mathbb{E}^h_p[\alpha])^3], \qquad \ldots \tag{15}$$

with the general pattern being that the nth derivative of η with respect to p is the nth cumulant of α under the distribution h_p (see Lemma EFC in Appendix SA.2.2).

If h_p is normal, then the third and higher cumulants disappear, implying that $\eta(p)$ is exactly linear and $\epsilon(p) = p\eta(p)$ is exactly quadratic, as we found in (14). If h_p is non-normal, then it will have non-zero third and higher cumulants. However, because h_p is an exponential family, its cumulants cannot grow too quickly; the log partition function is an exceptionally smooth (analytic) function (see, for example, Brown, 1986, Chapter 2). As a consequence, $\eta(p)$—the first derivative of the log partition function—should be well-approximated by a low-dimensional polynomial around some central price like the mean, $\bar{p}$, if there isn't too much price dispersion.[8]

[8]Fox et al. (2012) exploited the observation that the type-specific choice probabilities are analytic to provide nonparametric identification conditions for mixed logit models. Our use of analyticity is related

We now use this reasoning to build a quadratic approximation to the function $\epsilon(p,r) = p\eta(p,r)$ in the general case with random coefficients on both price and non-price characteristics. If η is smooth, then ϵ will be smooth as well. If our goal is to approximate ϵ as a quadratic built from η, then we also need to include a p^{-1} term in $\eta(p,r)$, to account for a potential missing intercept. In particular, we take our approximation of η to be

$$\eta_{\mathrm{Q}}(p,r) \equiv \frac{\theta_{00} + \theta_{01}'(r-\bar{r})}{p} + \theta_{10} + \theta_{11}'(r-\bar{r}) + \theta_2 p, \tag{16}$$

so that the implied elasticity approximation is the quadratic function

$$\epsilon_{\mathrm{Q}}(p,r) \equiv \theta_{00} + \theta_{01}'(r-\bar{r}) + \theta_{10}p + \theta_{11}'(r-\bar{r})p + \theta_2 p^2. \tag{17}$$

In the next proposition, we choose the coefficients $\theta_{00}, \theta_{01}, \theta_{10}, \theta_{11}$, and θ_2 to match the level of η and four of its derivatives at the mean characteristics $(\bar{p}, \bar{r})$.[9]

Proposition 4. Recall that $\mathbb{C}^h_{p,r}[\alpha,\beta]$ denotes the covariance operator with respect to the density $h_{p,r}$ defined in (8). Also define the following third order cumulants with respect to $h_{p,r}$:

$$\begin{aligned} &\mathbb{K}^h_{p,r}[\alpha,\alpha,\alpha] \equiv \mathbb{E}^h_{p,r}\left[(\alpha - \mathbb{E}^h_{p,r}[\alpha])^3\right] \\ \text{and} \quad &\mathbb{K}^h_{p,r}[\alpha,\alpha,\beta] \equiv \mathbb{E}^h_{p,r}\left[(\alpha - \mathbb{E}^h_{p,r}[\alpha])^2(\beta - \mathbb{E}^h_{p,r}[\beta])\right], \end{aligned}$$

where the latter cumulant is a vector if β is a vector. Suppose that we take

$$\begin{aligned} &\theta_{00} = \tfrac{1}{2}\bar{p}^3\ \mathbb{K}^h_{\bar{p},\bar{r}}[\alpha,\alpha,\alpha], && \theta_{01} = -\bar{p}^2\ \mathbb{K}^h_{\bar{p},\bar{r}}[\alpha,\alpha,\beta], \\ &\theta_{10} = \mathbb{E}^h_{\bar{p},\bar{r}}[\alpha] - \bar{p}\ \mathbb{V}^h_{\bar{p},\bar{r}}[\alpha] - \bar{p}^2\ \mathbb{K}^h_{\bar{p},\bar{r}}[\alpha,\alpha,\alpha], && \theta_{11} = \mathbb{C}^h_{\bar{p},\bar{r}}[\alpha,\beta] + \bar{p}\ \mathbb{K}^h_{\bar{p},\bar{r}}[\alpha,\alpha,\beta], \\ \text{and}\quad &\theta_2 = \mathbb{V}^h_{\bar{p},\bar{r}}[\alpha] + \tfrac{1}{2}\bar{p}\ \mathbb{K}^h_{\bar{p},\bar{r}}[\alpha,\alpha,\alpha]. \end{aligned}$$

only in the sense that both are derived from the exponential structure produced by the idiosyncratic logit (type I extreme value) shocks. The connection to exponential tilting is not used by Fox et al. (2012) and the substance of the contributions is entirely distinct; our analysis is not about identification but about the structure of the mixed logit model, regardless of how the parameters are determined.

[9] As with Proposition 2, evaluating around the mean seems like a good heuristic, but we could choose to evaluate around any product-invariant point for each market.

Table 4: Relative error of quadratic approximations

	$\text{ERR}(\eta)$			$\text{ERR}(\bar{\eta}_{\text{Q}})$			$\text{ERR}(\eta_{\text{OLS}})$			$\text{ERR}(\epsilon_{\text{OLS}})$		
Percentile:	10	50	90	10	50	90	10	50	90	10	50	90
US autos	1.0	1.1	1.5	7.5	17.3	32.5	3.1	4.2	7.1	2.5	3.8	5.2
China autos	0.1	0.1	0.1	0.4	0.5	0.7	0.5	0.6	0.8	0.3	0.3	0.4
France autos	0.3	1.1	1.7	0.3	1.1	1.7	0.1	0.4	0.7	0.1	0.4	0.7
Synthetic cereal	4.2	4.7	5.1	7.0	7.8	8.6	6.2	7.1	8.3	6.2	7.1	8.3

Notes: See notes for Table 2. In this table, we first compute mean relative error for each market, then report the 10th, 50th, and 90th percentiles across markets for each application. The column $\text{ERR}(\epsilon_{\text{OLS}})$ *is the relative error of the fit from (12) with* $\text{LHS}_j = \epsilon_j^\star(\boldsymbol{p}, \boldsymbol{r}, \boldsymbol{c}, \boldsymbol{\xi})$. *The column* $\text{ERR}(\eta_{\text{OLS}})$ *is the relative error of the fit from the corresponding semi-elasticity regression where the regressand is* $\eta_j^\star(\boldsymbol{p}, \boldsymbol{r}, \boldsymbol{c}, \boldsymbol{\xi})$ *and the regressors are* $1/p_j, 1,$ *and* p_j. *Note that the relative errors for these OLS fits are different, in contrast to the relative errors for semi-elasticities and elasticities formed through direct multiplication by price.*

Then $\eta_{\text{Q}}(\bar{p}, \bar{r}) = \eta(\bar{p}, \bar{r})$ and

$$\frac{\partial}{\partial p}\eta_{\text{Q}}(\bar{p}, \bar{r}) = \frac{\partial}{\partial p}\eta(\bar{p}, \bar{r}), \qquad \frac{\partial}{\partial r_\ell}\eta_{\text{Q}}(\bar{p}, \bar{r}) = \frac{\partial}{\partial r_\ell}\eta(\bar{p}, \bar{r}) \quad \text{for } \ell = 1, \ldots, L,$$
$$\frac{\partial^2}{\partial p^2}\eta_{\text{Q}}(\bar{p}, \bar{r}) = \frac{\partial^2}{\partial p^2}\eta(\bar{p}, \bar{r}). \qquad \frac{\partial^2}{\partial p\,\partial r_\ell}\eta_{\text{Q}}(\bar{p}, \bar{r}) = \frac{\partial^2}{\partial p\,\partial r_\ell}\eta(\bar{p}, \bar{r}) \quad \text{for } \ell = 1, \ldots, L,$$

where $\ell = 1, \ldots, L$ indexes the characteristic vector r.

The third set of columns in Table 2 report the relative error $\text{ERR}(\eta_{\text{Q}})$ of the quadratic approximation in Proposition 4. The results are nearly the same as for the small-π approximation η. While Proposition 1 allowed for the possibility that own-price elasticities vary flexibly as a function of own price for a given product, the quality of the quadratic approximation shows that this is not actually true. Even if r_j contained a set of product indicators—a specification likely too rich to be able to estimate in practice—the quadratic approximation says that the own-price elasticity would still behave like a quadratic function in own price.

The fourth set of columns in Table 2 reports the relative error $\text{ERR}(\bar{\eta}_{\text{Q}})$ of a price-only approximation that sets $r = \bar{r}$:

$$\bar{\eta}_{\text{Q}}(p) \equiv \eta_{\text{Q}}(p, \bar{r}) = \frac{\theta_{00}}{p} + \theta_{10} + \theta_2 p \quad \text{and} \quad \bar{\epsilon}_{\text{Q}}(p) \equiv \epsilon_{\text{Q}}(p, \bar{r}) = \theta_{00} + \theta_{10}p + \theta_2 p^2. \quad (18)$$

If α and β are independent, then we expect $\bar{\eta}_{\text{Q}}$ to be close to η_{Q}, because the coefficients θ_{01} and θ_{11} are determined by the covariance and third cross-cumulants between α and β, which should be small under independence.

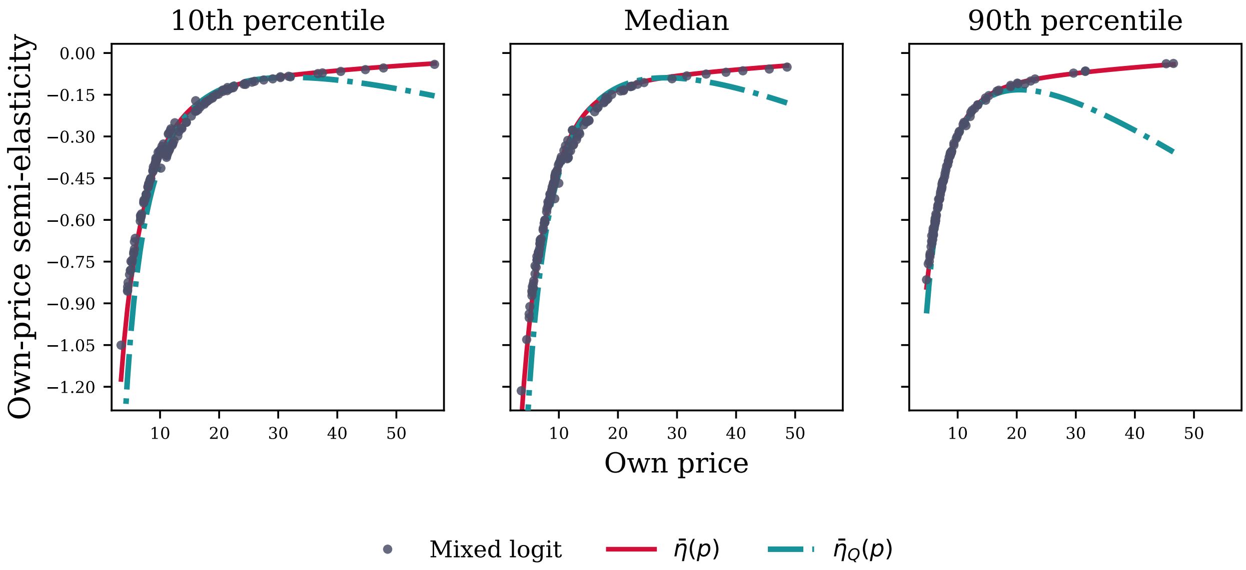
Table 2 bears out this prediction, together with the consequence that $\bar{\eta}_{\text{Q}}$ tends to be a good approximation for the true semi-elasticities under independence. Figures SA.1–SA.4 illustrate by plotting $\bar{\epsilon}_{\mathrm{Q}}$ against actual elasticities for nine markets in the US auto, China auto, France auto, and cereal applications, respectively. Table 4 summarizes by showing the distribution of market-level average relative error for each application compared to both the small-π approximation η and OLS approximations produced from (12). An important point evident in Table 4 is that even for the US auto application—the one case in which our quadratic approximation appears to work more poorly (although see below)—the OLS fits still have low relative error. We conclude that own-price elasticities are well-approximated by quadratic functions of own price in the widely-used class of mixed logit specifications that have independent random coefficients. Usually Proposition 4 does an excellent job of describing this approximation, but the conclusion is true even when it does not.

Moreover, a closer examination of the US auto application shows that the ostensibly worse results are an artifact of using relative error. The reason is that its own-price semi-elasticities are nearly zero for high prices, making relative error a sensitive measure of precision in this application. Figure 4 shows that the *absolute* error in the price-only semi-elasticity approximation $\bar{\eta}_{\mathrm{Q}}$ is still small even for the market with the 90th percentile of average relative error. To the extent that the approximation deteriorates, it does so for $\bar{\eta}_{\mathrm{Q}}$ at high prices, which is to be expected for a global approximation. The curves for $\bar{\eta}$ show that even so, the own-price semi-elasticities are closely tracked by a function of own prices alone.

Part of the difference in the US auto application stems from the coefficient on θ_2, which controls the curvature of $\bar{\eta}_{\mathrm{Q}}$. This coefficient has an intuitive structure that depends on the variance and third cumulant of α (under h at $\bar{p}$). The variance of α—a coefficient which is expected to be negative—reflects a selection effect: more expensive products are bought by less price-sensitive consumers, leading elasticities to tend towards zero for higher prices. The third cumulant reflects how this selection effect changes with prices (recall (15)). If α is symmetric, so that the third cumulant is approximately zero, then the selection effect doesn't depend on the evaluation point $\bar{p}$ and the quadratic is upward-opening.

On the other hand, if the distribution of α has a sufficiently strong left skew, then the third cumulant can be sufficiently negative to make $\theta_2 < 0$, causing the quadratic approximation to become downward-opening. A left skew implies from (18) that semi-elasticities always increase towards zero as prices increase but at a decreasing rate. Multiplying these semi-elasticities by increasing prices produces elasticities that mechanically start to become more negative as prices increase. This can lead to $\theta_2 < 0$ if the average price $\bar{p}$ is sufficiently large. Figure 4 shows how this happens in the US

Figure 4: Own-price semi-elasticities for the US automobile application



Notes: The figure shows the semi-elasticities, price-only approximations, and quadratic price-only approximations for markets with the 10th, 50th, and 90th percentile of average (across goods) relative error.

auto application, where the distribution of price sensitivity is a re-scaled version of the inverse income distribution, which is highly skewed.

3.6 Implications for derived quantities

The stylized own-price elasticities produced by commonly-used mixed logit models are often used as downstream inputs into economic quantities of interest.

For example, suppose that each product is produced by a single-product firm with constant marginal cost $\textsc{mc}_j$ and that firms engage in Bertrand-Nash competition in prices. Firm j's optimal price is

$$p_j = \frac{\textsc{mc}_j}{1 + 1/\epsilon_j^\star(\boldsymbol{p}, \boldsymbol{r}, \boldsymbol{c}, \boldsymbol{\xi})}. \tag{19}$$

Marginal costs can be backed out from (19) given prices and estimates of own-price elasticities:

$$\textsc{mc}_j = p_j \left(1 + \frac{1}{\epsilon_j^\star(\boldsymbol{p}, \boldsymbol{r}, \boldsymbol{c}, \boldsymbol{\xi})}\right). \tag{20}$$

Firm j's markup can also be inferred directly from their own-price elasticity:

$$\text{MUP}_j \equiv \frac{p_j - \text{MC}_j}{p_j} = -\frac{1}{\epsilon_j^\star(\boldsymbol{p}, \boldsymbol{r}, \boldsymbol{c}, \boldsymbol{\xi})}. \tag{21}$$

This type of analysis is commonly used to infer important economic quantities like markups from demand estimates when costs are not observed by using assumptions about firm conduct.

Our results show that using mixed logit models for this analysis can continue to bake in stylized findings, just as with the standard logit model. For the standard logit, inserting the small-share approximation (5) into (20) and (21) implies that firms' marginal cost differences are the same as their price differences and that markups are necessarily decreasing in prices. The price-only quadratic approximation $\bar{\eta}_{\text{Q}}$ given in (18) leads to conclusions that are only mechanically different: marginal costs are a concave (inverted-U-shaped) function of price when α has low skewness. Both models imply that marginal costs are determined by prices. Markups are similarly stylized: if α has low skewness, then they will *necessarily* be high for both low-priced and high-priced firms, and lowest for firms with mid-range prices, where the upward-opening quadratic makes demand most elastic.

The same type of conclusion also applies to the pass-through of cost shocks. In a small-π regime with independent random coefficients, marginal costs are

$$\text{MC}_j = p_j + \frac{1}{\eta_j^\star(\boldsymbol{p}, \boldsymbol{r}, \boldsymbol{c}, \boldsymbol{\xi})} \approx p_j + \frac{1}{\bar{\eta}(p_j)}.$$

Suppose that equilibrium feedback from rival firms adjusting prices is small. Then the implied pass-through of a cost shock to product j is

$$\text{PT}_j \equiv \frac{dp_j}{d\text{MC}_j} = \left(1 - \frac{\frac{d}{dp}\bar{\eta}(p_j)}{\bar{\eta}(p_j)^2}\right)^{-1} = \left(1 - \frac{\mathbb{V}^h_{p_j,\bar{r}}[\alpha]}{\mathbb{E}^h_{p_j,\bar{r}}[\alpha]^2}\right)^{-1}, \tag{22}$$

where the final equality uses (15), as applied to $\bar{\eta}(p) \equiv \eta(p, \bar{r})$. The standard logit falls out as a special case with $\mathbb{V}^h_{p,\bar{r}}[\alpha] = 0$, so that its pass-through is approximately one for all goods.

For the mixed logit, (22) says that the pass-through exceeds one, producing over-shifting for all goods that is implied by the model. This finding is consistent with empirical results in Compiani (2022), who finds that mixed logit estimates generate pass-through estimates that are considerably larger than those from his nonparametric method, albeit for an example with $J = 2$ where the small-π approximation may be poor. Moreover, from (15) we know that $\mathbb{E}^h_{p,\bar{r}}[\alpha]$ is an increasing function of p, so that $\mathbb{E}^h_{p,\bar{r}}[\alpha]^2$ is a decreasing function of p in the likely case that $\alpha < 0$. If $\mathbb{V}^h_{p,\bar{r}}[\alpha]$ is approximately constant as a function of p—which (15) shows happens when α has low skewness—then (22) implies that pass-through should be increasing with prices. While overshifting may be plausible in some settings, the assumptions that pass-through must exceed one and increase with prices are strong restrictions relative to the range of patterns documented in the pass-through literature (Fabinger and Weyl, 2018; Miravete et al., 2025).

Some quantities derived from elasticities depend on both own- and cross-price elasticities. A leading example is a diversion ratio, which is a scaled ratio of cross-price to own-price elasticities. Our results imply that these too will be stylized quantities to the extent that they depend on the own-price elasticity component. As we show in the next section, cross-price elasticities in the mixed logit model tend to be considerably less stylized and accord more with the intuition originally used to motivate the use of these models for discrete choice demand.

4 Cross-Price Elasticities

The cross-price elasticity of good k with respect to good j is

$$\epsilon^\star_{kj}(\boldsymbol{p}, \boldsymbol{r}, \boldsymbol{c}, \boldsymbol{\xi}) \equiv \frac{\partial s_k}{\partial p_j}\frac{p_j}{s_k} = -\frac{\mathbb{E}[\alpha p_j \pi_j(\alpha,\beta)\pi_k(\alpha,\beta)]}{\mathbb{E}[\pi_k(\alpha,\beta)]} \equiv p_j \eta^\star_{kj}(\boldsymbol{p}, \boldsymbol{r}, \boldsymbol{c}, \boldsymbol{\xi}), \tag{23}$$

where $\eta^\star_{kj}$ is the cross-price semi-elasticity. The original motivation for using mixed logit in discrete choice demand modeling focused heavily on the restrictive properties of the standard logit model for cross-price elasticities; see Berry (1994, pg. 245) and Berry et al. (1995, pg. 847). These restrictive properties can be seen in (23): if α and β are non-random then $\pi_k(\alpha,\beta)$ cancels in the numerator and denominator, so that the cross-price elasticity of good k only depends on the share and price of good j, but not any characteristics of k.

The next proposition modifies our analysis for own-price elasticities to shed new light on the way in which mixed logit models add more flexibility to cross-price elasticities.

Proposition 5. Cross-price semi-elasticities can be written as

$$\eta^\star_{kj}(\boldsymbol{p}, \boldsymbol{r}, \boldsymbol{c}, \boldsymbol{\xi}) = \mathbb{E}\left[\pi_j(\alpha,\beta)\frac{\pi_k(\alpha,\beta)}{\mathbb{E}[\pi_k(\alpha,\beta)]}\right]\eta^\times(p_k + p_j, r_k + r_j), \tag{24}$$

where $\eta^\times$ is a function that only depends on $(p_k + p_j, r_k + r_j)$ for any given market.

Proposition 5 shows that cross-price elasticities share some similarities to own-price elasticities but differ in two important ways. The similarity is in the function $\eta^{\times}$, which has similar properties to the function η for own prices.[10] The first important difference between cross- and own-price elasticities is that (24) is exact, not an approximation that depends on whether type-specific choice probabilities are small. The second important difference is the scaling of $\eta^{\times}$ by the expected choice probability of good j weighted by the choice probability of good k.

It is primarily this weighted probability term that gives the mixed logit model its cross-price flexibility over the standard logit model. This term has a natural economic interpretation in terms of the primitive indirect utility (1) with the idiosyncratic logit shock. Imagine a consumer who makes two choices y_1 and y_2 based on two independent draws of logit shocks, while keeping their draws of (α, β) fixed. Then

$$\mathbb{P}[y_2 = j | y_1 = k] = \frac{\mathbb{E}\left[\mathbb{P}[y_2 = j|\alpha, \beta]\, \mathbb{P}[y_1 = k|\alpha, \beta]\right]}{\mathbb{E}\left[\mathbb{P}[y_1 = k|\alpha, \beta]\right]} = \mathbb{E}\left[\pi_j(\alpha, \beta) \frac{\pi_k(\alpha, \beta)}{\mathbb{E}[\pi_k(\alpha, \beta)]}\right],$$

so that the weighted choice probability term can be interpreted as the conditional probability this consumer chooses j on their second draw given that they chose k on their first draw. The mixed logit links these choices together through the random coefficients (α, β), making a consumer who chose k more likely to then choose j if the characteristics with random coefficients are similar for goods k and j.

The weighting by $\pi_k(\alpha, \beta)$ is key to this interpretation. Its impact will persist even if α and β are independent. The following proposition gives an approximation to the differences that it creates relative to an unweighted average.

Proposition 6. Let $w_j \equiv (p_j, r_j)$ and $\nu \equiv (\alpha, \beta)$. Denote the choice-probability-weighted average characteristic evaluated at $\mathbb{E}[\nu]$ as $\tilde{w} \equiv \sum_{\ell=0}^{J} \pi_\ell(\mathbb{E}[\nu]) w_\ell$. Then

$$\frac{\mathbb{E}\left[\pi_j(\alpha, \beta) \frac{\pi_k(\alpha, \beta)}{\mathbb{E}[\pi_k(\alpha, \beta)]}\right]}{\mathbb{E}\left[\pi_j(\alpha, \beta)\right]} \approx 1 + (w_j - \tilde{w})' \, \mathbb{V}[\nu] \, (w_k - \tilde{w}),$$

where the neglected terms are of third and higher order in the central moments of ν.

Proposition 6 shows that the first-order impact of the weighting induced by the mixed logit is to add a term that depends on the similarity of the characteristics of goods j and k. Even if α and β are independent, so that $\mathbb{V}[\nu]$ is block diagonal, the differences be-

[10] In particular, $\eta^{\times}$ only depends on characteristics with random coefficients. The same machinery that we used for own-prices can be used to show that $\eta^{\times}$ is well-approximated by a quadratic function that only depends on prices—now the *sum* of prices $p_k + p_j$—if the random coefficients on prices are independent of those on other characteristics.

tween both price and non-price characteristics will contribute to this first-order impact. The implication is that cross-price elasticities in commonly-used mixed logit specifications can be decidedly less stylized than their standard logit counterparts, even while the own-price elasticities continue to exhibit the unattractive properties described in the previous section.

5 The Demand for Toothpaste

In this section, we apply the results of the previous sections to study the demand for toothpaste, using data from a large-scale pricing test conducted by a national US retailer. This unique setting is interesting for our purposes, because it provides us enough statistical power to implement linear IV estimates and flexible mixed logit specifications using variation from randomly-assigned prices.

5.1 The natural field experiment

Our data comes from a large US retailer that operates thousands of brick-and-mortar stores across all fifty states, selling both private-label and national-brand products across food, consumables, and general merchandise. The retailer's economics team ran a pricing test 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 prices were explicitly randomized, "field" in that it took place in the retail environment in which customers make their regular purchasing decisions, and "natural" in that customers were unaware that the price changes were part of an experiment.

The pricing test was run in 4,512 stores across 46 states, grouped into 2,236 clusters. Stores in the same cluster received identical prices. Prices in the test were assigned at the level of a "line": a set of items, such as flavor variants of the same size and brand, that must carry the same price within a store at any point in time. All products in a line were assigned the same price in each store.

For most lines, one-third of the clusters carrying that line were randomly assigned—independently across products—to a low, middle, or high price point. The middle price point was set to the modal pre-test price of the product across stores, while the low and high variants were set symmetrically around it, subject to some regulatory, cost, and other constraints; see Appendix SA.3.5 for more details. A small number of lines were randomized to sets of four or five prices with similar schemes. Randomization was stratified by geographic region and by tercile of pre-period unit sales. Tables SA.5 and SA.6 report balance tests which confirm that the randomization was successful. Table SA.7 shows that compliance with assigned prices was high; for the typical (median) product, 90% of stores had an average price exactly equal to the assigned price.

The pricing test included many types of products. We focus on toothpaste, which was chosen by the retailer's economics team at the outset of the pricing test. Estimating the demand for toothpaste with a discrete choice model is natural because toothpaste has readily-measurable product characteristics and a relatively clear market size. We take a market t to be a single store location and measure demand as aggregated over the course of the pricing test. In Appendix SA.3.5, we describe how we use the retailer's data together with census data and recommendations from the dental literature to turn quantities into market share. We take a product j to be a single UPC. This leads to a total of 190 toothpaste products, of which 167 products spread across 64 lines were included in the pricing test. Table SA.8 reports summary statistics for these products. We use price per ounce for p_{jt} throughout.

5.2 Linear IV estimates

In this section, we report linear IV estimates, which we interpret through the lens of a nonparametric demand model.

The nonparametric model defines $s_{jt}(\boldsymbol{p})$ as the market share for good j that would occur at price vector $\boldsymbol{p}$ in market t, distinct from the actual observed market share s_{jt}. We write $\log(s_{jt}/s_{0t})(\boldsymbol{p}) \equiv \log s_{jt}(\boldsymbol{p}) - \log s_{0t}(\boldsymbol{p})$ to denote the log share for good j relative to the outside option. Let $p_{jt}(z)$ denote the price that product j would have in market t if its line were assigned price z. Then the actual price of product j in market t is $p_{jt} = \sum_z \mathbb{1}[z_{\ell(j)t} = z] p_{jt}(z)$, where $z_{\ell(j)t}$ is the randomly-assigned line price and $\ell(j)$ is the line that product j belongs to. Let $\boldsymbol{p}_{\ell t}(z_\ell^a)$ denote the vector of prices for the products in line ℓ—all of which are the same—if the line were assigned to price z_ℓ^a. There are n_ℓ values $z_\ell^1, \ldots, z_\ell^{n_\ell}$ indexed in increasing order, where $n_\ell = 3$ for most lines, but is four or five for a few others. Let $\boldsymbol{p}_{-\ell t}(z_\ell^a)$ denote the vector of prices for products not in line ℓ if line ℓ were assigned to price z_ℓ^a.

We make the following nonparametric IV assumptions. First, price assignment is excluded from shares, an assumption already implicit in the notation $s_{jt}(\boldsymbol{p})$. Second, price assignment for line ℓ does not affect the prices of other products: $\boldsymbol{p}_{-\ell t}(z_\ell^a) = \boldsymbol{p}_{-\ell t}$. Third, price assignment is exogenous, so that $z_{\ell(j)t}$ is jointly independent of $s_{jt}(\boldsymbol{p})$ for any $\boldsymbol{p}$, $\boldsymbol{p}_{\ell(j)t}(z)$ for any z, and $\boldsymbol{p}_{-\ell(j)t}$. All of these assumptions are standard and supported by the details of the pricing test.

Now consider a product-specific instrumental variables regression of $\log s_{jt} - \log s_{0t}$ onto price p_{jt} and a constant, instrumenting for price with $z_{\ell(j)t}$, using data from a population of markets. Let IV_j denote the population coefficient on p_{jt}. In Appendix

SA.4, we show that

$$\text{IV}_j = \mathcal{E}\left[\sum_{a=2}^{n_{\ell(j)}} \omega_{jt}^a \breve{\eta}_{jt}^a\right], \tag{25}$$

where $\mathcal{E}$ denotes expectation over markets, $\breve{\eta}_{jt}^a$ are arc semi-elasticities, and ω_{jt}^a are weights. The arc semi-elasticities are defined as

$$\breve{\eta}_{jt}^a \equiv \frac{\log(s_{jt}/s_{0t})\left(\boldsymbol{p}_{\ell(j)t}(z_{\ell(j)}^a), \boldsymbol{p}_{-\ell(j)t}\right) - \log(s_{jt}/s_{0t})\left(\boldsymbol{p}_{\ell(j)t}(z_{\ell(j)}^{a-1}), \boldsymbol{p}_{-\ell(j)t}\right)}{p_{jt}(z_{\ell(j)}^a) - p_{jt}(z_{\ell(j)}^{a-1})},$$

which reflect the change in the log share of good j relative to the outside option when the prices in line $\ell(j)$ are shifted, normalized by that change in price—which is the same for all products in the line—while holding the prices of all goods outside of that line fixed.[11] The weights are defined as

$$\omega_{jt}^a = \frac{\mathcal{C}[z_{\ell(j)t}, \mathbb{1}[z_{\ell(j)t} \geq z_{\ell(j)}^a]]\left(\mathcal{E}[p_{jt}|z_{\ell(j)t} = z_{\ell(j)}^a] - \mathcal{E}[p_{jt}|z_{\ell(j)t} = z_{\ell(j)}^{a-1}]\right)}{\sum_{e=2}^{n_{\ell(j)}} \mathcal{C}[z_{\ell(j)t}, \mathbb{1}[z_{\ell(j)t} \geq z_{\ell(j)}^e]]\left(\mathcal{E}[p_{jt}|z_{\ell(j)t} = z_{\ell(j)}^e] - \mathcal{E}[p_{jt}|z_{\ell(j)t} = z_{\ell(j)}^{e-1}]\right)} \times \frac{\left(p_{jt}(z_{\ell(j)}^a) - p_{jt}(z_{\ell(j)}^{a-1})\right)}{\mathcal{E}\left[p_{jt}(z_{\ell(j)}^a) - p_{jt}(z_{\ell(j)}^{a-1})\right]}. \tag{26}$$

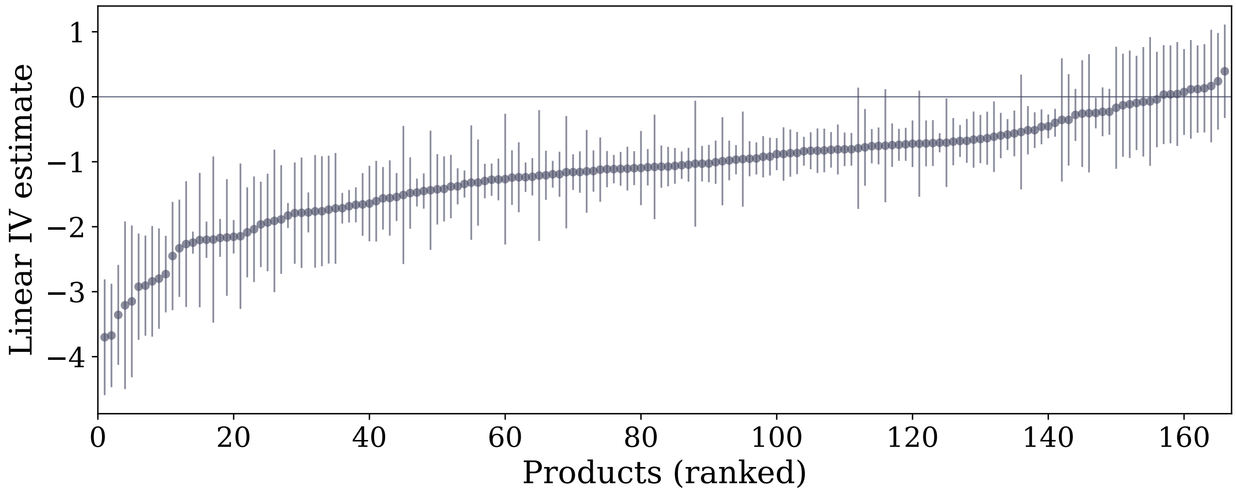
The first component reflects the averaging across multiple instrument contrasts that is performed by the linear IV estimator. The second component reflects non-compliance that occurs when $p_{jt}(z_{\ell(j)}^a)$ is a random variable that is not deterministically equal to the assigned price $z_{\ell(j)}^a$. The weights integrate to one. They are also non-negative under the monotonicity condition that higher assigned prices do not cause lower posted prices, although we do not need this condition for our purposes.

Figure 5 shows estimates of IV_j for each product j, multiplied by the quantity-weighted average price across markets to put them on the scale of an elasticity. The estimates are all larger than -3.5. A few are non-negative, but these estimates are not statistically different from zero. Most of the estimates are quite precise, a consequence of the strength of the randomly assigned price instruments.

While (25) and (26) show that interpreting these estimates of IV_j is difficult in general, they also show how features of our setting help simplify the interpretation. First,

[11] If $p_{jt}(z_{\ell(j)}^a) = p_{jt}(z_{\ell(j)}^{a-1})$, so that the denominator of $\breve{\eta}_{jt}^a$ is zero, then the weights ω_{jt}^a are also zero. This just reflects a multiply-and-divide step that we keep implicit in the notation for simplicity—see Appendix SA.4 for details.

Figure 5: Per-product linear IV estimates



Notes: The figure shows per-product linear IV estimates multiplied by the quantity-weighted average price across markets. Products are ordered by their point estimates, which are shown with 95% confidence intervals clustered at the store cluster level. We only include 166 *of the* 167 *products that were in the pricing test. The excluded product is an imprecisely-estimated outlier with a point estimate of 41.5 and a standard error of 30.9.*

compliance with the assigned prices was high (Table SA.7), which makes the second component of ω_{jt}^{a} approximately one. Second, the assigned prices were approximately evenly spaced with equal probability, making the first component of ω_{jt}^{a} approximately equal across a. Third, the magnitude difference between the assigned prices was relatively small, which would mean these arc semi-elasticities are approximately point semi-elasticities. Taken together, these features suggest that the estimates of IV_j in Figure 5 are approximately estimates of cross-market average elasticities of demand, evaluated at each product's typical price level. This approximate interpretation doesn't play a role in what follows, but it suggests that IV_j can be thought of as a coarse cross-market summary of the type of elasticity a researcher might use a logit demand model to estimate.

5.3 Mixed logit estimates

We also estimated five logit demand models for toothpaste. All specifications contain product indicators, but differ in how they incorporate random coefficients and what instruments they use. We implement all estimates using the `pyblp` best practices (Conlon and Gortmaker, 2020) with multiple starting values for the optimization.

Specification 1 is a standard logit. Specification 2 puts a random coefficient only on price. Specification 3 puts independent random coefficients on five additional non-price

Table 5: Parameter estimates for toothpaste mixed logit demand models

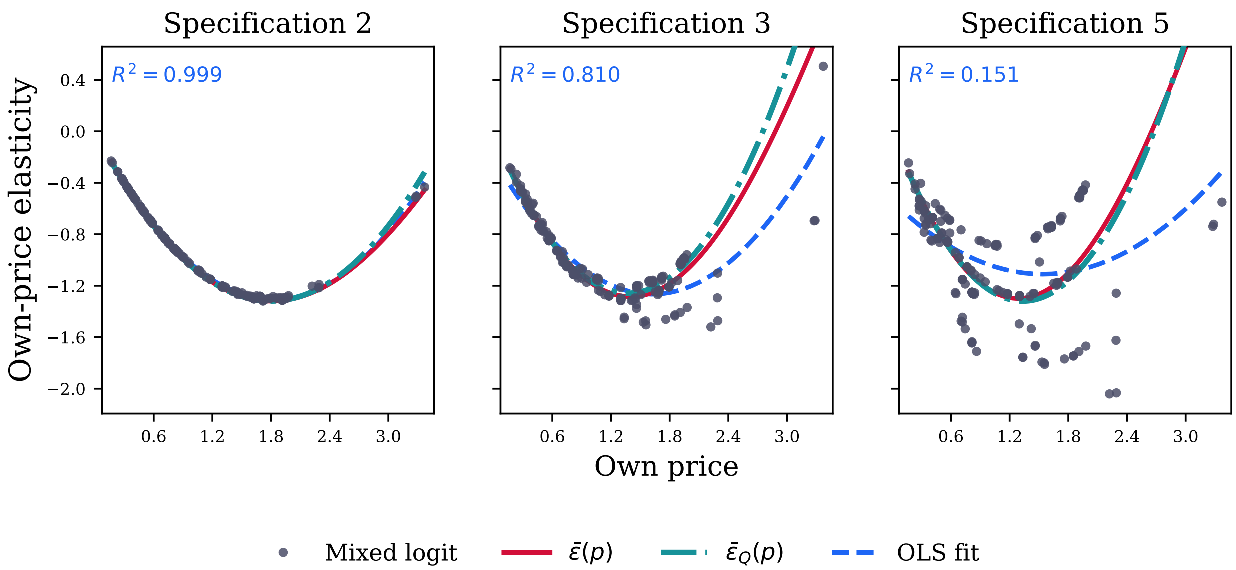
Random coefficients	(1) None	(2) Price only	(3) Independent	(4) Independent	(5) Corr. w/ price
Mean coefficient					
Price	−0.65 (0.02)	−1.37 (0.05)	−1.69 (0.11)	−1.85 (0.12)	−1.57 (0.08)
Std. dev. of random coefficients					
Price	—	0.67 (0.02)	0.89 (0.06)	1.00 (0.07)	0.92 (0.04)
Brand 1	—	—	1.48 (0.72)	1.79 (0.78)	0.44 (3.85)
Brand 2	—	—	3.11 (1.39)	0.00	0.66 (1.46)
Peroxide	—	—	3.07 (1.51)	4.93 (1.38)	0.21 (0.08)
Gel	—	—	1.38 (2.41)	0.08 (15.40)	2.49 (2.11)
Travel size	—	—	2.27 (0.69)	0.27 (4.80)	0.51 (10.59)
Income interaction (× income)					
Price	—	—	—	0.07 (0.07)	—
Brand 1	—	—	—	−0.26 (0.11)	—
Brand 2	—	—	—	−1.02 (0.38)	—
Peroxide	—	—	—	1.17 (0.29)	—
Gel	—	—	—	2.88 (0.71)	—
Travel size	—	—	—	−7.70 (5.23)	—
Covariance with price coefficient					
Brand 1	—	—	—	—	−0.38 (0.09)
Brand 2	—	—	—	—	−0.61 (0.09)
Peroxide	—	—	—	—	−0.20 (0.08)
Gel	—	—	—	—	−0.76 (0.40)
Travel size	—	—	—	—	−0.32 (0.08)
Product indicators	✓	✓	✓	✓	✓
Instruments					
Randomized line prices	✓	✓	✓	✓	✓
Differentiation instruments			✓	✓	✓
Income × characteristics				✓	

Notes: Standard errors in parentheses. All specifications are estimated using `pyblp` *(Conlon and Gortmaker, 2020). In Specification (4), the Brand 2 standard deviation is estimated to be zero.*

binary characteristics: two brand indicators for the two largest toothpaste brands, an indicator for hydrogen peroxide, an indicator for whether the product is gel-based, and an indicator for whether the product is travel-sized. Specification 4 adds income interactions for all characteristics. Specification 5 drops income interactions and estimates correlations between the price random coefficient and the random coefficients on all other non-price characteristics. We were not able to successfully estimate a specification that had both income interactions and unrestricted correlations between random coefficients or a specification that had a completely unrestricted correlation structure.

Specifications 1 and 2 are estimated using the same assigned-line-price instruments that we used to construct the linear IV estimates in Section 5.2. In Specifications 3,

Figure 6: Own-price elasticities for toothpaste



Notes: The figure shows the cross-market average elasticities, price-only approximations, quadratic price-only approximations, and OLS quadratic fit for Specifications 2, 3, and 5. Specification 2 only has a random price coefficient, Specification 3 has independent random coefficients on all product characteristics, and Specification 5 allows for correlated random coefficients between price and other non-price characteristics; see Table 5 and the main text for more detail. The R-squared figures are for the quadratic OLS fits. Averages are weighted by quantity sold, so that markets that don't carry the product receive zero weight.

4, and 5, we also add the Gandhi and Houde (2019) differentiation instruments for the five non-price characteristics. In Specification 4, we additionally interact household income with price and the product characteristics, as in Miller and Weinberg (2017).

Table 5 reports the parameter estimates from the five models. All models have tightly estimated negative price coefficients. All specifications with random coefficients have average price coefficients that are two to three times as negative as the standard logit. The standard deviations of the price coefficient in these specifications are equal or greater in magnitude than the single logit estimate. Specification 3 adds random coefficients on non-price characteristics, the variances of which we estimate to be large and significant, with the exception of the gel indicator. Adding interactions with income in Specification 4 changes the estimated variances considerably, driving the Brand 2 variance to zero, and produces significant income interactions for several of the non-price characteristics. However, the interaction between income and price is small and statistically insignificant. Allowing for correlated random coefficients in Specification 5 also substantially changes the estimated variances of the non-price characteristics, while introducing non-trivial and significant covariance estimates.

Figure 6 plots the own-price elasticities that come out of Specifications 2, 3, and 5, averaged across markets. The elasticities follow a strong upward-opening quadratic shape for Specification 2, which only has a random coefficient on price. The independent random coefficients in Specification 3 loosen the grip of the quadratic fit slightly, but not by much, especially given the large magnitude of the non-price random coefficient variances. These two panels are strongly concordant with the theory developed in Section 3. Specification 4 (not shown) turns out to be similar to Specification 3 because the income interaction with price is estimated to be small, leading to little correlation between the price and non-price random coefficients. Specification 5 breaks the quadratic shape by allowing for correlated random coefficients. Because the characteristics are binary, one can still make out the tendency towards a quadratic structure created by certain groups of similar products, consistent with Proposition 4. However, the plot shows that there is no longer a strong functional relationship between own-price elasticities and own prices alone.

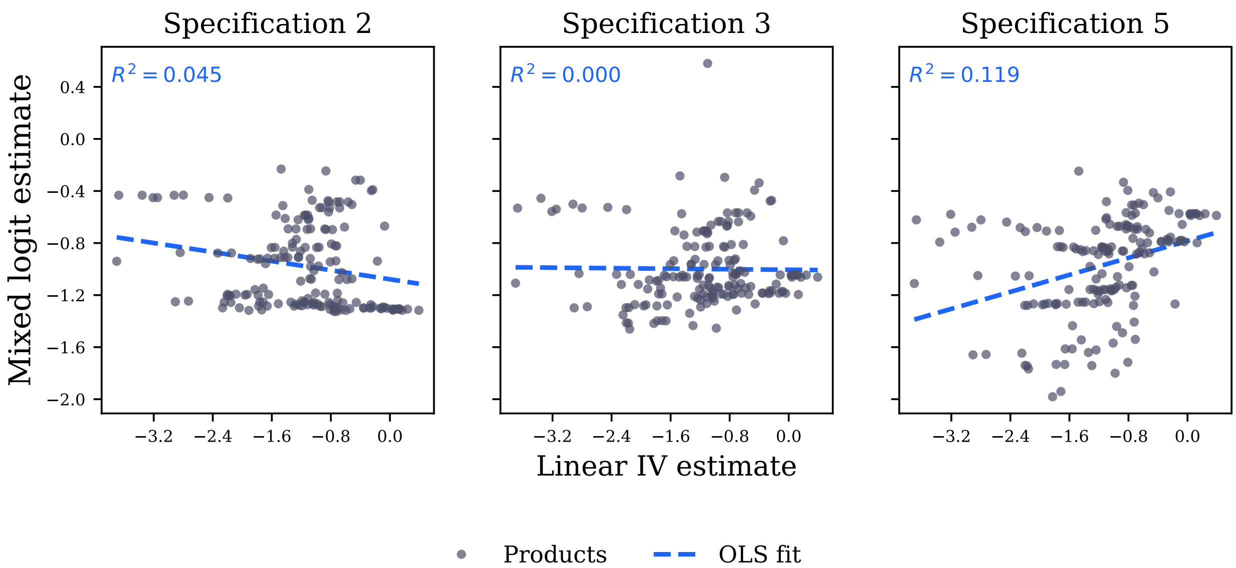
5.4 Comparing the linear IV and mixed logit estimates

In (25) we showed that the linear IV estimand can be interpreted as a weighted average of cross-market arc semi-elasticities. This interpretation rests on standard exclusion and exogeneity conditions; in particular, it does not require any parametric assumptions. In a setting with full compliance between assigned and posted prices, the mixed logit model should be able to reproduce the linear IV estimand if it is correctly specified.

Compliance is high in our setting, but not full: for the typical product, 90% of stores posted exactly the assigned price, and most deviations were within a few percent of it (Table SA.7). The structure of the linear IV estimand in (25) reveals two ways in which non-compliance can affect its interpretation. First, the second component of the weights (26) is random, so may not be equal to one. Second, the arc semi-elasticities $\breve{\eta}^a_{jt}$ for each instrument contrast are between posted prices rather than assigned prices.

These effects should be small in our setting, especially given that compliance is so high. If the posted prices that a store would set under each assignment (the collection of random variables $p_{jt}(z)$) are independent of the store's demand conditions, then the weights are independent of the arc semi-elasticities, so that the first source of non-compliance has minimal impact. This independence is plausible in our setting because non-compliance appears to have been caused primarily by idiosyncratic managerial errors, with pre-planned sales or markdowns forming a small share of non-compliance. Non-compliance still creates a potential deviation in how the arc semi-elasticities differ when comparing assigned and posted prices. However, because non-compliance

Figure 7: Mixed logit estimates of the linear IV estimand



Notes: Each point is a toothpaste product, plotting its linear IV estimate from Figure 5 against the mixed logit estimate of the IV estimand, (25). The dashed line is an OLS fit of the mixed logit estimate on the linear IV estimate, with its R-squared shown. We exclude one of the 167 products that has an outlier linear IV point estimate of 41.5 with a standard error of 30.9.

deviations are typically an order of magnitude smaller than the differences between adjacent assigned prices, it seems unlikely that deviations will lead to large differences between average assigned-price arc semi-elasticities and average posted-price arc semi-elasticities. Non-compliance occurred for both high and low assigned prices, so the impact of these deviations may also average out to zero across stores.

Figure 7 plots linear IV estimates against the corresponding mixed logit estimates of the linear IV estimand. The logic just outlined suggests that the two estimates should be similar if the mixed logit model is correctly specified. Instead, we find that the mixed logit estimates are only weakly correlated with the linear IV estimates. In particular, the R-squared from Specification 3, which only has independent random coefficients, is essentially zero. Adding correlated random coefficients in Specification 5 improves the correlation somewhat, leading to an R-squared of approximately 0.12.

In Appendix SA.5, we develop formal statistical tests of the null hypotheses that the linear IV estimands are equal to their mixed logit counterparts, accounting for sampling variation in both the linear IV estimates and the estimated mixed logit models. The product-specific tests reject at the 5% level for 31% of the products in Specification 5, and for a larger share of products in the other specifications. A joint test of equality across all products rejects at any conventional significance level for all specifications (Table SA.9). Restricting the test to products with higher compliance actually *increases* the proportion of products for which equality is rejected, suggesting that the non-compliance caveat discussed above is not particularly relevant.

6 Conclusion

In this paper we provided a new analysis of the elasticities produced by logit and mixed logit discrete choice models of demand. Our analysis shows both theoretically and empirically that the own-price elasticities produced by these models can be stylized. When the random coefficients on price and non-price characteristics are independent, the own-price elasticities tend to be well-approximated by quadratic functions of own price. This is not that different from the standard logit model, which tends to produce own-price elasticities that are well-approximated by linear functions of own price. These properties imply that downstream economic quantities, such as markups, pass-through, and diversion ratios, will also be stylized and driven primarily by modeling assumptions.

Our results provide several takeaways for practitioners.

First, it is important to keep in mind the essential role of the small-π approximation in Proposition 1. In settings with large market shares—or large and highly-variable type-specific choice probabilities—the small-π approximation error ι_j^{SCP} in Proposition 1 might be substantial. Our analysis of the η function in Proposition 1 is only relevant for describing own-price elasticities for settings when the small-π approximation error is small. This may be particularly relevant in settings with a small number of goods.

Second, a recurring theme of our findings is that in order to produce flexible own-price elasticities it is important to allow for correlation between price and non-price random coefficients. The most direct way to do this is by estimating off-diagonal terms of the variance-covariance matrix of random coefficients. Our application to toothpaste demand shows that this is feasible—albeit in a data-rich environment—and produces own-price elasticities that are less stylized. Correlation is often an implicit feature of proposals for flexibly-parameterized mixed logit models, as in Fox et al. (2011), Fox and Gandhi (2016), Train (2016), and Nevo et al. (2026). A simpler way to induce correlation is to include demographic interactions with price and non-price characteristics that either use the same demographics or use different demographics drawn from the joint distribution of demographics. This motivation for including demographics does not seem to have been proposed before in the literature.

Third, all of the approximations that we provide and errors that we quantify involve objects that are easy to compute from data and parameter values. Researchers can take their fitted model and conduct the same numerical analysis that we have done to assess the extent to which the elasticities in their model are (or are not) driven by the functional form restrictions imposed by the mixed logit model.

References

Angrist, J. D. and G. W. Imbens (1995): "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity," *Journal of the American Statistical Association*, 90, 431–442.

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

——— (1999): "Voluntary Export Restraints on Automobiles: Evaluating a Trade Policy," *The American Economic Review*, 89, 400–430.

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

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

Björnerstedt, J. and F. Verboven (2016): "Does Merger Simulation Work? Evidence from the Swedish Analgesics Market," *American Economic Journal: Applied Economics*, 8, 125–164.

Bourreau, M., Y. Sun, and F. Verboven (2021): "Market Entry, Fighting Brands, and Tacit Collusion: Evidence from the French Mobile Telecommunications Market," *American Economic Review*, 111, 3459–3499.

Brascamp, H. J. and E. H. Lieb (1976): "On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation," *Journal of Functional Analysis*, 22, 366–389.

Brown, L. D. (1986): *Fundamentals of statistical exponential families with applications in statistical decision theory*, SPIE.

Cacoullos, T. and V. Papathanasiou (1989): "Characterizations of distributions by variance bounds," *Statistics & Probability Letters*, 7, 351–356.

Compiani, G. (2022): "Market counterfactuals and the specification of multiproduct demand: A nonparametric approach," *Quantitative Economics*, 13, 545–591.

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

Crawford, G. S., O. Shcherbakov, and M. Shum (2019): "Quality Overprovision in Cable Television Markets," *American Economic Review*, 109, 956–995.

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

D'Haultfœuille, X., I. Durrmeyer, and P. Février (2019): "Automobile Prices in Market Equilibrium with Unobserved Price Discrimination," *The Review of Economic Studies*, 86, 1973–1998.

Fabinger, M. and E. G. Weyl (2018): "Functional Forms for Tractable Economic Models and the Cost Structure of International Trade," Tech. Rep. CIRJE-F-1092, CIRJE, Faculty of Economics, University of Tokyo.

Fox, J. T. and A. Gandhi (2016): "Nonparametric identification and estimation of random coefficients in multinomial choice models," *The RAND Journal of Economics*, 47, 118–139.

Fox, J. T., K. i. Kim, S. P. Ryan, and P. Bajari (2011): "A simple estimator for the distribution of random coefficients," *Quantitative Economics*, 2, 381–418.

——— (2012): "The random coefficients logit model is identified," *Journal of Econometrics*, 166, 204–212.

Gandhi, A. and J.-F. Houde (2019): "Measuring Substitution Patterns in Differentiated-Products Industries," Working paper.

Gandhi, A. and A. Nevo (2021): "Empirical models of demand and supply in differentiated products industries," in *Handbook of Industrial Organization*, Elsevier, vol. 4, 63–139.

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

Hausman, J. A. and D. A. Wise (1978): "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," *Econometrica*, 46, 403–426.

Hortaçsu, A. and J. Joo (2023): *Structural econometric modeling in industrial organization and quantitative marketing: theory and applications*, Princeton University Press.

Imbens, G. W. and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475.

Li, S. (2018): "Better Lucky Than Rich? Welfare Analysis of Automobile Licence Allocations in Beijing and Shanghai," *The Review of Economic Studies*, 85, 2389–2428.

Miller, N. H. and M. C. Weinberg (2017): "Understanding the Price Effects of the MillerCoors Joint Venture," *Econometrica*, 85, 1763–1791.

Miravete, E. J., K. Seim, and J. Thurk (2025): "Elasticity and Curvature of Discrete Choice Demand Models," Working paper.

Mogstad, M. and A. Torgovitsky (2024): "Instrumental variables with unobserved heterogeneity in treatment effects," in *Handbook of Labor Economics*, Elsevier, vol. 5, 1–114.

Nevo, A. (2000): "A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand," *Journal of Economics & Management Strategy*, 9, 513–548.

——— (2001): "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, 69, 307–342.

Nevo, A., J. Tao, and A. Gandhi (2026): "Flexible Estimation of Differentiated Product Demand Models Using Aggregate Data," .

Prékopa, A. (1973): "On logarithmic concave measures and functions," *Acta Scientiarum Mathematicarum (Szeged)*, 34, 335–343.

Tebaldi, P., A. Torgovitsky, and H. Yang (2023): "Nonparametric Estimates of Demand in the California Health Insurance Exchange," *Econometrica*, 91, 107–146.

Train, K. (2016): "Mixed logit with a flexible mixing distribution," *Journal of Choice Modelling*, 19, 40–53.

Train, K. E. (2009): *Discrete choice methods with simulation*, Cambridge university press.

Zhang, L. (2026): "Identification and Estimation of Market Size in Discrete Choice Demand Models," *Working paper*.

Supplemental Appendix

Contents

SA.1 Survey of Logit Demand Estimation

This section provides some additional details on how we constructed Table 1.

We began with the sample of 29 papers collected by Zhang (2026), who graciously shared their collection with us. We excluded four papers, three of which did not include price as a characteristic, and one that used a sufficiently non-standard two-stage model. When classifying specifications, we focus only on the random coefficients, ignoring distinctions about whether or what type of nests were used. The "Any use" column includes any result reported in the main text and appendices.

We defined a random coefficient as being a coefficient that is nondegenerate within a market. In the classic example of a mixed logit, this is because of unobserved heterogeneity for α and/or β. However, this definition also encompasses heterogeneity introduced by including draws of demographics, of which income is the most common example. We determined a market by the level t at which the important ξ_{jt} residual in (1) varies. Some papers use random coefficients on a constant term, which creates heterogeneity in utility for all inside goods versus the outside option. Our results in Proposition 1 show that such terms do not contribute to variation in own-price elasticities across goods, so we do not treat these cases as non-price random coefficients; the only characteristics we count coefficients on are those that vary across inside goods.

SA.2 Proofs for Sections 3 and 4

Section SA.2.1 defines some notation used in this appendix. Section SA.2.2 contains intermediate results related to exponential tilting that are used in our approximations. The remaining subsections then contain proofs of the results in the main text.

SA.2.1 Notation

The derivations in this appendix are centered around product characteristics with random coefficients, which we sometimes group together as $w_j = (p_j, r_j)$ with coefficients $\nu \equiv (\alpha, \beta)$ for brevity. We write the type-specific choice probability equation (2) as

$$\pi_j(\nu) \equiv \frac{\exp(w_j'\nu + \delta_j)}{\sum_{k=0}^{J} \exp(w_k'\nu + \delta_k)}, \quad \text{where} \quad \delta_j \equiv \gamma' c_j + \xi_j,$$

and we have continued to suppress the individual subscript i, which is captured through the random vector ν, as well as the market subscript t, because all of our analysis in this appendix works within a single market. The new variable δ_j is the sum of the characteristics with non-random coefficients, both observed $\gamma' c_j$ and unobserved ξ_j.

All of the random coefficient contributions are contained in $w_j'\nu$. Market-level shares are (as before) given by $s_j \equiv \mathbb{E}[\pi_j(\nu)] \equiv \mathbb{E}[\pi_j(\alpha, \beta)]$, with the expectation taken over $\nu \equiv (\alpha, \beta)$.

SA.2.2 The mixed logit connection to exponential tilting

A common object in the mixed logit model is a probability-weighted expectation like

$$\mathbb{E}\left[g(\nu)\frac{\pi_j(\nu)}{\mathbb{E}[\pi_j(\nu)]}\right] = \int g(v)\frac{\pi_j(v)}{\mathbb{E}[\pi_j(\nu)]} f(v)\ dv$$

for some function g, where $f(v)$ is the density of ν. Our analysis uses the observation that weighting $f(v)$ by $\pi_j(v)$ produces a member of an exponential family of densities defined as

$$h_\omega(v) \equiv h(v)\exp(\omega' v - A(\omega)), \tag{27}$$

where $h \equiv h_0$ and A are given by

$$\begin{aligned} h(v) &\equiv \frac{f(v)}{\sum_{j=0}^J \exp(w_j' v + \delta_j)} = f(v)\pi_0(v) \\ \text{and} \quad A(\omega) &\equiv \log \int h(v)\exp(\omega' v)\, dv = \log \mathbb{E}[\pi_0(\nu)\exp(\omega'\nu)]. \end{aligned} \tag{28}$$

The densities $h_\omega(v)$ form a proper exponential family in the natural parameter ω, with canonical sufficient statistic v, base measure h, and log partition function $A(\omega)$. Note that both h and A are defined in terms of all product and market characteristics $(w_1, \ldots, w_J, \delta_1, \ldots, \delta_J)$, which are held fixed as we vary the natural parameter ω. The log partition function is well-defined at any point in the convex hull of product characteristics: $\pi_0(v)\exp(w_j'v) = \exp(w_j'v)/\sum_{k=0}^J \exp(w_k'v + \delta_k) \leq \exp(-\delta_j)$ is bounded for all v and therefore also for any ω formed from a convex combination of $\{w_j\}_{j=1}^J$, due to the convexity of the exponential function.

The next lemma shows that weighting an expectation by the choice probability of good j is the same as integrating against h_{w_j}.

Lemma PWT. (Probability-weighting is the same as tilting) Let h_ω be the exponential family of densities defined in (27). Then

$$f(v)\frac{\pi_j(v)}{\mathbb{E}[\pi_j(\nu)]} = h_{w_j}(v) \quad \text{for all } v.$$

Proof of Lemma PWT. The equality follows from

$$\begin{aligned}
&f(v)\left(\frac{\pi_j(v)}{\mathbb{E}[\pi_j(\nu)]}\right) \\
&\quad= f(v)\left(\frac{\exp(w_j'v+\delta_j)}{\sum_{k=0}^J \exp(w_k'v+\delta_k)}\right)\left(\mathbb{E}\left[\frac{\exp(w_j'\nu+\delta_j)}{\sum_{k=0}^J \exp(w_k'\nu+\delta_k)}\right]\right)^{-1} \\
&\quad= f(v)\left(\frac{\exp(w_j'v)}{\sum_{k=0}^J \exp(w_k'v+\delta_k)}\right)\left(\mathbb{E}\left[\frac{\exp(w_j'\nu)}{\sum_{k=0}^J \exp(w_k'\nu+\delta_k)}\right]\right)^{-1} \\
&\quad= f(v)\left(\frac{\exp(w_j'v)}{\sum_{k=0}^J \exp(w_k'v+\delta_k)}\right)\left(\int \frac{\exp(w_j'\tilde{v})}{\sum_{k=0}^J \exp(w_k'\tilde{v}+\delta_k)} f(\tilde{v})\,d\tilde{v}\right)^{-1} \\
&\quad\equiv h(v)\exp(w_j'v)\left(\int h(\tilde{v})\exp(w_j'\tilde{v})d\tilde{v}\right)^{-1} \\
&\quad\equiv h(v)\exp(w_j'v)\exp(-A(w_j)) \equiv h_{w_j}(v).
\end{aligned}$$

Q.E.D.

Expectations for exponential families are described by the derivative of the log partition function, $A(\omega)$. Further derivatives of the log partition function characterize the higher order cumulants associated with h_ω. The following lemma records these properties, which are well-known. For this lemma, and throughout this appendix, we use the notation $\mathbb{E}^h_\omega$, $\mathbb{V}^h_\omega$, $\mathbb{C}^h_\omega$, and $\mathbb{K}^h_\omega$ to denote operators when ν has density h_ω.

Lemma EFC. (Exponential family cumulants) Let $\mathbb{E}^h_\omega$, $\mathbb{V}^h_\omega$, $\mathbb{C}^h_\omega$, and $\mathbb{K}^h_\omega$ denote expectation, variance, covariance, and third cumulant operators when ν has density h_ω. Then

$$\begin{aligned}
\mathbb{E}^h_\omega[\nu_\ell] &= \frac{\partial A(\omega)}{\partial \omega_\ell} \\
\mathbb{C}^h_\omega[\nu_\ell,\nu_m] &\equiv \mathbb{E}^h_\omega\left[(\nu_\ell-\mathbb{E}^h_\omega[\nu_\ell])(\nu_m-\mathbb{E}^h_\omega[\nu_m])\right] = \frac{\partial^2 A(\omega)}{\partial\omega_\ell\partial\omega_m} \qquad (29) \\
\text{and} \quad \mathbb{K}^h_\omega[\nu_\ell,\nu_m,\nu_n] &\equiv \mathbb{E}^h_\omega[(\nu_\ell-\mathbb{E}^h_\omega[\nu_\ell])(\nu_m-\mathbb{E}^h_\omega[\nu_m])(\nu_n-\mathbb{E}^h_\omega[\nu_n])] = \frac{\partial^3 A(\omega)}{\partial\omega_\ell\partial\omega_m\partial\omega_n}.
\end{aligned}$$

Proof of Lemma EFC. These properties are well-known. See, for example, Brown (1986, Chapter 2). Q.E.D.

SA.2.3 Proof of Proposition 1

Proof of Proposition 1. The semi-elasticity with respect to the ℓth characteristic can be found by differentiating and rearranging:

$$\begin{aligned}\frac{\partial s_j}{\partial w_{j\ell}}\frac{1}{s_j} &= \frac{\mathbb{E}[\nu_\ell \pi_j(\nu)(1-\pi_j(\nu))]}{\mathbb{E}[\pi_j(\nu)]} \\ &= \mathbb{E}\left[\nu_\ell \frac{\pi_j(\nu)}{\mathbb{E}[\pi_j(\nu)]}\right] - \mathbb{E}\left[\nu_\ell \pi_j(\nu)\frac{\pi_j(\nu)}{\mathbb{E}[\pi_j(\nu)]}\right] \\ &= \int v_\ell h_{w_j}(v)\,dv - \int v_\ell \pi_j(v) h_{w_j}(v)\,dv \equiv \mathbb{E}^h_{w_j}[\nu_\ell] - \mathbb{E}^h_{w_j}[\nu_\ell \pi_j(\nu)],\end{aligned}$$

where the second-to-last equality used Lemma PWT with the exponential tilting connection discussed in Section SA.2.2. Setting $\nu_\ell = \alpha$, we let

$$\eta(w_j) \equiv \eta(p_j, r_j) \equiv \mathbb{E}^h_{w_j}[\alpha] \quad \text{and} \quad \iota_j^{\text{SCP}} \equiv -\,\mathbb{E}^h_{w_j}[\alpha \pi_j(\alpha, \beta)].$$

Lemma SMS derives the stated bound on $|\iota_j^{\text{SCP}}|$. Q.E.D.

Lemma SMS. (Small-π approximations) Suppose that $\mathbb{E}[\nu_\ell^2]$ is finite. Then

$$\left|\mathbb{E}^h_{w_j}[\nu_\ell \pi_j(\nu)]\right| \le \mathbb{E}[\nu_\ell^2]^{1/2}\left(\frac{\mathbb{E}[\pi_j(\nu)^4]}{\mathbb{E}[\pi_j(\nu)]^2}\right)^{1/2}. \tag{30}$$

Proof of Lemma SMS. The inequality follows from

$$\begin{aligned}\left|\mathbb{E}^h_{w_j}[\nu_\ell \pi_j(\nu)]\right| &= \left|\frac{1}{s_j}\,\mathbb{E}\left[\nu_\ell \pi_j(\nu)^2\right]\right| \\ &\le \frac{1}{s_j}\,\mathbb{E}\left[\nu_\ell^2\right]^{1/2}\mathbb{E}\left[\pi_j(\nu)^4\right]^{1/2} = \mathbb{E}[\nu_\ell^2]^{1/2}\left(\frac{\mathbb{E}[\pi_j(\nu)^4]}{\mathbb{E}[\pi_j(\nu)]^2}\right)^{1/2},\end{aligned}$$

where the first equality used Lemma PWT, the inequality is Cauchy-Schwarz, and the final equality just rewrites terms. Q.E.D.

SA.2.4 Proof of Proposition 2

Proof of Proposition 2. Let $r^\star$ denote a fixed vector that may depend on the market but not the good j. (The statement of the proposition takes $r^\star = \bar{r}$ as a special case.) Define $r_j(\tau) \equiv r^\star + \tau(r_j - r^\star)$ for $\tau \in [0,1]$. Then the fundamental theorem of calculus gives

$$\eta(p_j, r_j) - \eta(p_j, r^\star) = \int_0^1 \frac{d}{d\tau}\eta(p_j, r_j(\tau))\,d\tau = \int_0^1 (r_j - r^\star)'\nabla_r \eta(p_j, r_j(\tau))\,d\tau, \tag{31}$$

where $\nabla_r \eta(p_j, r_j(\tau))$ is the vector of derivatives of η with respect to the components of r_j, all evaluated at $(p_j, r_j(\tau))$. The result follows by applying Lemma EFC, which shows that

$$\frac{\partial}{\partial r_\ell}\eta(p,r) = \frac{\partial}{\partial r_\ell}\left(\frac{\partial}{\partial p}A(p,r)\right) = \mathbb{C}^h_{p,r}[\alpha, \beta_\ell],$$

where A is the log partition function in the exponential family (27). Q.E.D.

SA.2.5 Proof of Proposition 3

Our proof of Proposition 3 uses several intermediate arguments, which we include as Lemmas CMD, SKG, and BLI at the end of this subsection.

Proof of Proposition 3. We start by deriving an alternative expression for the tilted covariance $\mathbb{C}^h_w[\alpha, \beta]$ that appears in the integrand of ι_j^{DEP} with $w = (p_j, r_j(\tau))$. Iterating expectations on α gives

$$\mathbb{C}^h_w[\alpha, \beta] = \mathbb{E}^h_w\left[\left(\alpha - \mathbb{E}^h_w[\alpha]\right)\mathbb{E}^h_w[\beta|\alpha]\right] \equiv \mathbb{C}^h_w\left[\alpha, \mu_w(\alpha)\right], \tag{32}$$

where $\mu_w(a) \equiv \mathbb{E}^h_w[\beta|\alpha = a]$ is a vector-valued function of a. In Lemma CMD, we use independence between α and β together with properties of the logit to show that the derivative of each component of $\mu_w(a)$ satisfies

$$\mu_w'(a) \equiv \frac{d}{da}\mu_w(a) = -\,\mathbb{C}^h_w[\beta, \bar{p}(\alpha,\beta)|\alpha = a] \quad \text{where} \quad \bar{p}(a,b) \equiv \sum_{j=0}^{J}\pi_j(a,b)p_j. \tag{33}$$

In Lemma SKG, we record a generalization of Stein's identity that when applied to (32) with (33) gives

$$\mathbb{C}^h_w\left[\alpha, \mu_w(\alpha)\right] = \mathbb{E}^h_w\left[\mathcal{K}^h_w(\alpha)\mu_w'(\alpha)\right] = -\,\mathbb{E}^h_w\left[\mathcal{K}^h_w(\alpha)\,\mathbb{C}^h_w[\beta, \bar{p}(\alpha,\beta)|\alpha]\right],$$

where $\mathcal{K}^h_w(\alpha)$ is the Stein kernel of α when α is distributed like h_w (defined in Lemma SKG); we omit the routine verification of the boundary conditions required by Lemma SKG. Returning to the definition of $\iota_j^{\text{DEP}}(r^\star)$ with $w = (p_j, r_j(\tau))$, we have

$$\iota_j^{\text{DEP}}(r^\star) = -\int_0^1 \mathbb{E}^h_{p_j, r_j(\tau)}\left[\mathcal{K}^h_{p_j,r_j(\tau)}(\alpha)\,\mathbb{C}^h_{p_j,r_j(\tau)}\left[(r_j - r^\star)'\beta,\, \bar{p}(\alpha,\beta)|\alpha\right]\right] d\tau. \tag{34}$$

To derive bounds on $\iota_j^{\text{DEP}}(r^\star)$, we start by applying the Cauchy-Schwarz inequality to the covariance term in (34):

$$\left|\mathbb{C}^h_{p_j,r_j(\tau)}[(r_j - r^\star)'\beta, \bar{p}(\alpha,\beta)|\alpha]\right| \\ \leq \left(\mathbb{V}^h_{p_j,r_j(\tau)}\left[(r_j - r^\star)'\beta|\alpha\right]\right)^{1/2}\left(\mathbb{V}^h_{p_j,r_j(\tau)}\left[\bar{p}(\alpha,\beta)|\alpha\right]\right)^{1/2}.$$

We use the Brascamp-Lieb inequality (Lemma BLI) to bound both objects. The inequality relies on the maintained assumption that β is normally distributed. Applying Lemma BLI to both variances gives

$$\mathbb{V}^h_{p_j,r_j(\tau)}\left[(r_j - r^\star)'\beta|\alpha\right] \leq (r_j - r^\star)'\,\mathbb{V}[\beta](r_j - r^\star) \\ \text{and} \quad \mathbb{V}^h_{p_j,r_j(\tau)}\left[\bar{p}(\alpha,\beta)|\alpha\right] \leq \mathbb{E}\left[\nabla_b\bar{p}(\alpha,\beta)'\,\mathbb{V}[\beta]\nabla_b\bar{p}(\alpha,\beta)\middle|\alpha\right], \tag{35}$$

where $\nabla_b\bar{p}(a,b)$ is the gradient of probability-weighted price with respect to b. This gradient has a nice interpretation that depends on logit properties:

$$\begin{aligned}\frac{\partial}{\partial b_\ell}\bar{p}(a,b) &= \sum_{j=0}^{J} p_j \frac{\partial}{\partial b_\ell}\pi_j(a,b) \\ &= \sum_{j=0}^{J} p_j\left(\pi_j(a,b)\left[r_{\ell j} - \sum_{k=0}^{J}\pi_k(a,b)r_{\ell k}\right]\right) \\ &= \left(\sum_{j=0}^{J}\pi_j(a,b)p_j r_{\ell j}\right) - \left(\sum_{j=0}^{J}\pi_j(a,b)p_j\right)\left(\sum_{j=0}^{J}\pi_j(a,b)r_{\ell j}\right) \equiv \mathfrak{c}_\ell(a,b).\end{aligned}$$

We write the gradient stacked over ℓ as

$$\nabla_b\bar{p}(a,b) \equiv \begin{bmatrix}\mathfrak{c}_1(a,b)\\ \vdots \\ \mathfrak{c}_L(a,b)\end{bmatrix} \equiv \mathfrak{c}(a,b). \tag{36}$$

Substituting (36) into (35) and combining with the Cauchy-Schwarz inequality above bounds the inner conditional covariance uniformly in α,

$$\left|\mathbb{C}^h_{p_j,r_j(\tau)}[(r_j - r^\star)'\beta, \bar{p}(\alpha,\beta)|\alpha]\right| \\ \leq \left((r_j - r^\star)'\,\mathbb{V}[\beta](r_j - r^\star)\right)^{1/2}\sup_{(a,b)}\left(\mathfrak{c}(a,b)'\,\mathbb{V}[\beta]\mathfrak{c}(a,b)\right)^{1/2}. \tag{37}$$

Because $\mathcal{K}^h_{p_j,r_j(\tau)}(\alpha) \geq 0$ (Lemma SKG) and $\mathbb{E}^h_{p_j,r_j(\tau)}[\mathcal{K}^h_{p_j,r_j(\tau)}(\alpha)] = \mathbb{V}^h_{p_j,r_j(\tau)}[\alpha]$, applying (37) inside (34) gives

$$\begin{aligned}|\iota_j^{\text{DEP}}(r^\star)| &\le \int_0^1 \mathbb{E}^h_{p_j,r_j(\tau)}\left[\mathcal{K}^h_{p_j,r_j(\tau)}(\alpha)\left|\mathbb{C}^h_{p_j,r_j(\tau)}[(r_j-r^\star)'\beta,\bar{p}(\alpha,\beta)|\alpha]\right|\right]\,d\tau\\ &\le \left(\int_0^1 \mathbb{V}^h_{p_j,r_j(\tau)}[\alpha]\,d\tau\right)\left((r_j-r^\star)'\,\mathbb{V}[\beta](r_j-r^\star)\right)^{1/2}\\ &\quad\times\sup_{(a,b)}\left(\mathfrak{c}(a,b)'\,\mathbb{V}[\beta]\mathfrak{c}(a,b)\right)^{1/2}.\end{aligned}$$

The remaining task is to bound the integral of $\mathbb{V}^h_{p_j,r_j(\tau)}[\alpha]$. If α is normally distributed, then we can appeal to Lemma BLI to conclude that $\mathbb{V}^h_{p_j,r_j(\tau)}[\alpha]\le\mathbb{V}[\alpha]$ for all τ, which yields the bound (10) with $r^\star$ set to $\bar{r}$.

If $-\alpha$ has a log-normal distribution, then we need to use a different argument. Let $\mathbb{E}^f_p$ denote expectation of (α,β) when (α,β) follow the density $f_p(a,b)$ that is proportional to $\exp(pa)f(a,b)$. Then for any function g, we can write

$$\begin{aligned}\mathbb{E}^h_w[g(\alpha)] &\equiv \frac{\mathbb{E}\left[g(\alpha)\pi_0(\alpha,\beta)\exp(\alpha p)\exp(\beta' r)\right]}{\mathbb{E}\left[\pi_0(\alpha,\beta)\exp(\alpha p)\exp(\beta' r)\right]}\\ &= \frac{\mathbb{E}\left[g(\alpha)\exp(\alpha p)\bar{\pi}_0(\alpha)\right]}{\mathbb{E}\left[\exp(\alpha p)\bar{\pi}_0(\alpha)\right]} \equiv \frac{\mathbb{E}^f_p[g(\alpha)\bar{\pi}_0(\alpha)]}{\mathbb{E}^f_p[\bar{\pi}_0(\alpha)]},\end{aligned}$$

where we have defined

$$\bar{\pi}_0(a)\equiv\mathbb{E}\left[\pi_0(\alpha,\beta)\left(\frac{\exp(\beta' r)}{\mathbb{E}[\exp(\beta' r)]}\right)\Big|\alpha=a\right].$$

Because $\bar{\pi}_0(a)$ is a weighted average of choice probabilities it is bounded by one, so we can conclude that

$$\mathbb{E}^h_w[g(\alpha)]\le\frac{\mathbb{E}^f_p[g(\alpha)]}{\mathbb{E}^f_p[\bar{\pi}_0(\alpha)]}.$$

Taking $g(a)\equiv a-\mathbb{E}^f_p[\alpha]$, we have that

$$\mathbb{V}^h_w[\alpha]\equiv\mathbb{E}^h_w\left[(\alpha-\mathbb{E}^h_w[\alpha])^2\right]\le\mathbb{E}^h_w\left[(\alpha-\mathbb{E}^f_p[\alpha])^2\right]\le\frac{\mathbb{E}^f_p\left[(\alpha-\mathbb{E}^f_p[\alpha])^2\right]}{\mathbb{E}^f_p\left[\bar{\pi}_0(\alpha)\right]}\equiv\frac{\mathbb{V}^f_p[\alpha]}{\mathbb{E}^f_p\left[\bar{\pi}_0(\alpha)\right]}.$$

If $-\alpha$ has a log-normal distribution, then $\mathbb{V}^f_p[\alpha]\le\mathbb{V}[\alpha]$ for any $p\ge 0$ because $\mathbb{V}^f_p[\alpha]$ is the second derivative of the cumulant generating function of α, which is decreasing for $p\ge 0$ and equal to $\mathbb{V}[\alpha]$ when $p=0$. Because $\bar{\pi}_0$ depends on r, we apply this inequality at each $w=(p_j,r_j(\tau))$ for $\tau\in[0,1]$ and bound the resulting denominators below by their infimum over τ, which yields (11) after rewriting

$$\mathbb{E}^f_{p_j}[\bar{\pi}_0(\alpha)] = \mathbb{E}\left[\pi_0(\alpha,\beta)\left(\frac{\exp(\alpha p_j + r_j(\tau)'\beta)}{\mathbb{E}[\exp(\alpha p_j + r_j(\tau)'\beta)]}\right)\right].$$

Q.E.D.

The following lemma is used in the proofs of the auxiliary lemmas.

Lemma TDI. (Tilted distribution under independence) Suppose that α and β are independent with density $f(a,b) = f_\alpha(a)f_\beta(b)$. Let $h_w(b|a)$ denote the conditional densities of β given $\alpha = a$ implied by h_w. Then

$$h_w(b|a) = \frac{\pi_0(a,b)\exp(b'r)f_\beta(b)}{\int \pi_0(a,z)\exp(z'r)f_\beta(z)\,dz}.$$

Similarly, if $h_w(a|b)$ is the conditional density of α given $\beta = b$ implied by h_w, then

$$h_w(a|b) = \frac{\pi_0(a,b)\exp(ap)f_\alpha(a)}{\int \pi_0(z,b)\exp(zp)f_\alpha(z)\,dz}.$$

Proof of Lemma TDI. Suppose that α and β are independent with density $f(a,b) = f_\alpha(a)f_\beta(b)$. Then the conditional density of β given α implied by h_w is

$$\begin{aligned} h_w(b|a) &\equiv \frac{h_w(a,b)}{\int h_w(a,z)\,dz} \\ &= \frac{\exp(ap+b'r)h(a,b)}{\int \exp(ap+z'r)h(a,z)\,dz} \\ &= \frac{\pi_0(a,b)\exp(b'r)f_\alpha(a)f_\beta(b)}{\int \pi_0(a,z)\exp(z'r)f_\alpha(a)f_\beta(z)\,dz} = \frac{\pi_0(a,b)\exp(b'r)f_\beta(b)}{\int \pi_0(a,z)\exp(z'r)f_\beta(z)\,dz}. \end{aligned}$$

The derivation of $h_w(a|b)$ is symmetric. Q.E.D.

Lemma CMD. (Conditional mean derivative under independence) Suppose that α and β are independent with density $f(a,b) = f_\alpha(a)f_\beta(b)$. Let $\mu_w(a) \equiv \mathbb{E}^h_w[\beta|\alpha = a]$ be the vector-valued function with ℓth component $\mathbb{E}^h_w[\beta_\ell|\alpha = a]$. Then

$$\frac{d}{da}\mu_w(a) = \begin{bmatrix} -\,\mathbb{C}^h_w[\beta_1, \bar{p}(\alpha,\beta)|\alpha = a] \\ \vdots \\ -\,\mathbb{C}^h_w[\beta_L, \bar{p}(\alpha,\beta)|\alpha = a] \end{bmatrix} \equiv -\,\mathbb{C}^h_w[\beta, \bar{p}(\alpha,\beta)|\alpha = a],$$

where $\bar{p}(a,b) \equiv \sum_{k=0}^J \pi_k(a,b)p_k$ is the probability-weighted average price.

Proof of Lemma CMD. Lemma TDI gives the conditional distribution of β given $\alpha = a$ implied by h_w, which we write concisely as

$$h_w(b|a) = \frac{\pi_0(a,b)\exp(b'r)f_\beta(b)}{\int \pi_0(a,z)\exp(z'r)f_\beta(z)\,dz} \equiv \frac{h_w^{\mathrm{NUM}}(a,b)}{\int h_w^{\mathrm{NUM}}(a,z)\,dz} \equiv \frac{h_w^{\mathrm{NUM}}(a,b)}{h_w^{\mathrm{DEN}}(a)}.$$

Then

$$\begin{aligned}
\frac{\partial}{\partial a} h_w^{\mathrm{NUM}}(a,b) &= \left(\frac{\partial}{\partial a}\pi_0(a,b)\right)\exp(b'r)f_\beta(b) \\
&= \pi_0(a,b)\left(\frac{\partial}{\partial a}\log \pi_0(a,b)\right)\exp(b'r)f_\beta(b) \\
&= -\pi_0(a,b)\bar{p}(a,b)\exp(b'r)f_\beta(b) = -\bar{p}(a,b)h_w^{\mathrm{NUM}}(a,b),
\end{aligned}$$

where the second-to-last equality follows from some well-known logit properties:

$$\begin{aligned}
\frac{\partial}{\partial a}\log \pi_0(a,b) &= \frac{\partial}{\partial a} - \log\left(\sum_{j=0}^{J}\exp(p_j a + r_j' b + \delta_j)\right) \\
&= \frac{-\sum_{j=0}^{J} p_j \exp(p_j a + r_j' b + \delta_j)}{\sum_{j=0}^{J}\exp(p_j a + r_j' b + \delta_j)} = -\sum_{j=0}^{J} p_j \pi_j(a,b) \equiv -\bar{p}(a,b).
\end{aligned}$$

Also,

$$\frac{d}{da} h_w^{\mathrm{DEN}}(a) = \int \frac{\partial}{\partial a} h_w^{\mathrm{NUM}}(a,z)\,dz = -\int \bar{p}(a,z)h_w^{\mathrm{NUM}}(a,z)\,dz.$$

It follows that

$$\begin{aligned}
\frac{\partial}{\partial a} h_w(b|a) &= \frac{h_w^{\mathrm{DEN}}(a)\frac{\partial}{\partial a}h_w^{\mathrm{NUM}}(a,b) - h_w^{\mathrm{NUM}}(a,b)\frac{d}{da}h_w^{\mathrm{DEN}}(a)}{h_w^{\mathrm{DEN}}(a)^2} \\
&= -\bar{p}(a,b)\frac{h_w^{\mathrm{NUM}}(a,b)}{h_w^{\mathrm{DEN}}(a)} + \left(\frac{h_w^{\mathrm{NUM}}(a,b)}{h_w^{\mathrm{DEN}}(a)}\right)\left(\frac{\int \bar{p}(a,z)h_w^{\mathrm{NUM}}(a,z)\,dz}{h_w^{\mathrm{DEN}}(a)}\right) \\
&= h_w(b|a)\,\mathbb{E}_w^h[\bar{p}(\alpha,\beta)|\alpha = a] - \bar{p}(a,b)h_w(b|a).
\end{aligned}$$

We conclude that for each component $\ell = 1, \ldots, L$,

$$\begin{aligned}
\left[\frac{d}{da}\mu_w(a)\right]_\ell &= \int \beta_\ell \frac{\partial}{\partial a} h_w(b|a)\, db \\
&= \left(\int \beta_\ell h_w(b|a)\, db\right) \mathbb{E}^h_w[\bar{p}(\alpha,\beta)|\alpha = a] - \int \beta_\ell \bar{p}(a,b) h_w(b|a)\, db \\
&= \mathbb{E}^h_w[\beta_\ell|\alpha = a]\, \mathbb{E}^h_w[\bar{p}(\alpha,\beta)|\alpha = a] - \mathbb{E}^h_w[\beta_\ell \bar{p}(\alpha,\beta)|\alpha = a] \\
&= -\,\mathbb{C}^h_w[\beta_\ell, \bar{p}(\alpha,\beta)|\alpha = a],
\end{aligned}$$

which was the claim. Q.E.D.

The following generalization of Stein's identity (also known as Stein's lemma) appears to be known, if perhaps somewhat obscure (see, for example, Cacoullos and Papathanasiou, 1989, Lemma 3.1). The proof involves a quick application of integration by parts, which we include for completeness.

Lemma SKG. (Stein's kernel generalization) Let X be a continuous random variable with density f supported on an interval (a, b) that could have $a = -\infty$ and/or $b = \infty$. Suppose that $\mathbb{E}[X]$ exists and define the Stein kernel

$$\mathcal{K}(x) \equiv \frac{1}{f(x)} \int_x^b (z - \mathbb{E}[X]) f(z)\, dz.$$

Let g be an absolutely continuous function and let g' denote its almost everywhere derivative. Suppose that $\lim_{x\to a} \mathcal{K}(x) f(x) g(x) = 0$ and $\lim_{x\to b} \mathcal{K}(x) f(x) g(x) = 0$. Then

$$\mathbb{C}[X, g(X)] = \mathbb{E}\left[\mathcal{K}(X) g'(X)\right],$$

assuming that the latter expectation exists. Moreover, $\mathcal{K}(x) \geq 0$ for all x and $\mathbb{E}[\mathcal{K}(X)] = \mathbb{V}[X]$ if the variance exists.

Proof of Lemma SKG. Observe that $(d/dx)(f(x)\mathcal{K}(x)) = -(x - \mathbb{E}[X]) f(x)$ and integrate by parts:

$$\begin{aligned}
\mathbb{C}[X, g(X)] &= \int_a^b (z - \mathbb{E}[X]) g(z) f(z)\, dz \\
&= -f(z)\mathcal{K}(z) g(z)\Big|_a^b - \int_a^b -\mathcal{K}(z) f(z) g'(z)\, dz = \mathbb{E}[\mathcal{K}(X) g'(X)],
\end{aligned}$$

where the final equality used the boundary conditions. Setting $g(x) = x$ gives $\mathbb{E}[\mathcal{K}(X)] = \mathbb{C}[X, X] = \mathbb{V}[X]$.

To see that $\mathcal{K}(x) \geq 0$, note that $\lim_{x\to a} f(x)\mathcal{K}(x) = 0$, that $\lim_{x\to b} f(x)\mathcal{K}(x) = 0$, and that $(d/dx)(f(x)\mathcal{K}(x)) = -(x - \mathbb{E}[X])f(x)$ is non-negative for $x \leq \mathbb{E}[X]$ and non-positive for $x \geq \mathbb{E}[X]$; that is, $x \mapsto f(x)\mathcal{K}(x)$ starts at zero, increases up to $x = \mathbb{E}[X]$, then decreases afterwards back to zero. Because $f(x)$ is non-negative, we conclude that $\mathcal{K}(x) \geq 0$ as well. Q.E.D.

Lemma BLI. (Brascamp-Lieb inequality) Suppose that α and β are independent. If α is normally distributed, then

$$\mathbb{V}_w^h[\alpha] \leq \mathbb{V}[\alpha].$$

If β is normally distributed and g is a continuously differentiable scalar-valued function with gradient $\nabla_\beta g(\beta)$, then

$$\mathbb{V}_w^h[g(\beta)|\alpha] \leq \mathbb{E}[\nabla g(\beta)' \, \mathbb{V}[\beta]\nabla g(\beta)|\alpha].$$

Proof of Lemma BLI. We begin with the second claim.

Lemma TDI shows that

$$\begin{aligned} h_w(b|a) &\propto \pi_0(a,b)\exp(b'r)f_\beta(b) \\ &= \exp(-(-b'r - \log\pi_0(a,b) - \log f_\beta(b))) \equiv \exp(-\Upsilon(a,b)). \end{aligned}$$

Because f_β is normal and therefore strictly log-concave, $-\log f_\beta(b)$ is a strictly convex function of b. Also,

$$-\log\pi_0(a,b) = \log\left(\sum_{j=0}^{J}\exp(p_j a + r_j' b + \delta_j)\right)$$

is the log-sum-exp function, which is strictly convex. It follows that $\Upsilon(a,b)$ is a strictly convex function of b for any fixed a, which means that $h_w(b|a)$ is strictly log-concave given a. Theorem 4.1 in Brascamp and Lieb (1976) then implies that

$$\mathbb{V}_w^h[g(\beta)|\alpha] \leq \mathbb{E}_w^h\left[\nabla g(\beta)'(\nabla_b^2\Upsilon(\alpha,\beta))^{-1}\nabla g(\beta)\middle|\alpha\right], \tag{38}$$

where $\nabla_b^2\Upsilon(a,b)$ is the Hessian of Υ with respect to b. Because $-\log\pi_0(a,b)$ is convex, we have that

$$\nabla_b^2\Upsilon(a,b) = \nabla_b^2\left(-\log\pi_0(a,b)\right) + \nabla_b^2\left(-\log f_\beta(b)\right) \succeq \nabla_b^2\left(-\log f_\beta(b)\right) = \mathbb{V}[\beta]^{-1},$$

where $\succeq$ denotes greater than in the positive semidefinite order and the final equality used the assumption that β is normally distributed. It follows that $\mathbb{V}[\beta] \succeq (\nabla_b^2 \Upsilon(a,b))^{-1}$. Returning to (38), we conclude that

$$\mathbb{V}_w^h[g(\beta)|\alpha] \leq \mathbb{E}\left[\nabla g(\beta)'(\nabla_b^2 \Upsilon(\alpha,\beta))^{-1}\nabla g(\beta)\middle|\alpha\right] \leq \mathbb{E}\left[\nabla g(\beta)'\,\mathbb{V}[\beta]\nabla g(\beta)\middle|\alpha\right],$$

which was the second claim.

For the first claim, note that the marginal density of α under h_w satisfies

$$h_w(a) \propto \exp(ap) f_\alpha(a) \upsilon(a) \quad \text{where} \quad \upsilon(a) \equiv \int \pi_0(a,b)\exp(b'r) f_\beta(b)\, db.$$

The integrand in the definition of $\upsilon(a)$ is strictly log-concave for any a, following the same arguments as above. Theorem 8 of Prékopa (1973) shows that integrating it over b preserves log-concavity, so that $\upsilon(a)$ is log-concave in a. Because α is normally distributed, it follows that

$$\frac{d^2}{da^2}\left(-\log h_w(a)\right) = \frac{1}{\mathbb{V}[\alpha]} + \frac{d^2}{da^2}\left(-\log \upsilon(a)\right) \geq \frac{1}{\mathbb{V}[\alpha]}.$$

Applying Theorem 4.1 of Brascamp and Lieb (1976), with the identity function then gives $\mathbb{V}_w^h[\alpha] \leq \mathbb{V}[\alpha]$, which was the first claim. Q.E.D.

SA.2.6 Proof of Proposition 4

Proof of Proposition 4. From Lemmas PWT and EFC, we have that

$$\eta(p,r) = \mathbb{E}_{p,r}^h[\alpha], \qquad \frac{\partial}{\partial r_\ell}\eta(p,r) = \mathbb{C}_{p,r}^h[\alpha,\beta_\ell],$$

$$\frac{\partial}{\partial p}\eta(p,r) = \mathbb{V}_{p,r}^h[\alpha], \qquad \frac{\partial^2}{\partial r_\ell \partial p}\eta(p,r) = \mathbb{K}_{p,r}^h[\alpha,\alpha,\beta_\ell],$$

$$\text{and} \quad \frac{\partial^2}{\partial p^2}\eta(p,r) = \mathbb{K}_{p,r}^h[\alpha,\alpha,\alpha].$$

Starting with the cross-derivative, we have

$$\frac{\partial^2}{\partial p \partial r_\ell}\eta_{\mathrm{Q}}(p,r) = -\frac{\theta_{01,\ell}}{p^2},$$

which matches the cross-derivative of η at $(\bar{p},\bar{r})$ if we set

$$\theta_{01} = -\bar{p}^2\,\mathbb{K}_{\bar{p},\bar{r}}^h[\alpha,\alpha,\beta].$$

The first derivative of the quadratic with respect to r_ℓ is

$$\frac{\partial}{\partial r_\ell}\eta_{\mathrm{Q}}(p,r) = \frac{\theta_{01,\ell}}{p} + \theta_{11,\ell},$$

which can be matched to its η counterpart by setting

$$\theta_{11} = \mathbb{C}^h_{\bar{p},\bar{r}}[\alpha,\beta] - \frac{\theta_{01}}{\bar{p}} = \mathbb{C}^h_{\bar{p},\bar{r}}[\alpha,\beta] + \bar{p}\,\mathbb{K}^h_{\bar{p},\bar{r}}[\alpha,\alpha,\beta].$$

Moving to the price derivatives, we have

$$\frac{\partial}{\partial p}\eta_{\mathrm{Q}}(p,r) = \frac{-\theta_{00} - \theta_{01}'(r-\bar{r})}{p^2} + \theta_2 \quad \text{and} \quad \frac{\partial^2}{\partial p^2}\eta_{\mathrm{Q}}(p,r) = \frac{2(\theta_{00} + \theta_{01}'(r-\bar{r}))}{p^3}.$$

Since we've already solved for θ_{01}, we can set this system of two equations equal to their η counterparts evaluated at $(\bar{p},\bar{r})$ then solve for the two unknowns:

$$\theta_{00} = \frac{1}{2}\bar{p}^3\,\mathbb{K}^h_{\bar{p},\bar{r}}[\alpha,\alpha,\alpha] \quad \text{and} \quad \theta_2 = \mathbb{V}^h_{\bar{p},\bar{r}}[\alpha] + \frac{1}{2}\bar{p}\,\mathbb{K}^h_{\bar{p},\bar{r}}[\alpha,\alpha,\alpha].$$

Finally, we get θ_{10} by setting $\eta_{\mathrm{Q}}(\bar{p},\bar{r}) = \eta(\bar{p},\bar{r})$ directly:

$$\begin{aligned}\theta_{10} = \eta(\bar{p},\bar{r}) - \frac{\theta_{00}}{\bar{p}} - \theta_2\bar{p} &= \mathbb{E}^h_{\bar{p},\bar{r}}[\alpha] - \tfrac{1}{2}\bar{p}^2\,\mathbb{K}^h_{\bar{p},\bar{r}}[\alpha,\alpha,\alpha] - \Big(\bar{p}\,\mathbb{V}^h_{\bar{p},\bar{r}}[\alpha] + \tfrac{1}{2}\bar{p}^2\,\mathbb{K}^h_{\bar{p},\bar{r}}[\alpha,\alpha,\alpha]\Big) \\ &= \mathbb{E}^h_{\bar{p},\bar{r}}[\alpha] - \bar{p}\,\mathbb{V}^h_{\bar{p},\bar{r}}[\alpha] - \bar{p}^2\,\mathbb{K}^h_{\bar{p},\bar{r}}[\alpha,\alpha,\alpha].\end{aligned}$$

Q.E.D.

SA.2.7 Proof of Proposition 5

Proof of Proposition 5. Multiplying and dividing by $\mathbb{E}[\pi_j(\alpha,\beta)\pi_k(\alpha,\beta)]$ in the cross-price component of (23) gives

$$\eta^\star_{kj}(\boldsymbol{p},\boldsymbol{r},\boldsymbol{c},\boldsymbol{\xi}) = \frac{\mathbb{E}[\pi_j(\alpha,\beta)\pi_k(\alpha,\beta)]}{\mathbb{E}[\pi_k(\alpha,\beta)]}\left(-\frac{\mathbb{E}[\alpha\pi_j(\alpha,\beta)\pi_k(\alpha,\beta)]}{\mathbb{E}[\pi_j(\alpha,\beta)\pi_k(\alpha,\beta)]}\right).$$

The first factor is the weighted probability term in (24). We now show that the second term can be written as a function $\eta^\times(p_k + p_j, r_k + r_j)$ that depends only on the sum of the characteristics for j and k that have random coefficients.

The argument uses the same exponential tilting connection developed in Appendix SA.2.2 for the own-price elasticities. Using the same shorthand notation as in that section, we have

$$\pi_j(v)\pi_k(v) = \pi_0(v)^2 \exp(w_j'v + \delta_j) \exp(w_k'v + \delta_k)$$

so that

$$\frac{\pi_j(v)\pi_k(v)}{\mathbb{E}[\pi_j(\nu)\pi_k(\nu)]} = \frac{\pi_0(v)^2 \exp((w_k + w_j)'v)}{\mathbb{E}\left[\pi_0(\nu)^2 \exp((w_k + w_j)'\nu)\right]}.$$

It follows that

$$-\frac{\mathbb{E}[\alpha\pi_j(\alpha,\beta)\pi_k(\alpha,\beta)]}{\mathbb{E}[\pi_j(\alpha,\beta)\pi_k(\alpha,\beta)]} = -\,\mathbb{E}^{h^\times}_{w_k+w_j}[\alpha] \equiv \eta^\times(w_k + w_j),$$

where $\mathbb{E}^{h^\times}_w$ denotes expectation when $\nu \equiv (\alpha, \beta)$ has density

$$h^\times_w(v) \propto f(v)\pi_0(v)^2 \exp(w'v).$$

Note that the densities $h^\times_\omega(v)$ form a proper exponential family in the natural parameter ω, with canonical sufficient statistic v, base measure $f(v)\pi_0(v)^2$, and log partition function

$$A^\times(\omega) \equiv \log \int f(v)\pi_0(v)^2 \exp(\omega'v)\, dv = \log \mathbb{E}\left[\exp(\omega'\nu)\pi_0(\nu)^2\right].$$

Q.E.D.

SA.2.8 Proof of Proposition 6

Proof of Proposition 6. Write the ratio as

$$\frac{\mathbb{E}\left[\pi_j(\nu)\frac{\pi_k(\nu)}{\mathbb{E}[\pi_k(\nu)]}\right]}{\mathbb{E}[\pi_j(\nu)]} = \frac{\mathbb{E}[\pi_j(\nu)\pi_k(\nu)]}{\mathbb{E}[\pi_j(\nu)]\,\mathbb{E}[\pi_k(\nu)]} = 1 + \frac{\mathbb{C}[\pi_j(\nu), \pi_k(\nu)]}{\mathbb{E}[\pi_j(\nu)]\,\mathbb{E}[\pi_k(\nu)]}. \tag{39}$$

A first-order Taylor series expansion of $\pi_j(\nu)$ around $\nu = \mathbb{E}[\nu]$ gives

$$\pi_j(\nu) = \pi_j(\mathbb{E}[\nu]) + \nabla\pi_j(\mathbb{E}[\nu])'\,(\nu - \mathbb{E}[\nu]) + \frac{1}{2}\,(\nu - \mathbb{E}[\nu])'\,\nabla^2\pi_j(\nu^\star)\,(\nu - \mathbb{E}[\nu])\,, \tag{40}$$

where $\nu^\star$ lies between ν and $\mathbb{E}[\nu]$. The gradient of the logit choice probability evaluated at $\mathbb{E}[\nu]$ is given by

$$\nabla\pi_j(\mathbb{E}[\nu]) = \pi_j(\mathbb{E}[\nu])\left(w_j - \sum_{\ell=0}^{J}\pi_\ell(\mathbb{E}[\nu])w_\ell\right) \equiv \pi_j(\mathbb{E}[\nu])\,(w_j - \tilde{w})\,. \tag{41}$$

The linear term in (40) vanishes under expectation, leaving

$$\mathbb{E}[\pi_j(\nu)] = \pi_j(\mathbb{E}[\nu]) + \frac{1}{2}\,\mathbb{E}\left[(\nu - \mathbb{E}[\nu])'\,\nabla^2\pi_j(\nu^\star)\,(\nu - \mathbb{E}[\nu])\right] = \pi_j(\mathbb{E}[\nu]) + O(\mathbb{V}[\nu]). \tag{42}$$

Similarly applying (40) to both $\pi_j(\nu)$ and $\pi_k(\nu)$ in the covariance term and using (41) gives

$$\mathbb{C}[\pi_j(\nu), \pi_k(\nu)] = \pi_j(\mathbb{E}[\nu])\pi_k(\mathbb{E}[\nu])\,(w_j - \tilde{w})'\,\mathbb{V}[\nu]\,(w_k - \tilde{w}) + O\left(\mathbb{E}\,\|\nu - \mathbb{E}[\nu]\|^3\right). \tag{43}$$

Substituting (42) and (43) into (39) we then have

$$\begin{aligned}
&\frac{\mathbb{E}[\pi_j(\nu)\pi_k(\nu)]}{\mathbb{E}[\pi_j(\nu)]\,\mathbb{E}[\pi_k(\nu)]} \\
&\quad = 1 + \frac{\pi_j(\mathbb{E}[\nu])\pi_k(\mathbb{E}[\nu])\,(w_j - \tilde{w})'\,\mathbb{V}[\nu]\,(w_k - \tilde{w}) + O\left(\mathbb{E}\,\|\nu - \mathbb{E}[\nu]\|^3\right)}{\pi_j(\mathbb{E}[\nu])\pi_k(\mathbb{E}[\nu])\,(1 + O(\mathbb{V}[\nu]))} \\
&\quad = 1 + \frac{(w_j - \tilde{w})'\,\mathbb{V}[\nu](w_k - \tilde{w})}{1 + O(\mathbb{V}[\nu])} + \frac{O(\mathbb{E}\,\|\nu - \mathbb{E}[\nu]\|^3)}{1 + O(\mathbb{V}[\nu])} \approx 1 + (w_j - \tilde{w})'\,\mathbb{V}[\nu](w_k - \tilde{w}),
\end{aligned}$$

which is the stated approximation. Q.E.D.

SA.3 Data and Mixed Logit Specifications

This section contains details on the data and mixed logit specifications for the five applications used in this paper. Sections SA.3.1–SA.3.4 are about data that has been used in other papers. Section SA.3.5 gives some additional details about the toothpaste data.

SA.3.1 Synthetic cereal data

The synthetic cereal data is the widely-used ready-to-eat cereal dataset distributed with the `pyblp` package (Conlon and Gortmaker, 2020), which comes from Nevo (2000) and was modeled after the actual data used by Nevo (2001). It consists of $J = 24$ branded cereals observed across $T = 94$ markets, defined as city-quarter pairs. Prices are measured in dollars per serving and range from \$0.05 to \$0.23 across products and markets. Each product has two additional observed characteristics: sugar content in grams per serving, which ranges between 0 and 20, and a binary mushy indicator for whether the cereal turns soft in milk.

Our baseline mixed logit specification uses exactly the same parameter values that

Table SA.1: Random coefficient distribution (baseline synthetic cereal)

	Mean	Constant	Price	Sugar	Mushy
Constant	—	1.20			
Price	−31.23	−0.79	13.84		
Sugar	—	0.53	−0.89	0.12	
Mushy	—	−0.60	0.09	0.17	0.98

Notes: Diagonal entries are standard deviations and off-diagonal entries are correlations. The means of the non-price random coefficients are not identified because they do not vary across markets and the specification includes product fixed effects.

Conlon and Gortmaker (2020) find in their `pyblp` tutorial example for correlated random coefficients.[12] The specification has a random coefficient on price and random coefficients on a constant term, sugar, and mushy, which constitutes r_{jt}. There is a full set of product fixed effects, constituting c_{jt}. Note that the means of the non-price random coefficients are not identified due to the product fixed effects; see Nevo (2001). The random coefficients are jointly normal and—unlike our other baseline specifications—are estimated with correlated random coefficients. The point estimates of the variance-covariance matrix are reported in Table SA.1. Note that the standard deviation of the price coefficient appears large only because of the units in which price is measured, which are an order of magnitude smaller than the units of mushy and two orders of magnitude smaller than the units of sugar.

Table 2 reports several modifications of the baseline specification—each obtained by transforming the estimated covariance matrix of (α, β) while holding everything else about the data and parameters fixed—together with, in the final rows, the estimated demographic specification. The specification of the rows is as follows:

- *Price only:* the variances of all non-price random coefficients are set to zero.
- *Independent:* the correlations are set to zero.
- *Scale:* starting with the independent specification, the variance $\mathbb{V}[\alpha]$ or $\mathbb{V}[\beta]$ is scaled up or down by the stated amount.
- *Lognormal:* starting from the independent specification, the marginal distribution of the price coefficient is taken to be log-normal with the same mean and variance.
- *Price and mushy:* set the variances of the constant and sugar to zero and adjust the correlation between price and mushy by the stated amount.

[12] See version 1.2 of https://pyblp.readthedocs.io/en/stable/_notebooks/tutorial/nevo.html.

Table SA.2: Random coefficient distribution (synthetic cereal independent, demographics)

	Mean	Std. dev.	Income	Income^2	Age	Child
Constant	—	0.558	2.292		1.284	
Price	−62.730	3.312	588.325	−30.192		11.055
Sugar	—	0.006	−0.385		0.052	
Mushy	—	0.093	0.748		−1.353	

Notes: The means of the non-price random coefficients are not identified because they do not vary across markets and the specification includes product fixed effects.

- *Independent, demographics:* this specification deviates from the baseline by changing the variance-covariance matrix, which is diagonal, and accommodating the demographic specification used in the `pyblp` tutorial, which is modeled after Nevo (2000). This adds a term ΠD_t to (α, β), where Π is a matrix of constant parameters and D_t are four demeaned variables constructed as joint demographic draws from the Current Population Survey: log of household income, its square, age, and a child indicator. The matrix Π is restricted so that income, income squared, and the child indicator enter the price coefficient, while income and age enter the constant, sugar, and mushy coefficients. Table SA.2 reports the estimates from this specification, which are taken from the `pyblp` tutorial.

SA.3.2 US automobile data

The US automobile data is the dataset originally used by Berry et al. (1995), which we obtained through the `pyblp` package (Conlon and Gortmaker, 2020). It is an unbalanced panel of new vehicle models observed over model years (1971–1990), each viewed as a market. The number of products varies by year with up to 150 models per year and a total of 2,217 model-year observations. Prices are measured in thousands of (1983) dollars and range from 3.4 to 68.6, with a mean of 11.8. The non-price characteristics are horsepower per weight, an air-conditioning indicator, miles per dollar, size, and a constant.

We use exactly the same specification and parameter values as in the "Best Practices" column of Table 8 in Conlon and Gortmaker (2020), which are reproduced in Table SA.3. Independent normal random coefficients are included on a constant term and the four non-price characteristics. The price coefficient is random, determined by the income distribution: $\alpha_i = -45.898/\textsc{inc}_i$ with $\textsc{inc}_i$ drawn from the empirical income distribution in each market (year), and -45.898 estimated as a constant parameter. These income draws are independent of the normal draws on the other

Table SA.3: Random coefficient means and standard deviations (US auto)

Characteristic	Mean	Std. dev.	$\times$ (1/income)
Constant	−6.679	2.962	—
Price	—	—	−45.898
Horsepower/weight	2.774	1.388	—
Air conditioning	0.572	1.424	—
Miles/dollar	0.340	0.072	—
Size	3.920	0.231	—

Notes: Values are from Conlon and Gortmaker (2020, Table 8).

random coefficients, so this is an example of a specification with independent random coefficients.

SA.3.3 China automobile data

Li (2018) studies the welfare consequences of automobile license allocations in Beijing and Shanghai. We obtain his data from his replication package. The data is a panel of new automobile models sold in four Chinese cities—Beijing, Nanjing, Shanghai, and Tianjin—over 60 months spanning 2008–2012, giving a total of $T = 240$ city-month markets. There are 562 distinct models of automobiles at roughly 354 per market.

His specification has independent random coefficients on a constant, log price, and log engine displacement. Li considers license-allocation policies that make the pool of eligible buyers vary across markets and time. We use Li's estimated preference parameters and back out the market-product unobservables ξ_{jt} under a simpler static version that abstracts from the license-allocation policies. The mean price-elasticities from our estimates are -7.3 compared to Li's reported -7.6 under his policy environment.

In Li's specification the coefficient on price (which is *log* price—see below) has two components: an unobserved normal draw with an estimated mean of -15.015 and standard deviation of 0.067, plus a shift by $1.620 \log(\text{INC}_i)$, where 1.620 is estimated in the data and INC_i is a draw from the income distribution, which varies by market. The constant coefficient is treated similarly, so that the constant and price random coefficients are correlated through the income draw. Proposition 1 shows that this has no impact under the small-π approximation, because the constant term does not vary across products.

Unlike our other applications, Li includes log price $\tilde{p}_{jt} \equiv \log p_{jt}$ in the utility instead of the level of price, p_{jt}. Our results apply directly to this case by noting that the semi-elasticity function with respect to log price is the elasticity with respect to price,

$$\eta_j^\star(\tilde{\boldsymbol{p}}, \boldsymbol{r}, \boldsymbol{c}, \boldsymbol{\xi}) \equiv \frac{\partial s_j}{\partial \log p_j}\frac{1}{s_j} = \frac{\partial s_j}{\partial p_j}\frac{p_j}{s_j} \equiv \epsilon_j^\star(\boldsymbol{p}, \boldsymbol{r}, \boldsymbol{c}, \boldsymbol{\xi}),$$

so that results about semi-elasticities with log prices $\tilde{p}_j$ immediately translate to results about elasticities without multiplying by price as in (4).

SA.3.4 The French automobile calibrated data

D'Haultfœuille et al. (2019) estimate demand in the French new-car market. Their empirical application relies on proprietary vehicle-registration microdata and a household budget survey that are not included in their replication package, so their estimated model cannot be reproduced. We instead use the Monte Carlo data-generating process included with their replication package, which provides a fully specified mixed logit that the authors calibrated to resemble their application. The specification used in their Monte Carlo is the same as in their application except that it has only one non-price characteristic.

The design of the Monte Carlo simulation has $T = 100$ markets of $J = 24$ products each. There is one product characteristic other than price, which stands in for the several non-price characteristics—all with non-random coefficients—in their application. The number of markets includes four different demographic groups, which the authors use to study third-degree price discrimination.[13] Only price has a random coefficient. Its standard deviation is set to 0.4 and its mean varies between -1.50 and -3.00 across demographic groups.

SA.3.5 Toothpaste at a large US retailer

This appendix includes more details about the pricing test studied in Section 5 and sample construction discussed in Section 5.1. The test spanned 23 categories and 1,884 products in total, collectively accounting for close to 2% of the retailer's revenue. Details on the model specifications we use are given in the main text.

Status-quo pricing. Pricing at the retailer is decentralized across category managers, each of whom has broad discretion over their merchandise subject to category-

[13] The demographic groups have different ξ_{jt} terms, which is the criterion we use to define a market in Appendix SA.1. D'Haultfœuille et al. (2019, footnote 1) write that the different demographic groups "cannot be seen as different markets because the [ξ_{jt}] are allowed to be correlated." This seems to be a statement about statistical inference because correlation among unobservables has nothing to do with market definition. In any case, the counterfactuals that the authors include require considering elasticities within a demographic group.

level financial targets and a basic set of pricing rules. For example, larger-sized variants of a product must be priced above smaller-sized variants, and every product must be priced above its wholesale supplier cost. There is no consensus approach to setting prices among the managers, nor to which estimates of price elasticities of demand to use in doing so. The retailer purchases own-price elasticity estimates for some products from third-party vendors, but most managers rely instead on their own judgment, past experience, and intuition. Providing the retailer with credible, ground-truth estimates of demand elasticities was one motivation for the pricing test that we study.

Extensive margin. The pricing test was designed to isolate the intensive margin—which product a customer buys in their chosen store—rather than the choice of where to shop. It did this by creating random variation across clusters *within* each product, thereby holding the average price level of each store approximately fixed. Because the pricing test held the national average price of each product approximately fixed, it also minimized price responses by competing retailers, which typically price nationally (DellaVigna and Gentzkow, 2019).

Price variation. Three price-point variants were constructed for most products in the test. A middle variant was set, typically to the modal price of the product across stores prior to the start of the test; low and high variants were set symmetrically about it on a log scale. The gap between the middle and the low/high variants was chosen to be as large as possible subject to the following constraints:

- The gap between the middle and the low/high prices can be no more than 0.15 log points.
- The gap between the middle and the low/high prices must be at least $0.05.
- Every price must be above the product's supplier cost and, where applicable, above the supplier-negotiated minimum advertised price.
- Larger-sized variants of a product must be priced above smaller-sized variants.

For some products, no amount of positive price variation could satisfy all of these constraints simultaneously; these products were excluded from the test. Each store was sent a list of its assigned prices to update its shelf tags and scanner system, and online shoppers saw the price for their selected store; the retailer implemented operational locks to prevent other price changes on test products for the duration of the test. Certain markdown and clearance sales, planned before the test began, were permitted to override the assigned prices, generating a small amount of noncompliance with the assigned variants.

Total market size. Implementing a discrete choice model requires turning quantity data into shares by estimating market size.[14] First, we use the retailer's data to estimate the average number of distinct households shopping at each store in a given week. Second, we scale this count by Public Use Microdata Area (PUMA) estimates of the number of individuals per household to obtain the total population shopping at each store. Third, we multiply by an estimate of per-capita toothpaste consumption, using an upper bound of three brushings per person per day and 0.5 grams of toothpaste per brushing (Creeth et al., 2013). The outside option $j = 0$ is then the share of this potential toothpaste demand that is not purchased at the retailer during the pricing test.

Demographics. To include demographics we draw household income from the income distribution of the store's PUMA, standardized to have mean zero and unit standard deviation across markets.

SA.4 Interpreting Linear IV Estimands

This appendix derives the weighted average representation (25) of the product-specific linear IV estimands discussed in Section 5.2. The derivation uses familiar arguments from the literature on instrumental variables with heterogeneous treatment effects (Imbens and Angrist, 1994; Angrist and Imbens, 1995; Mogstad and Torgovitsky, 2024).

We begin by decomposing the linear IV estimand into a weighted average of Wald estimands. Let $y_{jt}(\boldsymbol{p}) \equiv \log(s_{jt}/s_{0t})(\boldsymbol{p})$ denote the log share ratio as a function of the price vector, and let $y_{jt} \equiv \log(s_{jt}/s_{0t})$ denote its observed value at the realized prices. Then an argument from Angrist and Imbens (1995, Theorem 2) shows that

$$\text{IV}_j = \frac{\mathcal{C}[y_{jt}, z_{\ell(j)t}]}{\mathcal{C}[p_{jt}, z_{\ell(j)t}]} = \sum_{a=2}^{n_{\ell(j)}} \tilde{\omega}_j^a \text{WALD}_{ja}, \tag{44}$$

where $\mathcal{C}$ is covariance over markets in the same way that $\mathcal{E}$ is expectation over markets. The Wald estimand is defined as

$$\text{WALD}_{ja} \equiv \frac{\mathcal{E}[y_{jt}|z_{\ell(j)t} = z_{\ell(j)}^a] - \mathcal{E}[y_{jt}|z_{\ell(j)t} = z_{\ell(j)}^{a-1}]}{\mathcal{E}[p_{jt}|z_{\ell(j)t} = z_{\ell(j)}^a] - \mathcal{E}[p_{jt}|z_{\ell(j)t} = z_{\ell(j)}^{a-1}]},$$

[14] In this empirical application, we use "market" to refer to the geographic area around a store. We do not take this as a legal definition of a market and we recognize that a market may be much broader than a single geographic area.

and the weights are defined as

$$\tilde{\omega}_j^a \equiv \frac{\mathcal{C}[z_{\ell(j)t}, \mathbb{1}[z_{\ell(j)t} \geq z^a_{\ell(j)}]] \left(\mathcal{E}[p_{jt}|z_{\ell(j)t} = z^a_{\ell(j)}] - \mathcal{E}[p_{jt}|z_{\ell(j)t} = z^{a-1}_{\ell(j)}]\right)}{\sum_{e=2}^{n_{\ell(j)}} \mathcal{C}[z_{\ell(j)t}, \mathbb{1}[z_{\ell(j)t} \geq z^e_{\ell(j)}]] \left(\mathcal{E}[p_{jt}|z_{\ell(j)t} = z^e_{\ell(j)}] - \mathcal{E}[p_{jt}|z_{\ell(j)t} = z^{e-1}_{\ell(j)}]\right)}. \quad (45)$$

Next, we interpret each Wald estimand through the lens of the general nonparametric demand model. Random assignment and exclusion, together with the assumption that line $\ell(j)$'s assignment does not affect the prices of other lines, imply that

$$\mathcal{E}[y_{jt}|z_{\ell(j)t} = z] = \mathcal{E}\left[y_{jt}\left(\boldsymbol{p}_{\ell(j)t}(z), \boldsymbol{p}_{-\ell(j)t}\right)\right].$$

Similarly for the average price,

$$\mathcal{E}[p_{jt}|z_{\ell(j)t} = z] = \mathcal{E}[p_{jt}(z)].$$

It follows that

$$\text{WALD}_{ja} = \frac{\mathcal{E}\left[y_{jt}\left(\boldsymbol{p}_{\ell(j)t}(z^a_{\ell(j)}), \boldsymbol{p}_{-\ell(j)t}\right) - y_{jt}\left(\boldsymbol{p}_{\ell(j)t}(z^{a-1}_{\ell(j)}), \boldsymbol{p}_{-\ell(j)t}\right)\right]}{\mathcal{E}\left[p_{jt}(z^a_{\ell(j)}) - p_{jt}(z^{a-1}_{\ell(j)})\right]}.$$

Multiplying and dividing allows us to rewrite this as a single weighted average of arc semi-elasticities:

$$\text{WALD}_{ja} = \mathcal{E}\left[\frac{\left(p_{jt}(z^a_{\ell(j)}) - p_{jt}(z^{a-1}_{\ell(j)})\right)}{\mathcal{E}\left[p_{jt}(z^a_{\ell(j)}) - p_{jt}(z^{a-1}_{\ell(j)})\right]} \breve{\eta}^a_{jt}\right], \quad (46)$$

where the arc semi-elasticity is

$$\breve{\eta}^a_{jt} \equiv \frac{y_{jt}\left(\boldsymbol{p}_{\ell(j)t}(z^a_{\ell(j)}), \boldsymbol{p}_{-\ell(j)t}\right) - y_{jt}\left(\boldsymbol{p}_{\ell(j)t}(z^{a-1}_{\ell(j)}), \boldsymbol{p}_{-\ell(j)t}\right)}{p_{jt}(z^a_{\ell(j)}) - p_{jt}(z^{a-1}_{\ell(j)})}.$$

In a market where the posted price does not respond to the reassignment, so that $p_{jt}(z^a_{\ell(j)}) = p_{jt}(z^{a-1}_{\ell(j)})$, the denominator of $\breve{\eta}^a_{jt}$ is zero. However, the numerator is zero in such a market as well: all products in line $\ell(j)$ carry the common posted line price, so $\boldsymbol{p}_{\ell(j)t}(z^a_{\ell(j)}) = \boldsymbol{p}_{\ell(j)t}(z^{a-1}_{\ell(j)})$, while $\boldsymbol{p}_{-\ell(j)t}$ does not vary with the assignment, and hence y_{jt} takes the same value at both price vectors. Such markets contribute nothing to the numerator of the previous expression for WALD_{ja}, so we adopt the convention that the product of the compliance factor and $\breve{\eta}^a_{jt}$ is zero in these markets, under which (46) is exact. The compliance factor multiplying $\breve{\eta}^a_{jt}$ in (46) averages to one across markets by construction.

Finally, we substitute (46) into (44) and rearrange the order of summation and expectation to obtain

$$\text{IV}_j = \mathcal{E}\left[\sum_{a=2}^{n_{\ell(j)}} \frac{\tilde{\omega}^a_j \left(p_{jt}(z^a_{\ell(j)}) - p_{jt}(z^{a-1}_{\ell(j)})\right)}{\mathcal{E}\left[p_{jt}(z^a_{\ell(j)}) - p_{jt}(z^{a-1}_{\ell(j)})\right]} \breve{\eta}^a_{jt}\right] \equiv \mathcal{E}\left[\sum_{a=2}^{n_{\ell(j)}} \omega^a_{jt}\breve{\eta}^a_{jt}\right],$$

which coincides with the definition of ω^a_{jt} given in (26). To see that the weights integrate to one, note that $\tilde{\omega}^a_j$ is non-stochastic and sums to one, while the remaining part of ω^a_{jt} has mean one. The weights are non-negative if larger assigned prices do not lead to smaller posted prices, that is, if $p_{jt}(z^a_{\ell(j)}) \geq p_{jt}(z^{a-1}_{\ell(j)})$ for every market and every a. This makes the second factor of ω^a_{jt} non-negative directly. Averaging over markets, it makes the first-stage differences in (45) non-negative as well. The covariances in (45) are non-negative because the instrument values are indexed in increasing order.

SA.5 Specification Test Implementation

This appendix explains how we implement the formal statistical tests referenced in Section 5.4. Let $\widehat{\text{IV}}_j$ denote the linear IV estimator. Let $\widehat{\text{IV}}^{\text{ML}}_j$ denote a mixed logit estimator of the IV estimand (25), assuming that the role of compliance is small as discussed in the main text. Under weak statistical conditions, $\widehat{\text{IV}}_j$ will be consistent for IV_j. If the parametric assumptions underlying the mixed logit are correct, then $\widehat{\text{IV}}^{\text{ML}}_j$ will also be consistent for IV_j. Our test statistic is the difference $\hat{T}_j \equiv \widehat{\text{IV}}_j - \widehat{\text{IV}}^{\text{ML}}_j$.

The variance of each $\hat{T}_j$ depends on the variances of $\widehat{\text{IV}}_j$ and $\widehat{\text{IV}}^{\text{ML}}_j$ as well as their covariances. Let $\widehat{\Sigma}^{\text{IV}}_j$ denote the estimated variance of the first estimator, which we construct through standard cluster-robust estimation, and let $\widehat{\Sigma}^{\text{ML}}_j$ denote the estimated variance of the second estimator, which we construct using the parametric bootstrap routine in `pyblp` with 99 bootstrap draws.

We use two approaches to getting an upper bound on the variance of each $\hat{T}_j$, which we denote by $\widehat{\Sigma}_j$. Our first approach is to assume that the correlation between each $\widehat{\text{IV}}_j$ and $\widehat{\text{IV}}^{\text{ML}}_j$ is positive, which seems reasonable under the null hypothesis that both are consistent, because both estimators are constructed from the same underlying data. This results in taking $\widehat{\Sigma}_j = \widehat{\Sigma}^{\text{IV}}_j + \widehat{\Sigma}^{\text{ML}}_j$. Our second approach is to use a Cauchy–Schwarz inequality; this requires no additional assumptions and results in taking $\widehat{\Sigma}_j = \left((\widehat{\Sigma}^{\text{IV}}_j)^{1/2} + (\widehat{\Sigma}^{\text{ML}}_j)^{1/2}\right)^2$. If $\widehat{\Sigma}_j$ is too large, then the tests that we conduct will still be valid, just conservative.

We implement single-product tests using an asymptotic t-test with test statistic $\hat{T}_j/\widehat{\Sigma}_j^{1/2}$, which we compare against standard normal critical values. We implement the joint product tests using a Bonferroni correction applied to the combination of the single-product tests when based on the always-valid Cauchy–Schwarz bound. Table SA.9 gives the full results of the tests. A large proportion of single-product tests reject across all specifications, even when using the highly conservative Cauchy–Schwarz variance bound. The joint test rejects at any conventional significance level across all specifications, despite using the conservative Cauchy–Schwarz variance bound and a conservative Bonferroni correction. Panels B and C show that restricting attention to products with higher compliance does not change these conclusions.

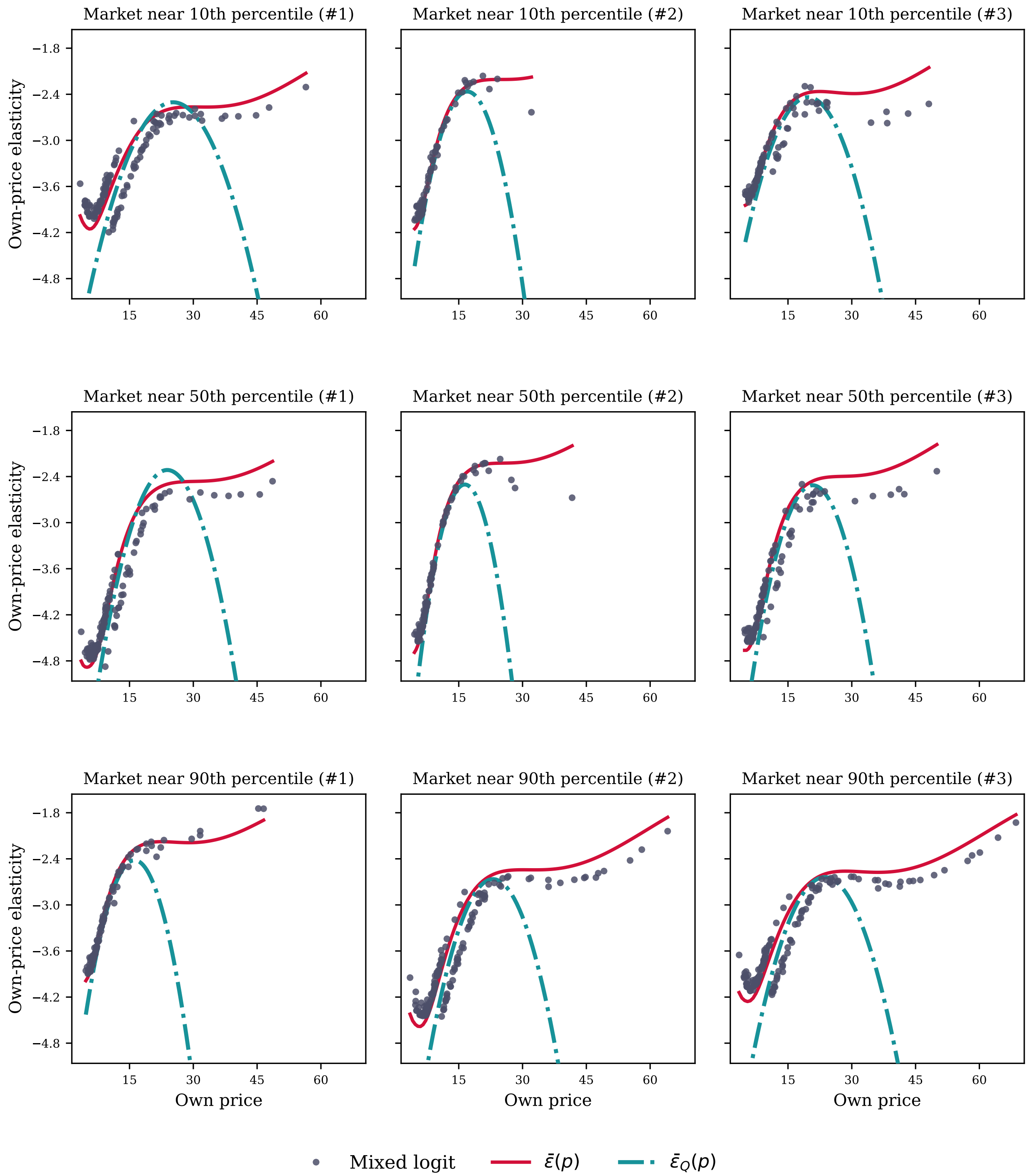
SA.6 Additional Exhibits

Table SA.4: Relative error of semi-elasticity approximations (market-level)

	ERR(η)			ERR($\bar{\eta}$)			ERR(η_Q)			ERR($\bar{\eta}_Q$)		
Percentile:	10	50	90	10	50	90	10	50	90	10	50	90
US autos	1.0	1.1	1.5	1.8	3.0	6.5	6.8	12.5	25.9	7.5	17.3	32.5
China autos	0.1	0.1	0.1	0.2	0.3	0.3	0.4	0.5	0.7	0.4	0.5	0.7
France autos	0.3	1.1	1.7	0.3	1.1	1.7	0.3	1.1	1.7	0.3	1.1	1.7
Synthetic cereal	4.2	4.7	5.1	6.7	7.6	8.6	4.6	5.1	5.8	7.0	7.8	8.6
Price only	2.0	2.4	2.8	2.0	2.4	2.8	2.0	2.3	2.7	2.0	2.3	2.7
Independent	4.2	4.7	5.2	5.1	5.9	7.0	3.7	4.3	5.0	5.3	6.1	7.0
Independent, lognormal	4.5	4.9	5.4	4.1	4.6	5.5	4.6	5.1	5.7	4.8	5.3	6.0
Independent, $\mathbb{V}[\alpha] \times 4$	5.7	17.1	86.5	20.0	34.3	133.8	5.3	12.8	71.8	19.1	32.9	135.5
Independent, $\mathbb{V}[\alpha] \times 4$*, lognormal*	5.0	5.3	6.0	4.9	5.8	7.1	4.9	5.4	6.2	5.8	6.5	7.3
Independent, $\mathbb{V}[\alpha] \div 4$	4.0	4.5	5.1	3.7	4.4	5.0	3.9	4.4	5.0	4.0	4.5	5.1
Independent, $\mathbb{V}[\beta] \times 4$	8.6	9.3	10.2	9.5	11.0	13.3	7.5	8.4	9.5	9.0	10.1	12.4
Independent, $\mathbb{V}[\beta] \div 4$	2.5	3.0	3.4	2.8	3.3	3.8	2.4	2.9	3.3	3.0	3.4	3.9
Price and mushy, correlation .2	2.6	3.1	3.6	1.9	2.6	3.3	2.5	2.9	3.5	2.5	3.0	3.5
Price and mushy, correlation .8	2.6	2.9	3.4	10.8	11.8	12.8	2.3	2.7	3.3	10.5	11.2	12.1
Independent, demographics	3.7	4.8	6.0	7.9	13.0	23.3	3.8	6.2	10.3	7.1	11.8	21.9

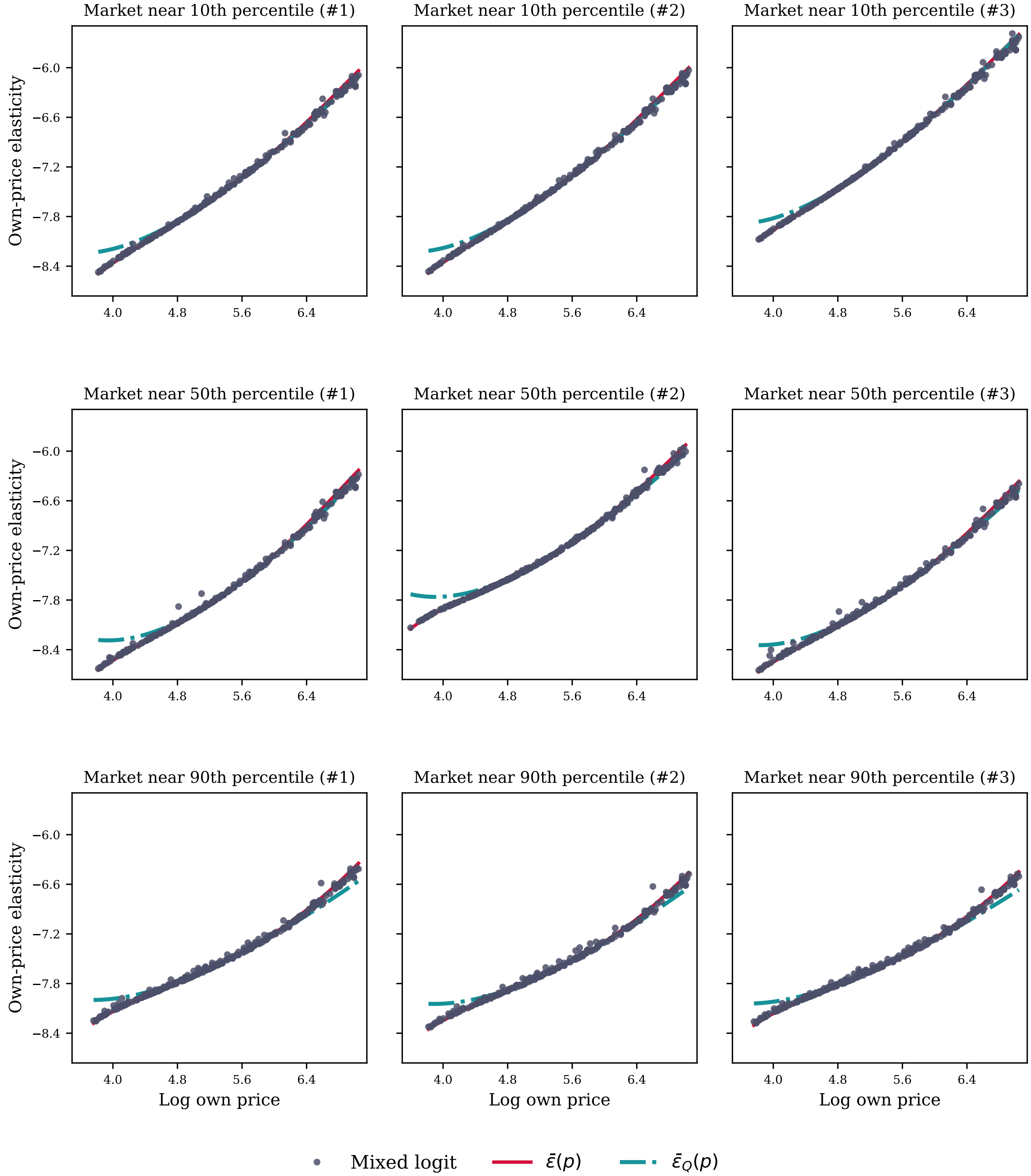
Notes: The relative error is defined in (7). In this table, we first take the average relative error across goods within each market. Then we report the 10, 50, and 90 percentiles of this average across markets. For other notes, see Table 2.

Figure SA.1: Elasticities for nine markets (US auto)



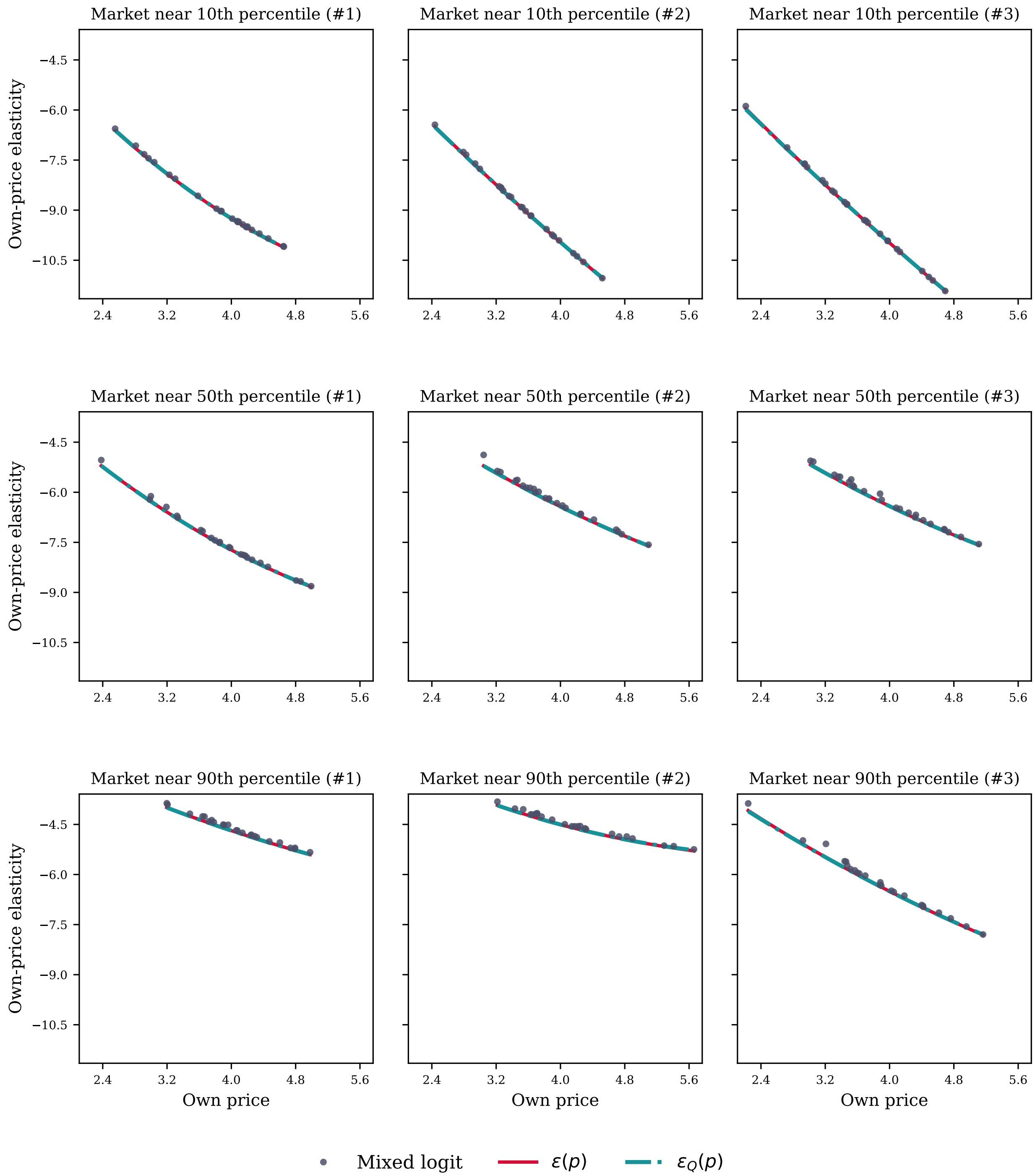
Notes: Actual elasticities against the price-only approximation $\bar{\epsilon}(p)$ and the quadratic price-only approximation $\epsilon_Q(p)$ for nine markets. The nine markets are chosen at the three that are closest to the 10th, 50th, and 90th percentiles in terms of average (over products) relative error of the small-π approximation ϵ.

Figure SA.2: Elasticities for nine markets (China auto)



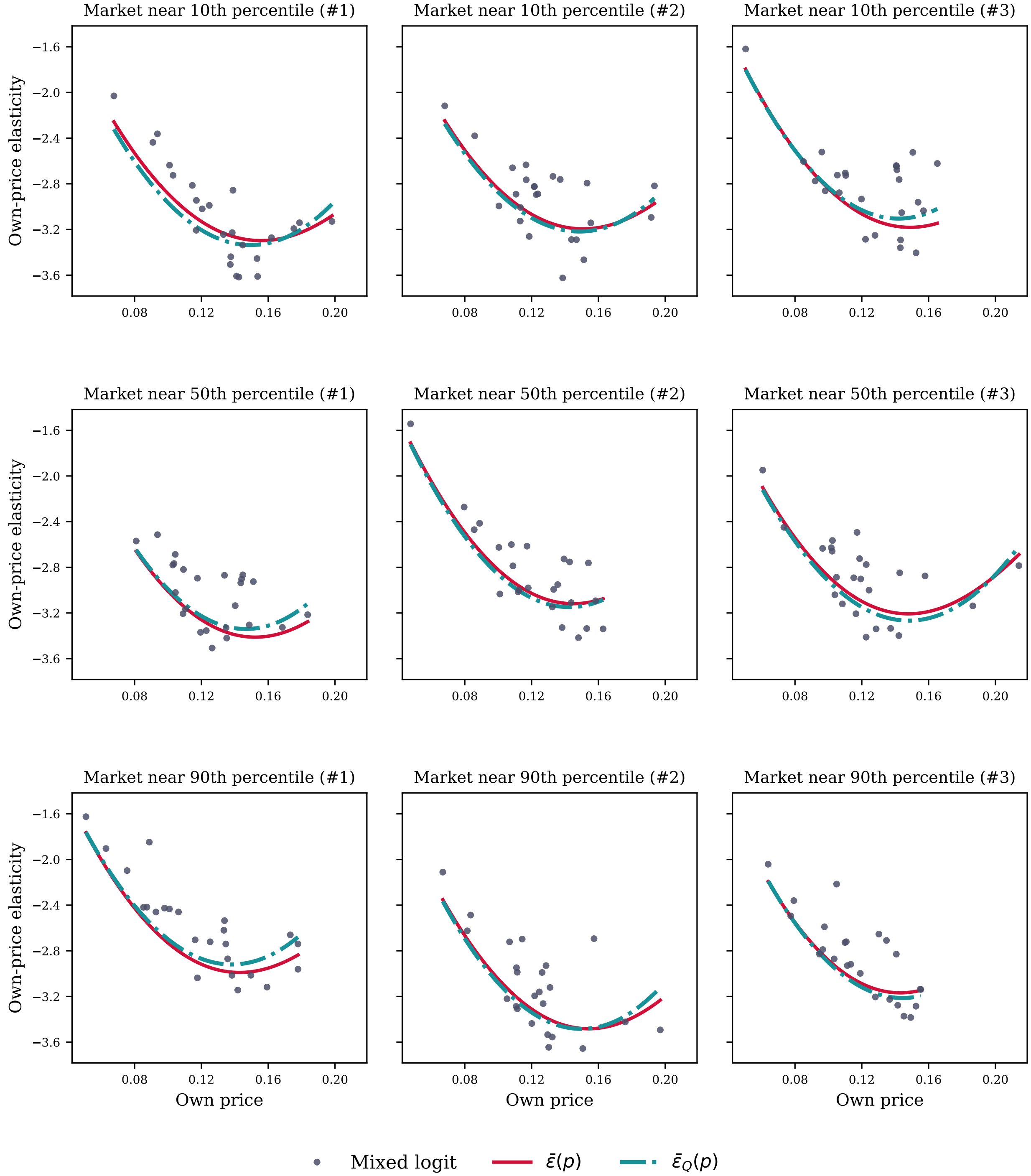
Notes: See notes for Figure SA.1.

Figure SA.3: Elasticities for nine markets (France auto)



Notes: See notes for Figure SA.1.

Figure SA.4: Elasticities for nine markets (synthetic cereal)



Notes: See notes for Figure SA.1. This figure uses the baseline synthetic cereal specification that has correlated random coefficients.

Table SA.5: Balance tests for toothpastes in lines with three price variants

	Pre-test Units Sold	Total Market Size (1000s oz.)	Brand 1	Brand 2	Peroxide	Gel	Travel Size	Avg. Per Capita Income ($1000s)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Intercept	41.076	154.320	0.315	0.308	0.120	0.080	0.053	40.633
	(0.332)	(0.708)	(0.007)	(0.007)	(0.003)	(0.002)	(0.002)	(0.119)
Variant 2	-0.023	-0.259	-0.011	0.001	0.002	-0.000	-0.000	-0.029
	(0.472)	(0.937)	(0.009)	(0.009)	(0.004)	(0.003)	(0.003)	(0.159)
Variant 3	-0.152	-0.382	-0.007	0.002	0.002	0.001	-0.000	-0.037
	(0.478)	(0.936)	(0.009)	(0.009)	(0.004)	(0.003)	(0.003)	(0.156)
Observations	596353	596353	596353	596353	596353	596353	596353	596353
R^2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
F	0.058	0.085	0.746	0.023	0.128	0.032	0.015	0.030

Notes: Each column reports a regression onto indicators for the assigned price variant. Each observation is a store-UPC pair. The sample is limited to the 161 products that are in lines with three-price variants in the pricing test. Standard errors clustered by store cluster and product line in parentheses.

Table SA.6: Balance tests for all toothpastes

	Pre-test Units Sold	Total Market Size (1000s oz.)	Avg. Per Capita Income ($1000s)
Assigned Price	0.054	-0.037	0.106
	(0.118)	(0.445)	(0.072)
Observations	617,000	617,000	617,000
UPC fixed effects	✓	✓	✓
Within R^2	0.000	0.000	0.000

Notes: Each column reports a regression onto the assigned price, controlling for UPC fixed effects. Standard errors clustered by store cluster and product line in parentheses.

Table SA.7: Compliance with the assigned toothpaste prices

	Percentile of products				
Share of stores with average price...	10th	25th	50th	75th	90th
equal to assigned	0.67	0.71	0.90	0.92	0.94
within 1% of assigned	0.72	0.75	0.95	0.97	0.97

Notes: For each of the 167 *products in the pricing test, we compute the share of stores whose average price over the test period equals the assigned price (up to a* $0.005 *rounding tolerance) and the share of stores whose average price is within 1% of the assigned price. The table reports percentiles of these two shares across products.*

Table SA.8: Summary statistics for toothpaste lines

	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

Notes: The 10th, 50th, and 90th percentiles of line-level statistics, calculated across store-weeks, for toothpaste products in the pre-test and test periods.

Table SA.9: Specification tests for the toothpaste mixed logit specifications

	% reject		Joint test	
Specification	Zero cov.	CS	Test stat.	p-value
Panel A: all products ($J = 167$)				
(1)	68.3	66.5	12.18	< 0.001
(2)	55.7	51.5	10.64	< 0.001
(3)	46.1	35.3	9.45	< 0.001
(4)	40.7	30.5	8.90	< 0.001
(5)	31.1	18.0	10.01	< 0.001
Panel B: products with at least 80% compliance ($J = 106$)				
(1)	74.5	71.7	12.18	< 0.001
(2)	59.4	53.8	10.64	< 0.001
(3)	47.2	35.8	9.45	< 0.001
(4)	42.5	30.2	8.90	< 0.001
(5)	35.8	21.7	10.01	< 0.001
Panel C: products with at least 90% compliance ($J = 78$)				
(1)	74.4	71.8	12.18	< 0.001
(2)	65.4	57.7	10.64	< 0.001
(3)	52.6	41.0	9.45	< 0.001
(4)	47.4	34.6	8.90	< 0.001
(5)	38.5	23.1	10.01	< 0.001

Notes: The first set of columns reports the percentage of products for which the product-specific t-test rejects at the 5% level, using either an assumption of zero covariance between the estimators or a conservative Cauchy–Schwarz bound. The joint test is a Bonferroni test with test statistic given by the largest absolute t-statistic across products when using the conservative Cauchy–Schwarz bound. The p-value for this test is J multiplied by the smallest single-product p-value. Panels B and C restrict the sample to products for which at least 80% or 90% of stores posted a price within $0.005 of the assigned price.